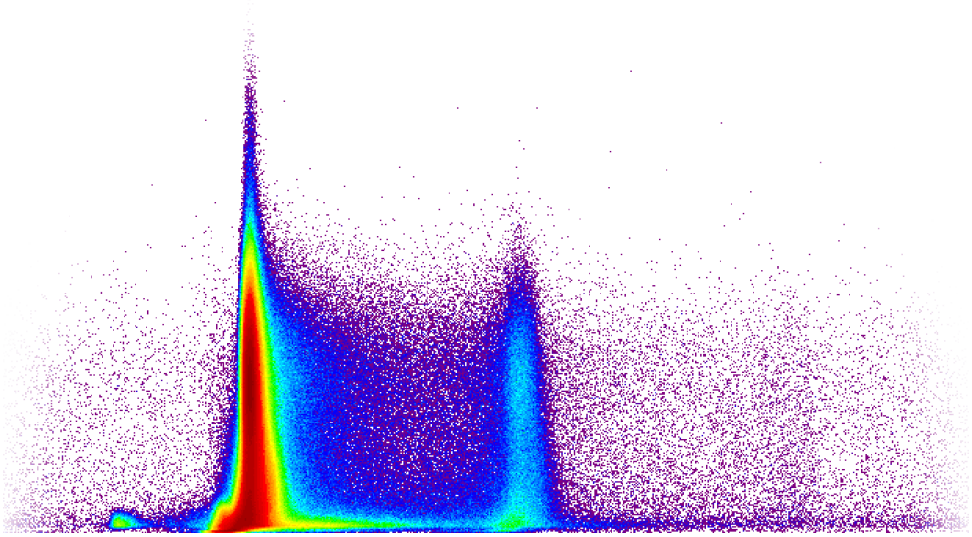
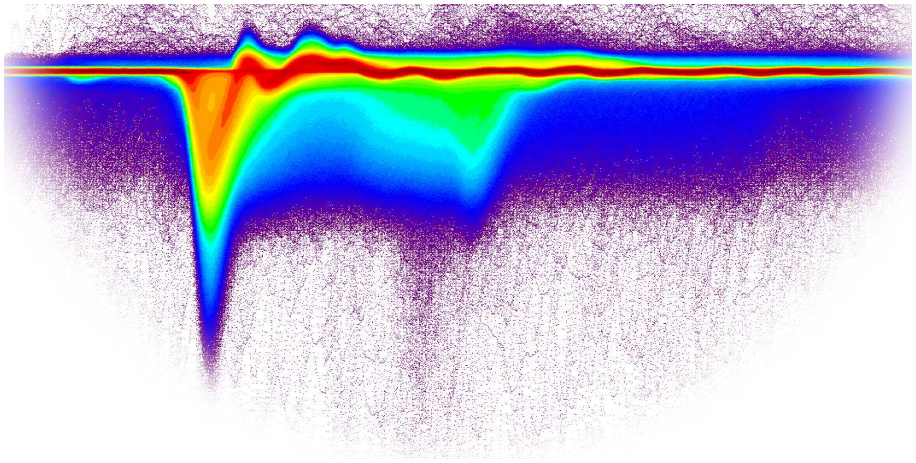


Comprehensive Photosensor Research for Large Liquid Scintillator Neutrino Detectors



Dissertation

von

Marc André Tippmann

Technische Universität München
Fakultät für Physik

Technische Universität München

Fakultät für Physik

Lehrstuhl für Experimentelle Astroteilchenphysik, E15

Prof. Dr. Lothar Oberauer

Comprehensive Photosensor Research for Large Liquid Scintillator Neutrino Detectors

Marc André Tippmann

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende/-r:	Prof. Dr. Alejandro Ibarra
Prüfende/-r der Dissertation:	1. Prof. Dr. Lothar Oberauer
	2. Prof. Dr. Andreas Ulrich

Die Dissertation wurde am 31.05.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 22.07.2021 angenommen.

Abstract

A comprehensive theory of the effects causing irregular transit times, undersized charges, early afterpulses (delay 0 to $\approx 100\,\mathrm{ns}$) and light emission in photomultipliers is presented (time shift theory, TST). Through the combination with measurements and an extensive literature research, the suitability of existing photosensor types and series for future large liquid scintillator neutrino detectors (LSNDs) was established.

TST is based on fundamental particle interactions with photomultiplier (PMT) structures and the transport processes in between them. This modular concept allows to identify all possible interaction chains, predict whether an effect produces afterpulses, and calculate the probability, charge, and time for both primary pulses and afterpulses. Ab initio calculations of 185 effects were carried out for a Hamamatsu R5912 and were found to be in excellent agreement with measurements. In every measured distribution of the primary and afterpulse properties all clusters could be explained by TST within the used approximations, and all predicted effects were observed. For the first time, the charge-resolved transit time distribution of primary pulses and the charge–delay distribution of early afterpulses were found to be fully understood. A series of new interactions was predicted and identified experimentally. These findings are supported by literature, where a subset of 15 of over 200 effects was discussed before. Published data also shows many other effects which were measured here. The great variance of measured pulse shapes was found to originate from pileup, in accordance with predictions from TST. In addition, the difference between rising and falling flank of an average pulse was explained by pileup pulses from frequent interactions on the middle dynodes. The influence of varying voltage and different PMT series observed in measurements was completely understood with a qualitative TST analysis. The theory was also extended to calculate the delays of ionic afterpulses (ca. $100\,\mathrm{ns}$–$20\,\mu\mathrm{s}$) and confirmed with experimental data. New effects from ion backscattering were postulated but could not be conclusively verified.
While developed in the context of PMTs, TST is applicable to every vacuum tube based photosensor and could be extended to encompass all photodetectors. An improved understanding of the underlying physical processes of irregular pulse formation might be used in photosensor development to reduce the time and charge resolution as well as the rates of early pulses, late pulses and afterpulses.
An extensive review of PMT properties and measurement methods was given in order to derive TST and assess the relevance for LSND detector requirements. New findings were presented for timing, early afterpulses, light emission from PMTs, ionic afterpulses, late afterpulses (delays beyond ionic afterpulses), charge, pulse shape and photosensor measurement statistics. The studies on light emission led to the development of improved light shields by Hamamatsu.

A new photosensor analysis method based on pulse fitting was developed and a setup (SPAX) focused on timing and the measurement of very small pulses was designed and constructed in order to verify and extend TST. The experimental setup is based on a dark box with a Faraday cage, a fast light source, an analog coincidence preserving the trigger time and a flash analog-to-digital converter. Compared to a classic measurement without pulse shape analysis, pulse fitting allows to recognize pileup from early afterpulses.

Pileup was found to be present for 63–68% of pulses at a threshold of 5–13% of the mean pulse height (unit pe). The mean reduced χ^2 of pulse fits ranged between 1.2 and 1.3. SPAX reached hardware thresholds of as low as 4.5% pe, which were further lowered to 1.4% pe using pulse fitting. Implementing pulse fitting in an LSND event reconstruction could allow to improve the time and charge resolution, the photodetection efficiency and photon counting capabilities.

The compliance of existing PMT series with the requirements of future LSNDs using the LENA detector as example was evaluated based on a thorough literature study and the measurement of missing properties for candidate series using two different setups. Only 2–3 series passed all criteria (ET Enterprises 9354, Hamamatsu R7081 and possibly R12860), but further measurements are required, since only one sample per series was investigated. Based on the presented findings, future Monte Carlo analyses can be carried out to assess the physics potential of a detector based on given PMT series.

In addition, a comprehensive overview of the properties of alternative photosensor types possibly relevant for LSNDs was compiled. To this end, a photosensor classification which is based on the physical interaction steps during the detection process was developed. Additionally, existing detection, amplification and readout methods were detailed, and the various hybrid photosensor types were standardized. All sensors sensitive to UV or visible light which could be found in literature – 84 sensor types with 77 subtypes and series – were analyzed for suitability for LSNDs. This is the first time that such an extensive review and compilation of sensor properties was carried out. A strict application of LENA requirements excludes all sensors but PMTs, hybrid avalanche photodiodes and possibly the IEBCMOS, which are all based on vacuum phototubes – a sensor class combining many favorable properties. Promising emerging photosensors were identified, many of which could permanently change the field of photodetection by combining excellent spatial and time resolutions. Possible approaches for future sensor developments were analyzed and auspicious existing and new sensor classes were recommended. Furthermore, two novel amplification structures are proposed which might be able to achieve a transit time spread in the single-digit picosecond range: the tynode channel plate and the cynode.

Moreover, auxiliary sensor equipment was developed for LENA based on early candidate PMT series to improve performance and compatibility.
A light concentrator (LC) optimized for detector homogeneity was devised. Here an LC which is constructed using the full detector and photosensor geometry but a slightly reduced viewing window provided the best homogeneity and light collection. A significant improvement of the detector homogeneity was achieved both by using LCs with optimized shape and by a better arrangement of sensors on the detector walls. In combination, this could almost eliminate the spatial variation of the detected light yield and coupled with a calibration with radioactive sources should notably improve the obtainable energy resolution of the detector.
Possible LC materials were studied regarding reflectance and chemical compatibility and an uncoated Ag film on an acrylic glass substrate was found to provide the best properties over prolonged times of operation.
Using these results, an optical module was designed which incorporates PMT, LC, buffer liquid, magnetic shielding and voltage divider in a pressure housing with minimum weight. A conic shape of the encapsulation proved to be best suited.

Zusammenfassung

Eine umfassende Theorie der Effekte, die irreguläre Laufzeiten, kleine Ladungen, frühe Nachpulse (Verzögerung 0 bis $\approx 100\,\mathrm{ns}$) und Lichtemission in Photomultipliern (PMTs) erzeugen, wird vorgestellt (Time Shift Theory, TST). Dies wurde in Kombination mit Messungen und einer ausführlichen Literaturrecherche verwendet, um die Eignung existierender Lichtsensortypen und -serien für zukünftige große Flüssigszintillator-Neutrinodetektoren (FSNDs) festzustellen.

TST basiert auf den fundamentalen Teilchenwechselwirkungen mit PMT-Strukturen und den Transportprozessen dazwischen. Dieses modulare Konzept erlaubt es, alle möglichen Interaktionsketten zu identifizieren, vorherzusagen ob ein Effekt Nachpulse produziert und die Wahrscheinlichkeit, Ladung und Zeit für sowohl Primärpulse als auch Nachpulse zu berechnen. Ab initio Berechnungen von 185 Effekten, die für einen Hamamatsu R5912 durchgeführt wurden, zeigten hervorragende Übereinstimmung mit Messungen. In jeder gemessenen Verteilung der Primär- und Nachpuls-Eigenschaften konnten alle Effektgruppen durch TST innerhalb der verwendeten Näherungen erklärt werden und alle vorausgesagten Effekte wurden beobachtet. Hierdurch wurden die Ladungs-Laufzeit-Verteilung von Primärpulsen und die Ladungs-Verzögerungs-Verteilung von frühen Nachpulsen erstmals vollständig verstanden. Eine Reihe neuer Interaktionstypen wurde vorhergesagt und zum ersten Mal experimentell verifiziert. Diese Ergebnisse werden durch die Literatur gestützt, wo 15 von über 200 Effekten bereits beschrieben wurden. Publizierte Daten zeigen außerdem viele weiteren der hier gemessenen Effekte. Die große Variation gemessener Pulsformen konnte wie von TST prognostiziert auf die Überlappung mehrerer Pulse zurückgeführt werden. Zudem konnte die Differenz zwischen steigender und fallender Flanke eines durchschnittlichen Pulses durch die Überlappung mit Pulsen von häufig auftretenden Interaktionen auf den mittleren Dynoden erklärt werden. Der Einfluss variierender Spannung und verschiedener PMT-Serien auf durchgeführte Messungen konnte durch eine qualitative TST-Analyse gänzlich verstanden werden. Die Theorie wurde außerdem auf die Berechnung ionischer Nachpulse (ca. $100\,\mathrm{ns}$–$20\,\mu\mathrm{s}$) erweitert und experimentell bestätigt. Neue Effekte durch Rückstreuung von Ionen wurden postuliert, konnten jedoch nicht eindeutig nachgewiesen werden.
Obwohl für PMTs entwickelt, ist TST auf alle Vakuumröhren-basierten Lichtsensoren applizierbar und könnte auf alle Photodetektoren erweitert werden. Das bessere Verständnis der zugrundeliegenden physikalischen Prozesse irregulärer Pulsbildung könnte in der Lichtsensorentwicklung verwendet werden, um sowohl die Zeit- und Ladungsauflösung als auch die Raten von Früh-, Spät- und Nachpulsen zu reduzieren.
PMT-Eigenschaften und deren Messmethoden wurden ausführlich diskutiert, um TST herzuleiten und ihre Relevanz für die Anforderungen von FSNDs abschätzen zu können. Neue Ergebnisse für das Zeitverhalten, frühe Nachpulse, die Lichtemission von PMTs, ionische Nachpulse, späte Nachpulse (Verzögerungen jenseits von ionischen Nachpulsen), die Ladung, die Pulsform und die Statistik von Photodetektormessungen wurden vorgestellt. Die Studien zur Lichtemission führten zur Entwicklung verbesserter Lichtabschirmungen durch Hamamatsu.

Um TST zu verifizieren und erweitern, wurde eine neue Lichtsensor-Analysemethode, die auf Pulsanpassung basiert, entwickelt und ein Aufbau (SPAX) für die Messung des Zeitverhaltens und von sehr kleinen Pulsen konzipiert und konstruiert. Letzterer basiert auf einer Dunkelkammer samt Faradaykäfig, einer schnellen Lichtquelle, einer analogen Koinzidenzschaltung, welche die Auslösezeit nicht beeinflusst, und einem Flash-Analog-Digital-Umsetzer. Verglichen mit einer klassischen Messung ohne Pulsformanalyse ermöglicht es die Pulsanpassung, die Überlappung von frühen Nachpulsen zu erkennen,

was für 63–68% aller Pulse bei einer Schwelle von 5–13% der mittleren Pulshöhe (pe) der Fall war. Das mittlere reduzierte χ^2 der Pulsanpassung lag zwischen 1.2 und 1.3. Mit SPAX konnten Hardware-Schwellen bis hinab zu 4.5% pe erreicht werden, welche durch Pulsanpassung bis auf 1.4% pe gesenkt werden konnten. Die Implementierung von Pulsanpassung in die Ereignisrekonstruktion von FSNDs könnte es erlauben, die Zeit- und Ladungsauflösung, Photodetektionseffizienz und Photonenzählstatistik zu verbessern.

Die Konformität existierender PMT-Serien mit den Anforderungen zukünftiger FSNDs wurde anhand des Beispiels des LENA-Detektors basierend auf einer sorgfältigen Literaturstudie und der Messung fehlender Eigenschaften für in Frage kommende Serien mit zwei Messaufbauten evaluiert. Nur 2–3 Serien erfüllten alle Kriterien (ET Enterprises 9354, Hamamatsu R7081 und eventuell R12860), allerdings sind weitere Messungen erforderlich, da nur ein Exemplar pro Serie vermessen werden konnte. Basierend auf den vorgestellten Resultaten können zukünftige Monte-Carlo-Analysen durchgeführt werden, um das Physikpotential eines Detektors für gegebene PMT-Serien zu eruieren.

Zusätzlich wurde eine umfassende Übersicht der Eigenschaften alternativer Lichtsensortypen, die potentiell for FSNDs relevant sind, zusammengestellt. Hierzu wurde eine auf den physikalischen Interaktionsschritten während des Nachweisprozesses basierende Photodetektorklassifikation entwickelt, sowie existierende Nachweis-, Verstärkungs- und Auslesemethoden beschrieben und die zahlreichen Arten hybrider Lichtsensoren vereinheitlicht. Alle in der Literatur verfügbaren auf UV- oder sichtbares Licht sensitiven Sensoren – 84 Sensortypen mit 77 Untertypen und Serien – wurden auf ihre Eignung für FSNDs analysiert. Dies ist das erste Mal, dass eine derart eingehende Behandlung und Zusammenstellung von Sensoreigenschaften durchgeführt wurde. Eine strenge Anwendung der LENA-Anforderungen schließt alle Sensoren außer PMTs, hybride Avalanche-Photodioden und möglicherweise den IEBCMOS aus, welche alle auf Vakuum-Photozellen basieren – einer Sensorklasse, die viele vorteilhafte Eigenschaften kombiniert. Aussichtsreiche neue Lichtsensoren wurden identifiziert, von denen viele das Feld der Photodetektion durch die Kombination exzellenter Orts- und Zeitauflösung nachhaltig verändern könnten. Mögliche Ansätze für zukünftige Sensorentwicklungen wurden analysiert und vielversprechende existierende und neue Sensorklassen empfohlen. Überdies werden zwei neue Verstärkungsstrukturen vorgestellt, die in der Lage sein könnten, Zeitauflösungen im einstelligen Pikosekunden-Bereich zu erreichen: die Tynode Channel Plate und die Cynode.

Um die Leistungsfähigkeit und Kompatibilität der Sensoren in LENA zu verbessern, wurden außerdem Hilfsmittel basierend auf frühen PMT-Kandidatenserien entwickelt.
Zum einen wurde ein Lichtkonzentrator (LK) konzipiert, der für Detektorhomogenität optimiert ist. Hierbei zeigte ein LK, der mittels der vollen Detektor- und Lichtsensorgeometrie jedoch für ein leicht verringertes Sichtfeld konstruiert wurde, die beste Homogenität und Sammeleffizienz. Die Detektorhomogenität konnte sowohl durch die Verwendung von LKs mit optimierter Form als auch der besseren Verteilung der Sensoren auf den Detektorwänden signifikant verbessert werden. In Kombination konnte die räumliche Variation der nachgewiesenen Lichtausbeute nahezu eliminiert werden und zusammen mit einer Kalibration mit radioaktiven Quellen sollte die erreichbare Energieauflösung des Detektors merklich verbessert werden können.
Ferner wurden in Frage kommende LK-Materialien bezüglich ihrer Reflektanz und chemischen Kompatibilität untersucht, wobei ein unbeschichteter Silberfilm auf einem Acrylglassubstrat die besten Eigenschaften über ausgedehnte Betriebszeiten zeigte.
Unter Verwendung dieser Ergebnisse wurde ein optisches Modul entwickelt, das PMT, LK, Pufferflüssigkeit, magnetische Abschirmung und Spannungsteiler in eine Druckeinhausung mit minimalem Gewicht integriert. Eine konische Form der Hülle erfüllte hierbei die Anforderungen am besten.

Contents

1 Introduction

1.1 Neutrino physics

In the last decades great discoveries have been made in the field of neutrino physics [1]. After the postulation of neutrinos in 1930 by Pauli (e.g. [3]), the first experimental proof of their existence was found by Reines and Cowan in 1956 for anti-electron-neutrinos ($\overline{\nu}_e$) [4] (awarded the Noble Prize in 1995), while muon neutrinos (ν_μ) were first detected by Lederman, Schwartz and Steinberger in 1962 [5] (Noble Prize 1988), and tau neutrinos (ν_τ) were discovered with the DONUT detector in 2001 [6]. Measurements of the Z^0 decay width at the LEP showed that there are three (light) [2] neutrino families, i.e. electron, muon and tau, which partake in the weak interaction [7].
The discovery of neutrino oscillations for atmospheric neutrinos [3] by the Super Kamiokande detector (Super-K) in 1998 [8] and for solar neutrinos by the SNO experiment in 2001 [9] proved that neutrinos are not massless [4] as predicted by the standard model of particle physics and solved the atmospheric neutrino anomaly and the solar neutrino problem (Noble Prize for both in 2015). Neutrino oscillations occur, since the eigenstates of the weak force differ from the mass eigenstates. Therefore, the three known neutrino flavors are superpositions of the three mass eigenstates which causes a quantum mechanical oscillation of the flavor states during propagation [5]. For this reason, the detected flavor of a neutrino can differ from that with which it is created after propagation. This oscillation can be described with six free parameters: a 3×3 mixing matrix (PMNS matrix [6]) with three mixing angles θ_{12}, θ_{23} and θ_{13} and a phase δ_{CP} which can lead to CP (charge–parity) violations, and two mass-squared differences [7] Δm^2_{21} and Δm^2_{32}, where $\Delta m^2_{ij} = m^2_i - m^2_j$. Oscillations are only possible if the mass differences are different from zero, which shows that the standard model has to be extended. Through the work of many experiments with complementary (anti)neutrino sources (solar, atmospheric, reactor, accelerator), distances and energies, all mixing angles and mass-squared differences have now been measured with relatively high accuracy, and first values for δ_{CP} were

[1] The content of this section is based on [1, 2].
[2] $m \leq m_Z/2$
[3] Created by the decay of pions and kaons produced by the interaction of cosmic rays with nuclei in the atmosphere.
[4] At least two neutrino masses must be different from zero.
[5] If the kinetic energy surpasses the mass difference – which is basically always the case due to the low neutrino masses.
[6] Pontecorvo–Maki–Nakagawa–Sakata matrix
[7] $\Delta m^2_{31} \approx \Delta m^2_{32}$

obtained: [1, 10, 11]

$$\begin{aligned}
\sin^2(\theta_{12}) &= 0.307 \pm 0.013 \\
\sin^2(\theta_{23}) &= 0.545 \pm 0.021 \quad \text{(NO)} \\
&= 0.547 \pm 0.021 \quad \text{(IO)} \\
\sin^2(\theta_{13}) &= (2.18 \pm 0.07) \cdot 10^{-2} \\
\Delta m^2_{21} &= (7.53 \pm 0.18) \cdot 10^{-5}\,\text{eV}^2 \\
\Delta m^2_{32} &= (2.453 \pm 0.034) \cdot 10^{-3}\,\text{eV}^2 \quad \text{(NO)} \\
&= (-2.546^{+0.034}_{-0.040}) \cdot 10^{-3}\,\text{eV}^2 \quad \text{(IO)} \\
\delta_{\text{CP}}[^\circ] &= 197^{+27}_{-24} \quad \text{(NO)} \\
&= 282^{+26}_{-30} \quad \text{(IO)}
\end{aligned}$$

where the still unknown neutrino mass ordering (MO) is denoted by NO for normal ordering ($m_1 < m_2 < m_3$) and IO for the inverted ordering ($m_3 < m_1 < m_2$).

Parallel to the determination of neutrino properties, neutrinos have also been used very successfully as messenger particles to probe the physics of their sources, and in many cases an experiment contributes to the better understanding of both neutrinos and their origins.
Solar neutrinos were first detected by Davis in the radiochemical Homestake experiment in 1970, which proved the presence of thermonuclear fusion in the Sun (Noble Prize 2002) [12] and provided the first indication for neutrino oscillations through a significantly lower measured rate of ν_e than expected (solar neutrino problem). The first real time observation of solar neutrinos (^{8}B) with energy information was performed by Kamiokande [13]. Borexino achieved the direct detection of neutrinos from all *pp* chain processes but one (*hep*) and from the CNO cycle [14–16]. Through the observation of the interaction of neutrinos with solar matter (MSW [8] effect) the sign of Δm^2_{21} was determined and it was found that $m_2 > m_1$.
Neutrinos could also be detected from the supernova 1987A, which occurred in the Large Magellanic Cloud, by the Kamiokande (12 events), IMB (8) and BUST (5) detectors (Noble Prize 2002) [17–19]. The observed energy and time distribution confirmed the theory of core collapse supernovae.
Highly energetic astrophysical neutrinos were first observed with IceCube [20] and the blazar TXS 0506+056 was the first neutrino point source which could be identified through multi-messenger observations [21].
While through these discoveries the field of neutrino astronomy is now firmly established, the application of neutrinos as probes is not limited to astrophysical objects, and KamLAND [22] and Borexino [23] also measured the flux of $\overline{\nu}_e$ from radioactive decays in the interior of the Earth (geoneutrinos).

Open questions

Despite these considerable advancements, many fundamental questions remain to be solved.

[8] Mikheyev–Smirnov–Wolfenstein

Particle physics

The absolute mass of neutrinos is not known yet. From oscillation experiments results a lower bound for the electron neutrino mass of $8.5 \cdot 10^{-3}$ eV (NO) or $4.8 \cdot 10^{-2}$ eV (IO). Cosmological observations provide an indirect upper limit of currently about 0.14 eV (95% CL [9]) for the total mass of $m_1 + m_2 + m_3$, where the limit strongly depends on the selection of measurements included in the calculations but becomes increasingly robust. This starts to put pressure on the inverted mass hierarchy, which requires a total mass of above 0.11 eV. Planned experiments probing large scale structures (DESI [24], Euclid [25], LSST [26], SKA [27]) or possibly new cosmic microwave background experiments (CMB-S4 [28], Pixie [29], CMBPol [30] or CORE [31]) should allow the detection of the total neutrino mass at 3–4 σ and can possibly indicate the mass hierarchy. The best upper limits from direct measurements are obtained from energy–momentum conservation in beta decays producing a ν_e or $\overline{\nu}_e$ by searching for distortions in the electron energy spectrum near the end-point. The KATRIN experiment (^{3}H, end-point energy 18.6 keV, effective $\overline{\nu}_e$ mass) [32] has recently started data taking and has produced the best upper limit of 1.1 eV (90% CL) with an estimated reachable sensitivity of 0.2 eV. Additional experiments comprise Project 8 [33] and measurements using ^{163}Ho (ECHo [34], HOLMES [35] and NuMECS [36]), which allows to determine the effective mass for the ν_e. The mass could also be determined by the newly proposed PTOLEMY experiment [37] which aims at the direct detection of cosmological relic neutrinos from the big bang.

The mass ordering and the mass hierarchy still need to be determined conclusively. Regarding the latter, three solutions are possible: the normal hierarchical spectrum (NH) with $m_1 < m_2 \ll m_3$, the inverted hierarchical spectrum (IH) with $m_3 \ll m_1 < m_2$, and the quasidegenerate spectrum (QD) with $m_1 \approx m_2 \approx m_3$. At the moment, data from long and medium baselines, solar neutrinos and atmospheric neutrinos disfavors the IO with 2–3 σ, depending on the analysis and the inclusion of atmospheric neutrinos [1]. Recent data, however, has weakened the preference for NO [38]. The JUNO detector [39] is constructed with the explicit goal to clarify this question via precision spectral measurement of three-flavor oscillations of reactor neutrinos. Measurements of the matter effects of accelerator neutrinos in the Earth with DUNE [40] and Hyper-Kamiokande (Hyper-K) [41] are also sensitive to the mass ordering. Further results from T2K [42] and NOvA [43] and measurements of atmospheric neutrino matter effects with KM3NeT-ORCA [44] or IceCube-Gen2 [45] will also contribute to the solution of this puzzle.

The nature of the neutrino, i.e. whether it is a Majorana or Dirac particle, also has yet to be determined. The most sensitive method is the observance of neutrinoless double beta decay ($0\nu\beta\beta$) with the reaction

$$(A, Z) \rightarrow (A, Z + 2) + 2\, e^- \tag{1.1}$$

in isotopes with double beta decay, which only is possible if the neutrino is a Majorana particle [10] and violates lepton number conservation. This also would allow to obtain information about the neutrino mass through the effective Majorana neutrino mass m_{ee}. Currently used detector types encompass enriched ultra-high-purity Ge detectors,

[9] Confidence level

[10] This introduces two additional Majorana phases to the PMNS matrix.

liquid scintillator neutrino detectors (LSNDs) [11], time projection chambers (TPCs), cryogenic bolometers, and tracker-calorimeters. Ge detectors (GERDA [46], Majorana [47]) have a good energy resolution and low backgrounds. LSNDs (KamLAND-Zen [48], SNO+ [49]) have a simple structure, can reuse existing large low-background detectors and allow to use high detector masses. TPCs (EXO-200 [50], NEXT [51]) and cryogenic bolometers (CUORE [52], CUPID [53], AMoRE [54]) can use two detection channels – ionization/heat and scintillation. Finally, tracker-calorimeters (NEMO [55]) allow a full topological event reconstruction for better background rejection and additional information about the events. So far the most sensitive results were obtained with ^{136}Xe (KamLAND-Zen, $m_{ee} < 61$–$165\,\text{meV}$ at 90% CL [48]) and ^{76}Ge (GERDA, $m_{ee} < 79$–$180\,\text{meV}$ at 90% CL [56]). Future experiments such as SNO+ (^{130}Te) will allow to further extend the sensitivity.

For δ_{CP} the accelerator experiment T2K shows first indications for CP violation ($\delta_{\text{CP}} \stackrel{!}{=} 0$ or π) but is in tension with NOvA data, which lowers the significance to 1–$2\,\sigma$. Measurements with the future detectors DUNE and Hyper-Kamiokande are expected to confirm the presence of CP violation or constrain the phase to smaller than $\mathcal{O}(10°)$.

There are several indications for one or more sterile neutrinos with masses in the eV scale, namely an excess of events in the short baseline (SBL) experiments LSND [57] and MiniBooNE [58], the Gallium anomaly in form of missing events in all Ga-based radiochemical experiments (SAGE, GALLEX/GNO) [59], and the reactor anomaly, which refers to the prediction of a 3% percent higher than measured reactor $\overline{\nu}_e$ flux by new calculations [60, 61].
For the MiniBooNE and LSND anomalies, other explanations like unrecognized background are being discussed and several experiments plan to test the reported excesses (see below). The Gallium and reactor anomalies might be caused by $\nu_e/\overline{\nu}_e$ disappearance due to oscillations to sterile flavors, but the former could also be due to uncertainties in the interaction cross section and the latter could have its origin in systematic or theoretic uncertainties. Recent measurements of the ratio between the cumulative spectra of ^{235}U and ^{239}Pu with enhanced precision show that the reactor antineutrino flux from ^{235}U has to be revised down by 5.4% [62, 63]. When considering the contribution of this isotope to the total flux, new predictions are consistent with measured fluxes. If these results are confirmed, this would resolve the reactor antineutrino anomaly and weaken the evidence for sterile neutrino oscillations in the $1\,\text{eV}$ mass range.
Since the number of light neutrino flavors coupling to the weak force (active neutrinos) was determined to be three with great certainty from the Z^0 decay width, all potential additional neutrino types with $m < m_Z/2$ must be sterile. When assuming one sterile neutrino, there is significant tension between appearance and disappearance experiments, which excludes this configuration with up to $4.7\,\sigma$. For two or more sterile flavors the exclusion significance is slightly lower depending on the analysis but the tension remains. Cosmological observations disfavor the existence of one or more sterile neutrinos which were in thermal equilibrium in the early universe.
In order to test the observed anomalies, a great number of SBL experiments are currently underway or planned which use accelerators (SBN [64], JSNS^2 [65]), reactors (NEOS [66], DANSS [67], STEREO [68], PROSPECT [69], Neutrino-4 [70], SoLiD [71])

[11] The name of the detector class is identical to that of the LSND accelerator experiment.

or radioactive sources (BEST [72]; this actually uses the upgraded SAGE detector). At the moment, the status is unclear: A large parameter space of the reactor and Gallium anomalies is excluded, but the Neutrino-4 signal is in tension with the other results.

Astroparticle and geophysics

Regarding astroparticle physics, neutrinos could also resolve many open issues.

A renewed observation of a supernova (SN) neutrino burst from a core-collapse supernova (type II supernova) in the Milky Way or its satellite galaxies with a modern detector and high statistics would allow to study both the dynamics of the supernova and neutrino properties (lifetime, magnetic moments, number of neutrino families) [73]. However, such galactic supernovae are very rare and occur with a rate of only 1–3/century.
Alternatively, the cumulative (anti)neutrino flux from all core-collapse SNe in the universe – the diffuse supernova neutrino background (DSNB) – can be measured, which would allow to obtain the probability and average neutrino energy spectrum of type II SNe [2]. The currently best limits for the DSNB flux were set by Super-K and are now close to the most optimistic flux models. JUNO and Super-Kamiokande-Gd [74] could be able to detect this neutrino source for the first time. For the planned Hyper-K and THEIA [75] hundreds of DSNB events are expected for 10 a operation.

Despite the detailed results for solar neutrinos, this field still is of large interest. The determination of the solar metallicity could resolve the discrepancy between predictions from helioseismology and the composition of the solar atmosphere. Current measurements show a mild preference for high-metallicity but cannot exclude low-metallicity [15, 76]. Furthermore, one can use the transition region between matter dominated and vacuum oscillations to probe for nonstandard interactions – i.e. new physics. [77] This transition region is also sensitive to light sterile neutrinos in the meV-scale.

No experiment has yet been able to directly measure the cosmic neutrino background, which was created in the big bang. While this in fact constitutes the most abundant neutrino flux and its presence was verified with high significance by cosmological experiments, direct detection is extremely difficult due to the small neutrino energy (mean temperature of about 2 K). First experiments are planned which aim at this ambitious goal, e.g. PTOLEMY [37, 78].

Finally, the measured geoneutrino rate still has too large uncertainties to conclusively exclude or select geophysical models. SNO+, JUNO, and the Jinping Neutrino Experiment (JNE) [79] could allow to exactly determine the contribution of radioactive decays to the heat flux of the Earth, the fractions of the U and Th chains, and the contribution of the earth crust and mantle.

1.2 Neutrino detectors

Since neutrinos interact with matter only via the weak force[12], their interaction cross-sections are very small. Thus, for neutrino detectors typically very large

[12] And gravity, but this is not suitable for direct detection of individual neutrinos.

target masses are used and, since even then the neutrino interaction rate is small, background suppression is a major issue. In order to achieve this balancing act, many different detector types with distinct advantages and requirements have been used, of which the most relevant types will be briefly introduced in the following. This review is based on [1, 2], where also the references to the publications detailing the respective detectors can be found.

Liquid scintillator detectors

A liquid scintillator neutrino detector (LSND) contains a hydrocarbon liquid which acts as the target material. At MeV energies neutrinos are detected primarily through the elastic scattering on electrons [13]. When the freed electron traverses the scintillator, a small part of its kinetic energy is converted into scintillation light through the excitation of the scintillator molecules. The photons emitted on deexcitation are detected by photosensors [14], typically photomultipliers (PMTs [15]), placed on the detector walls. One or more wavelength-shifters (WLSs) can be added to avoid re-absorption in the liquid and increase the amount of photons reaching the detector walls.
Electron antineutrinos can also be detected via the so-called inverse beta decay (IBD) for $\overline{\nu}_e$ energies above 1.8 MeV:

$$\overline{\nu}_e + p \rightarrow e^+ + n \tag{1.2}$$

The neutron is captured on a free proton after around 250 µs which produces a 2.2 MeV γ. This can be used for a delayed coincidence with the prompt signal from the positron, which allows to efficiently discriminate IBD events from background. To improve the IBD tagging the scintillator can be loaded with Gd, which has a far higher neutron capture cross-section (49 kbarn instead of 0.3 barn for the protons in the organic compound), a shorter capture time of ca. 30 µs, and produces a cascade of gamma rays with 8 MeV mean total energy.

Typically, in an LSND the target volume is surrounded by one or more layers of buffer liquid (scintillator without WLSs, mineral oil, or water), which acts as a shield for the radioactive background from the detector walls and environment, where the volumes are separated by thin transparent vessels (e.g. nylon or acrylic). Often the liquid scintillator detector is surrounded by a water Cherenkov detector (WCD, see below) which serves as both an active veto against cosmic muons and external background and as additional shielding.

LSNDs have low energy thresholds (in principle down to a few 10 keV, but limited to ca. 200 keV by the intrinsic ^{14}C contamination [80]), good energy resolution (typically around $7\%/\sqrt{E(\mathrm{MeV})}$), a position resolution of a few 10 cm at 1 MeV, and relatively low costs. For an optical coverage (OC) of the detector surface with photosensors of typically 20–40%, the number of detected photons per MeV of deposited energy (detected photon yield, dpy) amounts to a few 100 photons/MeV.

[13] Charged current / neutral current interactions with nuclei also take place but with far lower rates.
[14] Where photodetectors and composite detectors could be confused, for the former the term photosensor or simply sensor is used in this work.
[15] PMT signifies photomultiplier tube.

While at MeV energies the isotropic emission of the scintillator weakens the direction reconstruction capability, for GeV energies tracking capabilities similar to WCDs were shown [75, 81–83].
These properties make LSNDs particularly suited for the detection of low energy neutrinos. However, in this energy range radioactive backgrounds from the detector materials, the environment and cosmogenics are a serious issue and careful material selection and scintillator purification are required. The remaining backgrounds can be reduced by applying a fiducial volume (FV) cut, temporal exclusion of the region around muon tracks, use of coincidences and pulse shape discrimination (PSD).

LSNDs were in fact the first detector type to be used for neutrino detection (Cowan–Reines neutrino experiment) and have since then contributed to many important findings in neutrino physics. BUST detected neutrinos from the SN 1987A, KamLAND contributed to the solution of the solar neutrino problem and observed geoneutrinos for the first time, Borexino first directly measured the flux of ^{7}Be, pp, pep and CNO neutrinos as well as the complete solar neutrino spectrum and verified the MSW effect, and θ_{13} was determined by the Double Chooz, Daya Bay and RENO detectors. KamLAND-Zen sets the currently best limit for $0\nu\beta\beta$ and NOvA contributes to the determination of the mass ordering and δ_{CP}, while the anomalies observed by the accelerator experiments LSND and MiniBooNE[16] together with other indications (see above) prompted the search for sterile neutrinos. This, in turn, is carried out with the LSNDs NEOS, STEREO, PROSPECT and NEUTRINO-4 amongst others. Further examples of LSNDs include LVD, MACRO, CTF, CHOOZ, HWPF, and KARMEN.
Due to their favorable properties, LSNDs continue to provide valuable insights, as evidenced by detectors which are currently being constructed or are planned for the future. This includes JUNO (20 kt, under construction, see 1.3.2), which plans to determine the mass ordering with medium baseline reactor antineutrinos through an extremely good energy resolution, and THEIA (100 kt, see 1.3.3), which uses a water-based liquid scintillator to exploit both Cherenkov and scintillation light for event reconstruction; both with a broad physics program enabled by their large target volume. In addition, the Jinping Neutrino Experiment (5 kt, see p. 24) plans a precise measurement of solar and geoneutrinos as well as the detection of the DSNB and a SN burst with an extremely well-shielded detector using a slow water-based liquid scintillator to better separate Cherenkov and scintillation light. The proposed BLVST detector (10 kt, see p. 26) has similar goals. Finally, SNO+ (0.78 kt, under construction) is a $0\nu\beta\beta$-experiment, and JSNS^2 is a planned SBL detector searching for sterile neutrinos.

Radiochemical detectors

Similar to the IBD (eq. (1.2)), a ν_e can also be detected through the conversion of a nucleus via the charged current (CC) interaction

$$\nu_e + (A, Z) \rightarrow e^- + (A, Z+1) \tag{1.3}$$

Here, target isotopes are chosen such that the product isotope is unstable and decays via electron capture back to the original state (the inverse reaction of (1.3)) with

[16] MiniBooNE used both the Cherenkov and scintillation light which is produced in mineral oil for event reconstruction.

lifetimes of several ten days. The target atoms are present in form of a liquid compound, and the lifetime allows for a large fraction of the converted nuclei to extract them from the tank and detect their decays in proportional counters.

This technique was used in the Homestake experiment (conversion of ^{37}Cl to ^{37}Ar) to measure the flux of solar neutrinos with energies above 0.81 MeV, which confirmed that the solar energy is created by fusion and first detected a deficit in the measured vs. the predicted flux due to oscillations (solar neutrino problem). This was confirmed by GALLEX/GNO and SAGE (^{71}Ga→^{71}Ge) with a lower energy threshold of 0.233 MeV, which also allowed to detect the dominant pp-neutrino flux.

Radiochemical detectors are counting experiments without temporal or spatial resolution and no energy resolution except for the threshold. While they were of crucial importance in the early days of neutrino physics, they now have been superseded by other detector types.

Water Cherenkov detectors

A water Cherenkov detector (WCD) uses Cherenkov light for detection, which is produced when a charged particle moves faster than the speed of light in a medium and is detected by photosensors (usually PMTs) surrounding a detector volume of water or ice. Neutrinos can be detected via the leptons they create through CC or neutral current (NC) interactions. From the photon hit pattern on the walls the Cherenkov ring can be reconstructed, which allows to determine the direction, interaction vertex, energy and to some extent the identity (rings from e.g. electrons are blurred more than those from muons) of the charged particle, from which the neutrino properties can be inferred.

The achievable energy threshold is higher than for LSNDs (e.g. 4–5 MeV for Super-K) due to radioactive backgrounds, and a mass-dependent Cherenkov threshold prevents detection for low energies (0.8 MeV for electrons, 160 MeV for muons). At solar neutrino energies the energy resolution for electrons is about $30\%/\sqrt{\mathrm{OC}\cdot E(\mathrm{MeV})}$, the dpy is ca. $15\cdot\mathrm{OC}\cdot E(\mathrm{MeV})$ for electrons (about 6/MeV for Super-K), and the angular resolution is 25–30° (limited by multiple scattering). At GeV energies the vertex resolution is a few 10 cm, the angular resolution a few degrees, and multiple photon hits per PMT become probable which raises the electron energy resolution to ca. $2\%/\sqrt{\mathrm{OC}\cdot E(\mathrm{GeV})}$.

Three subtypes can be distinguished, depending on whether the WCD is located underground in a tank, underwater in the deep-sea or under-ice in the antarctic ice shield. Deep-sea and deep-ice WCDs use the natural medium (sea water or ice) as target to produce extremely large detector volumes measured in km^3 for the detection of high-energy astrophysical neutrinos, sometimes called neutrino telescopes.

Examples for underground detectors include Kamiokande and IMB, which detected neutrinos of the SN 1987A and observed a deficit of atmospheric $\nu_\mu/\overline{\nu}_\mu$ (atmospheric neutrino anomaly). Kamiokande also reported the first real-time measurement of solar neutrinos. Super-Kamiokande (50 kt), later solved the atmospheric neutrino anomaly by proving the existence of neutrino oscillations. SNO [17] resolved

[17] Sudbury Neutrino Observatory

the solar neutrino problem using heavy water (D_2O) as target which allowed to independently determine the ν_e and total neutrino flux through different CC and NC interactions with deuterium in addition to the elastic scattering on electrons. The K2K near-detector was also a WCD. The Hyper-K detector – the successor of Super-K – is now under construction with 260 kt mass (possibly also a second detector of equal mass [41]). A rich physics program enabled by the combination with a long baseline (LBL) neutrino beam is planned, which includes δ_{CP}, mass ordering, SN burst neutrinos, DSNB and the search for proton decay (the original goal of Kamiokande).
The primary target of neutrino telescopes is the detection of high-energy astrophysical neutrinos to study their sources and the production mechanisms of cosmic rays. They can also be used to analyze neutrino oscillations, determine the mass hierarchy with atmospheric neutrinos through matter oscillations in the Earth, and to search for dark matter (indirectly) and exotic particles. Typical energy thresholds are about 100 GeV, where lower values can be reached through denser instrumentation of a part of the detector. The outer layers can be used as a muon veto to reduce the background from atmospheric neutrinos at the expense of sensitive volume.
The DUMAND project set the stage for later underwater WCDs, while the Baikal neutrino telescope (10^{-4} km³, in Lake Baikal) was the first successful realization of this concept. ANTARES constructed a 0.01 km³ detector in the Mediterranean Sea, and together with NESTOR and NEMO (prototype tests) continues development in form of the KM3NeT[18] experiment [44]. The latter is currently being built in several steps in the Mediterranean Sea since 2013. The final configuration foresees six sparsely instrumented building blocks (ARCA) of 0.5 km³ each for astrophysical neutrinos, and a densely instrumented block studying the oscillation of atmospheric neutrinos (ORCA) to determine the mass ordering. In addition, the Baikal neutrino telescope is being extended to Baikal-GVD with 0.5–1.5 km³ volume since 2015. Furthermore, the P-ONE project plans to deploy a segmented multi-km³ detector in the Pacific Ocean [84].
AMANDA was the first ice-based neutrino telescope (0.015 km³, South Pole) and was later upgraded to IceCube with 1 km³. The latter first detected high-energy astrophysical neutrinos and could identify the first point source, the blazar TXS 0506+056, in a multi-messenger approach. DeepCore is a high-density sub-array of photosensors at the center of IceCube to lower the energy threshold. IceCube-Gen2 [45] is a planned extension with 5–10 km³ for the detection of both lower (denser DeepCore instrumentation to measure MO) and higher energies, which also includes a surface veto and a radio detector (see below).

Liquid argon time projection chambers

In a liquid argon time projection chamber (LAr-TPC) charged particles are detected both by scintillation light (S1) and the electrons created by ionization in the liquid. Through an electric field of several 100 V/cm the electrons are drifted to the surface of the liquid, where in case of a single-phase TPC their charge is read out with wire planes through charge collection or induction. In a dual-phase TPC, a stronger field extracts the electrons into a gas phase above the liquid. Here they are either amplified in thick macroscopic GEMs[19] (see p. 593) called LEMs[20] and read out

[18] Cubic Kilometre Neutrino Telescope
[19] Gas electron multiplier
[20] Large electron multiplier

with a segmented anode [40], or they are accelerated sufficiently to excite but not ionize the Ar gas, which creates an amount of scintillation light proportional to the ionization charge (S2) [85], which is detected by the same photosensors as for the scintillation light from the liquid (S1).

Independent of the LAr-TPC type, the combination of both signals enables a 3D reconstruction of the particle track with mm precision, since the ionization signal (charge or S2) is detected with x and y resolution and z can be inferred from the drift time between the prompt signal S1 and the delayed ionization signal. A drawback is that due to the long drift times (ms range) cosmic backgrounds are substantially higher than in most other detector types. While small dual-phase LAr-TPCs can reach keV energy thresholds due to the effective discrimination of background, for large detectors the achievable threshold is not well-known yet but is expected to lie in the range of 5–10 MeV [40]. This is limited by radioactive impurities – in particular Rn emanating from detector components and the intrinsic contaminations of ^{39}Ar [40; p. 53] and ^{85}Kr [86; p. 395] – and electronegative impurities in the liquid (e.g. oxygen or water) which reduce the drift times by absorbing ionization electrons and are the reason why continuous purification is required during operation [86; pp. 7–8, 451].

LAr-TPCs are a comparatively new neutrino detector type and so far only a few such detectors have been realized, among them ICARUS (600 t), ArgoNeuT (25 kg) and MicroBooNE (170 t), all single-phase TPCs used in accelerator experiments. However, the technology is very promising and will be used for the LBL experiment DUNE which is currently being prepared. The DUNE far detector is planned to consist of four modules of both single-phase and dual-phase (using LEMs) LAr-TPCs with at least 10 kt fiducial mass each [40]. This allows a rich physics program, including the determination of δ_{CP} and mass ordering, a precision measurement of neutrino mixing parameters, the observation of a SN neutrino burst and the search for proton decay [87]. SBN is a planned SBL experiment searching for sterile neutrinos which will use three LAr-TPCs at different distances, including MicroBooNE and ICARUS, which was refurbished and transported to the Fermilab for this purpose.
Comparatively small dual phase TPCs with S2 readout are also used for dark matter search with either liquid argon (LAr) or liquid xenon (LXe) as scintillator. LAr is used by WArP [88], ArDM [89](originally LEM readout [90]) and DarkSide-50 [91](Dark-Side-20k is planned [85]), while LXe is employed by ZEPLIN [92], LUX [93], LZ [94](merger of the LUX and ZEPLIN collaborations), PandaX-II, XENON1T [95](XENONnT is planned [96]) and the proposed DARWIN [97].
Some $0\nu\beta\beta$ experiments also are based on TPCs. EXO-200 [98] and its planned successor nEXO [99] use a single-phase LXe-TPC, whereas NEXT [51] (S2 readout) and PandaX-III [100](charge amplification with Microbulk Micromegas, see p. 597) will use dual-stage high-pressure gaseous Xe TPCs.

Accelerator neutrino detectors

Detectors which are constructed for the detection of a neutrino beam often have considerably different layouts than neutrino observatories, since the main design considerations are the reconstruction and containment of particle tracks. Often segmented detectors are used and a multitude of detector types has been employed,

which shall only briefly be addressed here. For a more complete list see [1; pp. 579–582, 294–297].
In the early days, spark chambers and bubble chambers were used, where the latter provided excellent spatial resolution but involved photographic film, which made data acquisition and analysis extremely slow and laborious and led to their replacement by detectors which are read out electronically.
More recently, iron tracking calorimeters have been used, which consist of alternating planes of target material (iron/steel, can be magnetized to measure momentum) and detectors (liquid/plastic scintillator or gaseous detectors, see p. 585). Examples include MINOS (later MINOS+), INGRID (on-axis near detector of T2K), and the planned 50 kt iron calorimeter at INO.
Emulsion detectors (or emulsion cloud chambers) use the excellent spatial resolution of photographic nuclear emulsion films by stacking them in alternating layers with active detectors. The latter is necessary to preselect candidate events so that the corresponding emulsion bricks can be analyzed. This technique was used to detect ν_τ for the first time in DONUT and for the first direct observation of ν_τ appearance in a ν_μ beam in OPERA, since the produced tau has only a mean path length of 87 µm due to its short lifetime. The current NINJA experiment also uses an emulsion detector.
WCDs are also used to detect accelerator neutrinos, most notably Super-K (K2K and T2K beams). The K2K near detector was also a WCD.
Scintillator detectors either use plastic scintillator (SciBar, reused for SciBooNE) or liquid scintillator (LSND, KARMEN, MiniBooNE, NOvA, JSNS² is planned) and are frequently segmented as this aids track separation.
All LAr-TPCs mentioned above which are used to detect neutrinos (ICARUS, ArgoNeuT, MicroBooNE, DUNE and SBN) are accelerator neutrino detectors.
Finally, there are hybrid detectors combining some of the listed techniques, e.g. MINERvA and ND280 (off-axis near detector for T2K).

Radio detectors [21]

For ultra high energy (UHE) neutrinos above ca 10^{16} eV the fluxes are so low that even detectors with km³ volumes are too small to detect significant rates. Such neutrinos can be produced by astrophysical objects (10^{15}–10^{18} eV), the interaction of UHE protons with cosmic microwave background photons [22] (cosmogenic neutrinos, 10^{15}–10^{18} eV), or beyond-standard model physics (above 10^{20} eV), which makes their study highly interesting. Since the attenuation length in water or ice lies in the order of several 10 m, the costs to instrument the extremely large volumes required for these observations would be prohibitively high, and new techniques were developed to allow the detection of UHE neutrinos via the emission of radio or acoustic waves (see below) by a particle cascade. These methods provide large attenuation lengths of several km, allow event reconstruction with a sparse array of sensors, and – like WCD-based neutrino telescopes – utilize natural media.

Radio detectors exploit the so-called Askaryan radiation: In an electromagnetic cascade a negative net charge accumulates from an electron vs. positron excess

[21] Content based on [1; pp. 598–601] [101; pp. 385–386].
[22] This causes the GZK (Greisen–Zatsepin–Kuzmin) cutoff.

produced by Compton scattering of gammas and positron annihilation. The relativistic movement of this net charge in the cascade front produces coherent Cherenkov radiation in the radio frequency range (100 MHz–1 GHz), which can be detected by antenna arrays for energies above 10^{16} eV. As detector media the glacial ice in Antarctica or Greenland, salt domes, or the lunar regolith can be used, all of which offer large target masses and are transparent for radio frequency signals with large attenuation lengths (a few km in ice). The energy threshold of the detector is lowest when embedded in the target medium (typically 10^{17} eV) and otherwise rises with the distance between target and sensors, although directional antennas and the combination of multiple antennas (beam-forming) can significantly improve it. The threshold also depends on background sources (anthropogenic background is particularly critical).

So far, only upper boundaries have been found for the cosmogenic neutrino flux, where above 10^{20} eV the best limits stem from the ANITA radio detector, while for lower energies Auger and IceCube provide more stringent exclusion limits.

Ice-based radio detectors include RICE (buried antennas [102]), ARA (surface and buried antennas), ARIANNA (surface antennas on the Ross Ice Shelf[23]), and ANITA (balloon-borne over Antarctica), while IceCube-Gen2 will also include a radio array and new detectors are proposed by the RNO and ARIANNA collaborations. Alternatively, the moon can be used as target, which, however, leads to high thresholds of above 10^{20} eV due to the large distance between source and antennas. The detector volume depends on the used frequency band and is larger for low frequencies (more isotropic emission and longer attenuation length), whereas higher frequencies provide lower energy thresholds due to the higher field strength. NuMoon and LOFAR [103] use low, Lunaska/Parkes high frequencies. The planned SKA radio telescope (low and medium frequencies, $\approx 1\,\mathrm{km}^2$ area) [104] is expected to lower the threshold to close to 10^{20} eV.
Finally, GRAND [105] proposes to build an air-shower-based radio detector with an area of initially $10\,\mathrm{km}^2$ (China) and later $200\,\mathrm{km}^2$ distributed over several radio-quiet locations with an energy threshold $< 10^{17}$ eV to detect cosmogenic neutrinos along with many other physics targets.

Acoustic detectors[24]

Acoustic neutrino detection is another approach for large scale detectors which might enable the measurement of UHE neutrinos with energies $\gtrapprox 10^{18}$ eV.
When an UHE neutrino interacts with a nucleus in the target medium (e.g. sea water or ice), this produces a hadronic cascade with several m lengths and a few cm radius which leads to a local heating of the medium. Depending on the expansion coefficient this expands or contracts the medium which generates characteristic pressure waves (thermo-acoustic effect). Since the creation of the particle shower is virtually instantaneous compared to hydrodynamic time scales, this leads to a coherent superposition of the sound waves, which results in an emission lateral to

[23] This allowed to also detect downward-going neutrinos via reflections of the radio waves on the ice–water surface.
[24] Content based on [106, 107] [101; pp. 385–386].

the cascade direction. The signal can be detected by piezoelectric acoustic sensors and consists of two subsequent peaks with different sign (expansion followed by contraction or vice versa [25]) with a peak around 10 kHz in the frequency spectrum in sea water. The attenuation length amounts to several km in sea water.

Tests of this technique with prototype detectors were performed in water and ice, but salt domes and permafrost have also been suggested as medium. Recently, it was also proposed by [109] to use rock instead, since the signal amplitude could be an order of magnitude larger than in water due to the higher density and speed of sound. Furthermore, the attenuation length is larger than for water, and the higher density provides larger neutrino interaction rates.

Acoustic neutrino detectors so far were based on (sea)water and ice and either used existing arrays of acoustic sensors (mostly military; e.g. ACoRNE and SAUND) or included dedicated arrays in existing WCD neutrino telescopes (IceCube (SPATS), Baikal neutrino telescope, NEMO (OνDE, SMO), ANTARES (AMADEUS)). While the size of the first generation acoustic arrays is far too small to yield competitive limits on the UHE neutrino flux compared to e.g. radio detectors [110], they can be used to develop the detection technique and investigate occurring backgrounds. For KM3NeT and Baikal-GVD [111] it is planned to include acoustic sensors to calibrate the position of the optical modules, which can also be used for UHE neutrino detection.

Coherent recoil detectors [26]

Neutrinos can interact with a nucleus as a whole through a process called coherent elastic neutrino-nucleus scattering (CEνNS) for small neutrino energies. This enhances the cross section by about a factor of N^2 where N is the number of neutrons in the nucleus. When using heavy nuclei, this allows to achieve relatively high detection rates with compact detectors. The recoil energy, however, is in the range of 1 keV or lower and decreases with the mass of the target nucleus. This makes the detection extremely challenging, since the only signal which is produced is this small nuclear recoil energy.
The first detection of CEνNS was reported in 2017 with 6.7 σ confidence by the COHERENT experiment, which used the neutrino emission of the Spallation Neutron Source at Oak Ridge National Laboratory and a low-background Na-doped CsI scintillation detector with 14.6 kg mass. The large neutrino energies of up to 53 MeV led to a high interaction cross section and the pulsed nature of the source allowed background discrimination.
Currently many projects attempt to observe CEνNS for reactor neutrinos. Due to the lower neutrino energies of typically 4 MeV the interaction cross section and recoil energies are two orders of magnitude smaller (both scale with E^2) which requires detectors which are significantly more sensitive and an excellent background control. However, the higher neutrino flux of a reactor overcompensates the smaller cross section and allows large statistics with small detector masses. This enables to test multiple scenarios beyond the standard model and could allow to build significantly

[25] For sea water the sign of both peaks *changes* at the temperature of highest density near 4 °C [108].
[26] Content based on [2, 112].

smaller low energy neutrino detectors in the future. The detector types employed in these experiments vary greatly, and include high-purity Ge-detectors (CONUS [113], νGeN [114], TEXONO [115]), cryogenic Ge-, Si- or Zn-based detectors (MINER [116], Ricochet [117]), dual-phase S2-detecting LXe TPCs (RED-100 [118]), fully depleted high-resistance CCDs [27] (CONNIE [119]) and $CaWO_4/Al_2O_3$ crystals with transition-edge sensors [28] at mK temperatures (NUCLEUS [112,120]). The last approach seems especially promising, as prototype detectors reached energy thresholds of merely 20 eV. With such ultra-low threshold detectors large statistics can be collected with g-scale detector masses and a $5\,\sigma$ CEνNS observation with a detector mass of 10 g is predicted after only two weeks of operation.

Large low energy neutrino detectors

Of the neutrino detector types introduced in the last pages, detectors with kiloton (kt) target masses can be realized with LSNDs, WCDs, LAr-TPCs, iron tracking calorimeters, emulsion cloud chambers, radio detectors and acoustic detectors.
Neutrino telescopes looking for high-energetic astrophysical neutrinos and their sources are limited to naturally occurring materials, leaving WCDs, radio detectors and acoustic detectors.
On the other hand, if the target is the real-time detection of low-energy neutrinos in the MeV range, this reduces the selection to LSNDs, WCDs, LAr-TPCs, and coherent recoil detectors.
The combination of both large target masses and low energy thresholds allows a unique multipurpose set of physics targets spanning solar, reactor, geo, SN, atmospheric and accelerator neutrinos, as well as the search for baryon number violation and exotic particles. [1; p. 592] Such large low energy neutrino detectors can be realized with LSNDs, WCDs and LAr-TPCs, of which liquid scintillator detectors reach the lowest thresholds. For good energy resolutions a high fraction of the detector surface has to be covered with photosensors, usually PMTs, which can be a significant cost factor. However, since the amount of sensors only scales with the surface, very large volume detectors can be realized with comparatively reasonable costs which can be further lowered by using light concentrators (LCs, see 5.3). To reduce backgrounds from cosmic rays and the environment, usually these detectors are placed underground and surrounded by a shielding outer liquid layer, which can also be instrumented and used as an active veto.
As was shown, LSNDs are a mature technology with many favorable properties, which continues to produce cutting-edge detectors. Therefore this work focuses on LSNDs and the optimization of their performance.

1.3 Large liquid scintillator neutrino detectors

There are five projects aiming at the realization of multi-kiloton LSNDs: LENA, JUNO, THEIA, JNE and BLVST. Of these, the detector development of LENA (white paper, two technical design studies) and JUNO (currently being built) has

[27] Compare p. 632.
[28] Compare p. 569

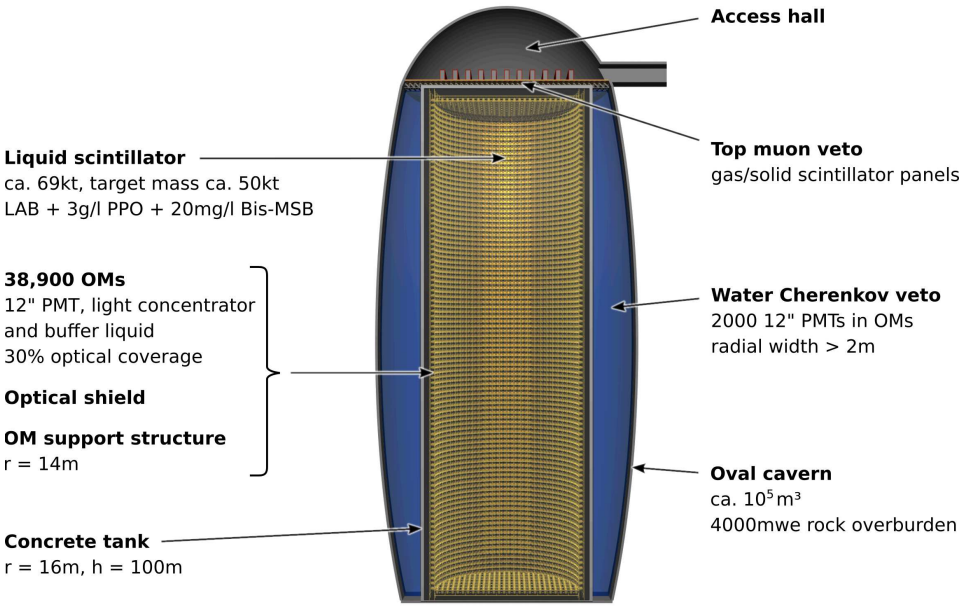


Figure 1.1: Layout of the LENA detector for the Pyhäsalmi site. Adapted from [122], courtesy of Lothar Oberauer.

progressed farthest, followed by THEIA (concept paper, white paper) and JNE (letter of intent), while BLVST is in an earlier stage of development [29]. These detectors will be shortly introduced in the following.
The present work started out in the context of LENA and was later extended to encompass all future large LSNDs, with a focus on LENA and its successor experiments JUNO and THEIA. For the studies and analyses of this work the LENA detector was used as a baseline concept.

1.3.1 LENA project

The LENA (Low Energy Neutrino Astronomy) project [30] planned to construct a next-generation LSND with an unprecedented mass scale of 50 kt fiducial volume allowing a broad physics program in a mine in Pyhäsalmi, Finland (see fig. 1.1).
After initial efforts by German research groups, an extensive investigation of viable sites and the physics potential was carried out in the framework of the LAGUNA and the succeeding LAGUNA-LBNO European design studies (FP7) [126–128]. This included a comprehensive development of the detector construction aiming towards a complete technical design. Seven possible sites were investigated, of which two showed sufficient shielding against cosmogenic background for LENA, and the Pyhäsalmi mine was selected as prime site. It could be shown that an upscaling of the well-established liquid scintillator technology to such large sizes is feasible. Unfortunately, the pursuit of a large low-energy neutrino detector in Europe ceased after the end of LAGUNA-LBNO and the DUNE experiment in the US became favored as a LBL far detector. The development efforts of the LENA detector were used for the JUNO

[29] Although first BLVST publications go back as far as 2006 according to [121].
[30] Content of this section based on [77, 81, 82, 123–125].

experiment and recently THEIA (see below), while LENA is currently not planned for construction. After extended studies and the development of new purification methods which allowed to achieve significantly greater attenuation lengths, JUNO selected a scintillator composition which is *identical* to that foreseen for LENA except for the amount of WLSs (see 1.3.2). This confirms the high level of the LENA technical design.

Detector layout

The layout of the detector is shown in fig. 1.1. The detector is centered around an upright cylindrical target volume of scintillator with 96 m height and 14 m radius, which provides a fiducial mass of about 50 kt. Due to the extremely large mass, the attenuation length of the scintillator becomes a limiting factor. Therefore a cylindrical detector shape instead of a spherical layout as for Borexino, SNO+ or JUNO was chosen to maximize the fraction of detected light for the expected attenuation lengths. The upright orientation is due to the rock conditions at the Pyhäsalmi site and provides better cavern stability. As scintillator linear alkylbenzene (LAB) with admixtures of 3 g/l PPO [31] and 20 mg/l Bis-MSB [32] as WLSs is used.
The target volume is surrounded by optical modules (OMs, see ch. 5), which each contain a Hamamatsu [33] R11780-HQE PMT with 12" diameter (see ch. 3), an attached reflective light concentrator (see 5.3), a magnetic shield and a voltage divider. All this is contained in a pressure housing, which protects the PMT from the high pressure close to the tank bottom and is filled with a non-scintillating mineral oil, which acts as a buffer liquid protecting the target volume from gamma rays emitted by the PMT glass. The entry aperture of the OMs is 14 m from the detector axis and an effective OC of 30% is foreseen.
The space between the OM apertures is closed by an optically opaque and possibly liquid-tight foil (optical shield), which separates the target volume from the 2 m thick region adjacent to the tank walls where natural radioactivity from the detector components produces large amounts of light. The outer region serves as a buffer volume shielding against radiation and contaminants emanating from the concrete tank walls.
The OMs are fixated in a stainless-steel scaffolding (OM support structure) which is mounted to the tank walls.
This inner detector (ID) is contained in a cylindrical concrete tank with an inner height of 100 m, an inner radius of 16 m and a wall thickness of 30–60 cm, which varies due to 30 cm wide cylindrical cavities inside the concrete which provide space for installations like cooling or cabling. Such a concrete tank was found to best withstand the external pressure from the surrounding WCD (water has a higher density than LAB). To ensure chemical inertness and prevent inward-leaking of water, the inside of the tank is lined with thin (1.5 mm) stainless steel plates and the outside is coated with spray-on plastic liner.
The tank is surrounded by the outer detector, a WCD serving as an active veto against inclined muons and which reduces the background from fast neutrons produced by muon interactions in the surrounding rock. For this, an oval cavern with an radial width of ≥ 2 m between tank and cavern wall is excavated, filled with water and instrumented with about 2000 12" PMTs. These sensors also are housed in

[31] 2,5-diphenyloxazole
[32] 1,4-Bis(2-methylstyryl)benzene
[33] Hamamatsu Photonics K. K.

pressure encapsulations but are not equipped with LCs to maximize their field of view. The oval cavern shape is beneficial for cavern stability.
In addition, a muon veto is placed on top of the detector to detect down-going muons and consists of an array of gas-filled detectors (drift chambers or limited streamer tubes) or plastic scintillator detectors.
The electronics is placed in a hall with ca. 15 m height above the detector and a small computer farm is located in a nearby auxiliary cavern.
The complete detector is located at a depth of 1400 m below ground, which corresponds to a cosmic ray attenuation of 4000 mwe (meter water equivalent).

Detector properties

LENA is predicted to have a ν_e energy threshold of $\approx 250\,\mathrm{keV}$ and for low energy events the energy and spatial resolution can be approximated as $6.0\%/\sqrt{E(\mathrm{MeV})}$ and $6.3\,\mathrm{cm}/\sqrt{E(\mathrm{MeV})}$ [82; p. 87].
For high-energy neutrinos, an energy resolution in the order of 9% can be reached for fully contained events [82; p. 70]. For fully contained muons with 1–10 GeV a novel iterative track reconstruction procedure obtained spatial resolutions of the track start and end point of $\lessapprox 25\,\mathrm{cm}$ lateral to the track and an angular resolution of about 1.4° at 1 GeV which improved to about 0.3° with rising energy. [83, 124] The energy resolution was found to approximately follow the function $10\%/\sqrt{E(\mathrm{GeV})} + 2\%$.

Physics potential

The physics goals of LENA exploit the low energy threshold and good energy resolution of the detector. [82, 124]
Solar neutrinos can be used to study variations of the central temperature, density and fusion rate of the Sun through variations of the neutrino flux [34], determine the solar metallicity by measuring the CNO neutrino flux, and search for non-standard interactions in the energy transition region between matter-dominated and vacuum oscillations.
A galactic core-collapse SN in 10 kpc distance would create $\approx 1.5 \cdot 10^4$ events during about 10 s, from which the ν_e, $\overline{\nu}_e$ and total neutrino flux can be distinguished through different CC and NC channels, and the temporal distribution could be obtained, which would allow a comprehensive test of SN theory and might reveal the mass ordering.
The DSNB would produce ca. 2–10 events/a from which its presence could be confirmed with $3\,\sigma$ significance after 10 a.
The geoneutrino $\overline{\nu}_e$ flux can be determined with $< 1\%$ precision after 10 a, which would significantly improve the estimates of radiogenic heat production within the Earth and give an explanation about the energy sources behind e.g. plate tectonics, and the contributions from ^{238}U and ^{232}Th and the U/Th ratio [35] can be determined with an accuracy of 2%, 4% and 6%, respectively.
With atmospheric neutrinos a high statistics measurement of θ_{23} and Δm^2_{32} can be carried out and neutrinos from below the horizon could allow to determine the mass ordering due to matter enhancement of the oscillation probabilities.
The mass ordering could also be established with a CERN-to-Pyhäsalmi LBL neutrino beam. After 10 a the inverted (normal) ordering could be rejected with 4.6–6.7 σ (4.2–5.8 σ) depending on the value of δ_{CP}.

[34] This is expected to shed new light on the mechanism behind the 11 a length of the solar cycle.
[35] This provides information on the accretion of elements in mantle and crust.

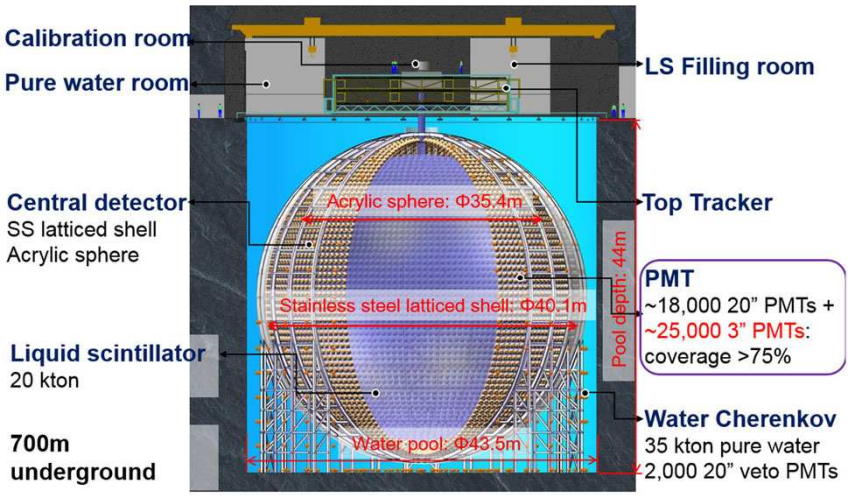


Figure 1.2: Layout of the JUNO detector, from [129].

LENA could also be used to very sensitively probe the existence of light sterile neutrinos by placing a radioactive source on top of the detector, since multiple oscillation lengths can be observed within the complete detector height of 96 m.
Finally, the large mass allows to search for proton decay channels which are difficult to access with WCDs. This would permit to either detect this long-sought process or otherwise place a stronger lower limit of $\tau > 4 \cdot 10^{34}$ on the proton lifetime after 10 a measurement.
As one can see, LENA would allow a rich physics program spanning particle, astroparticle and geophysics.

1.3.2 JUNO experiment

JUNO [36] (fig. 1.2) is a multi-purpose experiment with main focus on the determination of the neutrino mass ordering by measuring the oscillation fine structures of the energy spectrum of reactor antineutrinos after medium baselines [37].
To this purpose a LSND with 20 kt target mass is placed in a distance of 53 km to two large reactors in the Guangdong province in southern China. JUNO is currently under construction and regular data taking is foreseen to commence in 2023.

Mass ordering

In order to be able to determine the MO with a significance of 3–4 σ after 6 a, the detector must fulfill very challenging requirements. To sufficiently resolve the fine structures in the spectrum, an excellent energy resolution of $\leq 3\%/\sqrt{E(\mathrm{MeV})}$ is necessary. To achieve such low values, the statistic and systematic uncertainties must be minimized. First, the dpy should be as high as possible. A value of 1200/MeV is foreseen, for which the attenuation length of the scintillator must be > 20 m at 430 nm (peak emission wavelength), the OC of the detector should be above 75%, and photosensors with high photodetection efficiency (PDE) (above ca. 27%) have to be employed.

[36] Jiangmen Underground Neutrino Observatory; content of this section based on [39, 130–133].
[37] This is possible through the large value of θ_{13} and provides a clean measurement independent of δ_{CP} and θ_{23}.

In addition, a high statistics measurement with ca. 10^5 IBD events is needed for this sensitivity, which requires a large target mass of 20 kt. Furthermore, an energy scale uncertainty $> 1\%$ and an unknown non-linearity of the detector response could cause a false MO result, which is why a meticulous calibration will be performed and a double calorimetry with large and small photosensors is foreseen. Moreover, the baseline variation of the reactor cores must be smaller than 0.5 km, which is the case in the selected site, and the reactor flux rate and spectral shape of the antineutrino flux must be known with great accuracy, for which purpose the near detector JUNO-TAO (see below) is planned. Finally, low backgrounds are necessary, which requires e.g. scintillator purification and a large rock overburden.

Detector layout

The detector layout of JUNO is shown in fig. 1.2.
The central detector consists of an acrylic sphere with 35.4 m diameter and 12 cm wall thickness, which contains 20 kt liquid scintillator as target material. The acrylic vessel is held by a shell-like stainless steel support structure with 40.1 m diameter, on which also the photosensors for the central detector are fixed.
As scintillator currently LAB with 2.5 g/l PPO and 3 mg/l Bis-MSB is foreseen, and to achieve > 20 m attenuation length at the peak emission wavelength 430 nm, filtration in Al_2O_3 columns is necessary (an improvement from 14 m to 28 m was obtained). To reduce the amount of radioactive impurities, afterwards distillation, water extraction and steam stripping is used and the OSIRIS detector is monitoring the radiopurity of the scintillator before introduction into the detector.
The central detector is instrumented with about 17600 20" phototubes (12600 NNVT [38] MCP-PMTs [39] and 5000 Hamamatsu R12860-HQE PMTs) on a hexagonal grid, where 25600 3" PMTs (HZC Photonics [40] XP72B22) are placed in the gaps between the large sensors to achieve a total photocathode coverage of 78%. The 20" MCP-PMT was specifically developed for JUNO with the goal of a maximal PDE (on average 28.9% was reached) by using both a transmissive and a reflective photocathode and is detailed on p. 643. The R12860-HQE, on the other hand, has an excellent transit time spread (TTS) of 2.7 ns FWHM (full width at half maximum), a mean PDE of 28.1% and is treated in ch. 3. Besides increasing the OC by 2.5%, the 3" PMTs are used to cross-calibrate the 20" sensors to improve the linearity, to independently measure SN and solar neutrinos, and to extend the dynamic range of the detector for muons. All sensors have been fabricated and were tested in a massive characterization campaign [131, 134].
The central detector and the steel support structure are embedded in a cylindric WCD (called water pool) with 43.5 m diameter and 44 m height, which is filled with ca. 35 kt ultrapure water and is instrumented with about 2400 20" MCP-PMTs placed on the surface of the acrylic sphere. The WCD is used as both active muon veto and shielding and its surfaces are covered with diffusely reflective Tyvek sheets to increase the amount of detected light.
Compensation coils placed in the WCD reduce the geomagnetic field (GMF) to below 5 µT for sensors in the inner detector and < 10 µT in the WCD.
The detector top is covered to 60% by a top tracker consisting of three layers of plastic scintillator panels, which were reused from OPERA.
JUNO is located 700 m underground, which corresponds to about 1900 mwe.

[38] North Night Vision Technology Co., Ltd.
[39] Micro-channel plate PMT
[40] Hainan Zhanchuang Photonics Technology Co., Ltd.

JUNO-TAO [41]

To exclude uncertainties in the reactor $\overline{\nu}_e$ spectrum [42], a near detector called JUNO-TAO or TAO (Taishan Antineutrino Observatory) is planned in a distance of 30–40 m to one reactor core.
An extremely low energy resolution of about 2% at 1 MeV and $< 1\%$ above 3 MeV is targeted to study the fine structure of the reactor antineutrino spectrum, test nuclear databases, search for sterile neutrinos with a mass scale around 1 eV, and investigate the reactor spectrum anomalies like the 5 MeV distortion reported by Double Chooz, Daya Bay, RENO and NEOS.
TAO uses 2.6 t Gd-doped liquid scintillator as target (1 t FV) and is instrumented by ca. 10 m^2 of silicon photomultipliers (SiPMs, see p. 628) with a PDE of around 50% and an OC of $> 95\%$ to achieve a dpy of 4500/MeV. In order to reduce the dark count rate (DCR) of the SiPMs, the whole detector is cooled to $-50\,°C$. To lower the amount of neutrons from the reactor and other backgrounds, the inner detector is surrounded by a buffer of non-doped LAB, shielded by a thick layer of high density polyethylene and polyurethane, and enclosed in a muon veto in form of a WCD and plastic scintillator panels on top.
The experiment is expected to start operation in 2022.

Additional physics agenda

Naturally, such a large LSND with excellent energy resolution can also be used for a broad particle, astroparticle and geophysics program.
JUNO will allow to determine the neutrino oscillation parameters Δm^2_{21}, Δm^2_{32} and $\sin^2(\theta_{12})$ with $< 1\%$ uncertainties after 10 a and can probe the unitarity of the PMNS matrix to a level of about 1%, which could reveal the presence of hidden neutrino flavors.
Through a high statistics measurement of solar neutrino fluxes (ca. 10 k events/d), the solar metallicity could be tested, the MSW transition region could be studied and the CNO flux could be determined. However, these measurement place extremely high requirements on the radiopurity of the scintillator.
With atmospheric neutrinos a complimentary MO sensitivity can be obtained using the matter effect.
For a SN neutrino burst, the flavor content, energy spectrum and time evolution could be determined via the IBD (ca. 5000 events for 10 kpc distance) and elastic neutrino scattering (all flavors, about 3000 events).
The DSNB flux can be detected with up to 3σ significance after 10 a (about 3 events/a), which provides information about the star formation and death rates and the mean core-collapse SN neutrino spectrum.
About 400–500 geoneutrino events per year are expected, which can be used to study the composition of the Earth, its heat flux and the U/Th ratio.
Finally, one can search for proton decay via the kaon decay channel with a sensitivity to the proton lifetime of $2 \cdot 10^{34}$ a after 10 a.
Many of the secondary targets are similar to the physics program of LENA, although due the lower overburden of (ca. 1900 mwe) and the smaller size of JUNO the physics prospects for rare event searches are better for LENA.

[41] Content based on [135].
[42] For example micro-structures which could mimic the periodic oscillation structures and degrade MO sensitivity if they are not known.

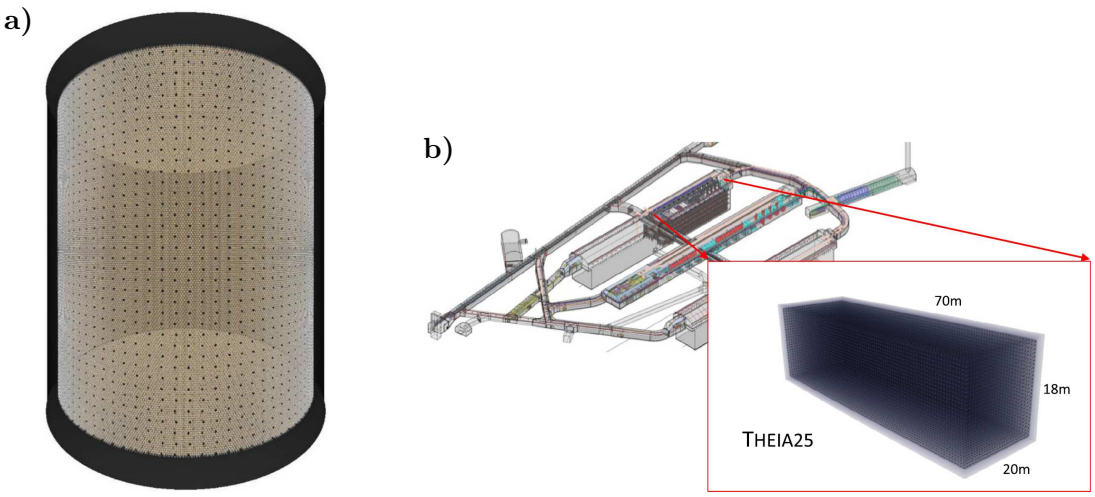


Figure 1.3: The THEIA detector: a) THEIA-100 with 100 kt water-based liquid scintillator. b) THEIA-25 with 25 kt, designed for the fourth DUNE cavern. Both models assume 86% optical coverage (OC) with 10" PMTs and 4% OC with LAPPDs (uniformly distributed). Both figures from [75].

1.3.3 THEIA project

THEIA [43] (fig. 1.3) is a planned hybrid Cherenkov/scintillation detector with 100 or 25 kt detector mass, for which several possible sites (SURF, Pyhäsalmi, Korea, ...) and experimental phases are currently being discussed [44].
This new detector type is based on the discrimination between Cherenkov and scintillation light (C/S separation), which allows to reconstruct the particle direction and identity via the Cherenkov signal (as in a WCD) while reaching the excellent energy resolution and low threshold of an LSND. The combination of both signals provides superior background discrimination both at low and high energies.
This detector class is made possible by the development of novel liquid scintillators types which produce significant levels of both Cherenkov and scintillation light, new photosensors (ultra-fast, wavelength-sensitive) and advanced image recognition methods (machine learning, pattern recognition), as will be explained in the following.

Liquid scintillator types

While already for an organic scintillator the combination of angular and time information can allow to discriminate between Cherenkov and scintillation light for high energy events [140], the attenuation lengths are far shorter than in water, the amount of Cherenkov light is comparatively low, the costs are higher, and the handling can be problematic.

Recently, water-based liquid scintillators (WbLS) were developed with the primary goal to increase the attenuation length. This is achieved by diluting a fraction of 1–10% organic liquid scintillator with ultrapure water. In order to solve the non-polar scintillator in water and form an emulsion, addition of a surfactant is needed, which results in the formation of micelles (nanoscopic droplets) of scintillator

[43] Named after the Titan Goddess of light. Content of section based on [75, 136–138].
[44] A large water-based liquid scintillator detector has first been proposed as part of the Advanced Scintillator Detection Concept [139] and was later developed further to the THEIA detector.

surrounded by surfactant. It was found that the light yield (ly) of the liquid is roughly proportional to the scintillator concentration [45] [141], which allows to adjust the light yield and energy threshold while preserving the directional capability and favorable properties of water. Compared to pure scintillators a WbLS has a better Cherenkov/scintillation light (C/S) ratio, faster emission times (shorter rise time and prompt decay constant) and attenuation lengths close to water (which can partially compensate the lower ly) and therefore is the prime candidate for the THEIA target medium.
Development and characterization of this scintillator type is ongoing. So far, studies were mostly based on LAS, a sulfonic acid derivative of LAB, which has both lipophilic and hydrophilic components and thus can act as both scintillator and surfactant at the same time. [142] Samples were stable for over 1.5 a and as a derivative of LAB, the production capabilities and costs are very attractive for large detectors. However, purification is required to reach sufficient attenuation lengths. For an already mixed WbLS this can be achieved by separating the water from the scintillator with nanofiltration and purifying both separately with standard techniques. The emission spectrum of LAS + PPO is comparable to pure LAB + PPO and the emission times are shorter than LAB + 2 g/l PPO, but PPO is not soluble in water and the micelles are only capable of binding a limited amount of PPO, for which reason pure LAB with high PPO concentrations shows shorter prompt decay constants [141, 143].
As an alternative to LAS, Triton-X100 – a surfactant with a benzene ring – could be used, which would allow PPO to become soluble in water. Triton-X100 is employed in commercial scintillator mixtures and is available with high purity for reasonable costs.

Instead of using a WbLS, the C/S separation could also be achieved by a slow scintillator, i.e. a liquid scintillator with a slow secondary fluor which increases the emission time and leads to an excellent temporal separation of both signals.

Photodetection

While a suitable scintillator type is required to provide both Cherenkov and scintillation light, both signal components must also be distinguished to provide an advantage. The C/S separation can be achieved through timing (Cherenkov light emitted instantaneously, scintillation delayed), spectrum (mostly monochromatic scintillation light, broad Cherenkov spectrum from the UV to the green) or angular distribution (heightened hit density at the Cherenkov angle, isotropic emission for scintillation). This is made possible through recent advances in photodetection technology.
If the disentanglement is to be achieved via timing, the photosensors must have a very low TTS. The newly available LAPPD [46] (see p. 641) offers ps TTS and mm position resolution, which is ideally suited for this purpose and allows to develop new reconstruction algorithms employing the correlation between arrival time and spatial distribution of photons.
Dichroic filters have a wavelength-selective transmission and can be used to separate photons by wavelengths. This is used in the dichroicon (see 4.1.2.3), which is a LC with dichroic mirrors which can be attached to e.g. an LAPPD and allows to detect long-wavelength Cherenkov photons on a different part of the (position-sensitive)

[45] In fact, at low concentrations the ly is *larger* than a linear downscaling from pure scintillator.
[46] Large Area Picosecond PhotoDetector

sensor than short-wavelength scintillation light. This even works in an undiluted scintillator.
For an angular resolution the sensor granularity must be very good, and to achieve this, affordable photosensors with high PDE and reasonably low TTS are required. Large PMTs have made significant progress with regard to quantum efficiency (QE) and TTS and many other properties in the last years (see chs. 2 and 3). OMs with multiple small PMTs and readout electronics (digital optical module) such as used by IceCube, KM3NeT, and Hyper-K offer a similar OC as large PMTs but lower spatial and temporal resolutions and in addition can resolve the direction. Furthermore, as in JUNO, multiple photosensor types can be used to combine low TTS with high OC and good granularity.
Angular resolution can also be achieved with directional light collectors. As detailed in 5.3, reflective LCs provide an angle-dependent acceptance. Alternatively, the use of lenses can allow directional sensitivity for a position-sensitive sensor or an array of small sensors behind the lens(es) [144].

Furthermore, new event reconstruction techniques allow unprecedented background rejection and effective tracking for high energy events.

Detector layout

Two possible scenarios exist for the detector layout. THEIA-100 (fig. 1.3a) has the shape of an upright cylinder with around 50 m diameter and 50 m height [47] and a total scintillator mass of 100 kt (FV depends on physics goal). Alternatively, a smaller version – THEIA-25 (fig. 1.3b) – with 25 kt mass could be deployed in the fourth DUNE cavern instead of an LAr-TPC. The detector is adapted to the cavern and takes the shape of a cuboid with 70 m length, 18–20 m width and 18 m height. THEIA-100 could achieve a broader scientific program with higher sensitivities than THEIA-25. Since the photosensors are expected to be a major cost factor, a later THEIA-100 could make use of existing instrumentation to reduce expenditure. Both variants must be located deep underground to reach their full scientific potential.
For both scenarios a phased approach is foreseen to cover a broad scientific program (see below). In the first phase, WbLS with low scintillator concentration (1–10%, water-like) is planned as target medium in combination with very fast photosensors but low OC to study LBL oscillations, the solar ^{8}B flux, proton decay and supernova neutrinos. For the second stage the detection of low energy neutrinos is targeted by using high OC (about 90%), high-concentration WbLS (oil-like) or slow scintillator – potentially loaded with ^{7}Li to increase the neutrino interaction cross-section – and possibly deploying dichroicons. This would provide a high ly, low threshold and C/S separation and would allow to also study solar (MSW transition region, CNO and pep fluxes) and geoneutrinos and resolve the flavor of SN neutrinos. In a third phase, an inner vessel with pure LAB + PPO [48] doped with a $0\nu\beta\beta$ isotope could be introduced to search for neutrinoless double beta decay on the < 10 meV scale. The density difference between pure scintillator and WbLS has to be considered in the design of the inner vessel; if possible a thin balloon would be preferable to an acrylic vessel as in SNO+ due to the easier deployment, higher transparency and lower radiological load.
So far the design is not finalized.

[47] The exact values vary between publications.
[48] Pure scintillator is used due to its higher ly and radiopurity.

Physics agenda

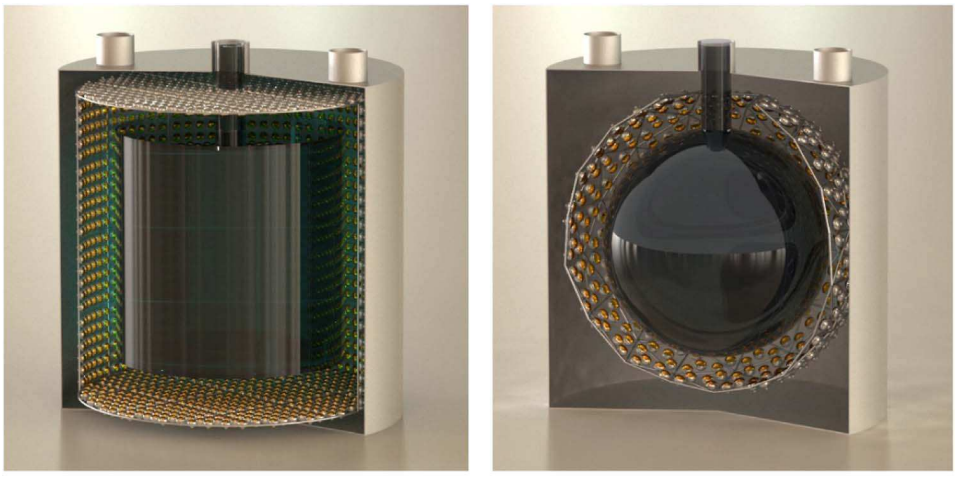
As for LENA and JUNO, a rich scientific program spanning particle, astroparticle and geophysics in a broad energy range is foreseen.
With the LBL neutrino beam in SURF, THEIA-100 could reach a sensitivity for δ_{CP} of above $5\,\sigma$ for 29% of the parameter space (64% for $3\,\sigma$) after 7 a. THEIA-25 would have a slightly better sensitivity than an LAr-TPC with 10 kt FV in the same cavern. Both variants could determine the mass ordering.
For a SN burst in 10 kpc distance over $2 \cdot 10^4$ events could be detected flavor-resolved with THEIA-100 and the SN could be located with 1° accuracy (5000 and 2° for THEIA-25). With this the temporal evolution of the spectrum, the neutronization burst from the initial SN phase (ca. 20 events) and for a close-by progenitor pre-supernova neutrinos from the Si burning could be detected.
A $5\,\sigma$ discovery of the DSNB would be possible in about 1.5 a for 100 kt (about 6 a for 25 kt) and after a few years spectral analyses become possible (for 25 kt after 10–20 a when combined with Super-K-Gd and JUNO). About 20 events per year are expected for 100 kt. This information would allow to assess the SN rate in dependence of the red shift, the equation of state of neutron stars, and the fraction of dark SNe from black-hole formation.
Regarding solar neutrinos, for 300 kt-years the CNO flux could be determined with $< 5\%$ accuracy (10% for 62.5 kt-years); with directionality the backgrounds could even be decreased substantially more. In addition, the MSW transition region can be probed with high statistics, the solar luminosity can be tested through a precision measurement of the pp and pep fluxes, the solar temperature at the core can be assessed, and the components of the CNO flux can be potentially separated to test to which extent the CNO cycle is in equilibrium. All in all, a capability similar to LENA can be reached and backgrounds can be better suppressed using the direction information.
Geoneutrinos can be detected with unprecedented accuracy (2650 events per 100 kt-years), which can be used to determine the U/Th ratio and by resolving the direction the distribution and transport of radioactive isotopes in the Earth can be analyzed.
In phase 3 (total isotopic mass of ≥ 30 t) a $0\nu\beta\beta$ lifetime sensitivity of $> 1.1 \cdot 10^{28}$ a can be reached for 211 t-years using ^{130}Te.
Finally, THEIA can search for nucleon decay using the kaon (sensitivity of $> 3.8 \cdot 10^{34}$ a for 800 kt-years) and $N \to 3\nu$ decay channels.
In many of the analyses the new photosensor types or the advantages provided by WbLS are not included yet, so the listed values are conservative estimates and the scientific potential is likely even better.

1.3.4 Other projects

Jinping Neutrino Experiment [49]

The Jinping Neutrino Experiment (JNE, fig. 1.4) proposes to build 2 LSNDs based on slow liquid scintillator with a total target mass of 5 kt in an underground laboratory (CJPL) with exceptionally low backgrounds in central China to study solar, geo, and

[49] Content based on [79, 145–149].



(a) Cylindrical scheme (b) Spherical scheme

Figure 1.4: Layout of the JNE detectors, from [150].

SN neutrinos. A 1 t prototype detector has been in operation since 2017 and a 100 t detector which is currently being designed is foreseen as a next step. Excavation of a large experimental hall (CJPL II) with ca. 10^5 m^3 is already under way.

China Jinping Underground Laboratory

The China Jinping Underground Laboratory (CJPL) is one of the world's deepest underground laboratories with a vertical rock overburden of about 2400 m and an effective overburden of 6720 mwe, resulting in a muon flux comparable to Sudbury [149]. It is located below a mountain ridge in the Sichuan province in China and is accessible via a traffic tunnel. The closest reactor is more than 900 km away and the rock has a very low radioactivity, which combined with the extremely low muon flux makes the CJPL ideal for low energy neutrino detectors, dark matter (DM) search (PandaX and CDEX are situated here) and $0\nu\beta\beta$ experiments.

Detector layout

The liquid scintillator is contained in an acrylic vessel which is either cylindrical (14 m height and diameter, FV 11.2 m for both) or spherical (14 m diameter, FV 12.8 m diameter); confer fig. 1.4. Both configurations result in a fiducial mass of 1 kt, where the spherical design has a lower surface per volume and requires less PMTs, while a cylinder offers to achieve a larger FV and seems to be favored in recent publications.
As target material either regular liquid scintillator (baseline design) or slow scintillator (water- or oil-based) will be used, possibly also WbLS. A slow scintillator would be tuned to balance the energy resolution (a dpy of 500/MeV was found to be required for the physics program) and C/S separation and would allow improved C/S separation through the different times of emission. In addition, the ly has a different dependence on particle momentum for Cherenkov and scintillation light, which can be used for particle identification. LAB with 0.07 g/l PPO and 13 mg/l Bis-MSB was studied and showed a detection window of about 3.5 ns for Cherenkov photons. First analyses of the C/S separability via fitting methods were promising but considered improvable e.g. through waveform analysis.

The central vessel is surrounded by a water buffer and viewed by PMTs on a supporting stainless steel structure placed 2–3 m from the inner vessel to shield gamma rays from the sensors. Each PMT is equipped with front-end electronics containing an FADC (flash analog-to-digital converter) and processing circuit.
The water and scaffolding are contained in a low-background steel tank with 20 m diameter and height.
Since the large depth constrains the cavern size and shape and limits the detector volume, it is planned to construct two detectors with in sum 5 kt target mass, which are placed in cylindrical excavations in the floor of the experimental hall. The total fiducial mass amounts to ca. 2 kt for solar neutrinos and 3 kt for geo- and SN neutrinos, because the enhanced background discrimination with IBD allows to use a larger FV.

Physics agenda

JNE mainly targets to study low energy neutrino fluxes.
Solar neutrinos will allow to study the MSW transition region and determine the solar metallicity with $5\,\sigma$ significance after 1500 d through a precise measurement of the CNO flux.
For geoneutrinos, rates of 31 events per kt-year are expected for U and 8 for Th, with which after 1500 d the U/Th ratio can be determined with 10% accuracy. At CJPL the geoneutrino flux is dominated by the crust (Himalaya), which can be combined with measurements of other experiments to derive the flux from the mantle.
The DSNB can be determined with $3.5\,\sigma$ after ca. 6 a and $5\,\sigma$ after about 13 a. Due to the extremely low amount of spallation induced by muons and the C/S separation, the background for DSNB neutrinos is greatly reduced in JNE. Neutrinos from a galactic SN could also be detected.
Finally, searches for DM annihilation via the produced neutrinos are possible.

BLVST [50]

The Baksan large volume scintillation telescope (BLVST, fig. 1.5) plans to deploy an LSND with regular scintillator and a target mass of around 10 kt (5–20 kt) at the BNO underground laboratory in the Baksan valley in Russia. The research activities focus on low energy neutrinos where the main scientific targets are geo, SN and solar neutrinos. Studies of a large LSND at BNO go back as far as 2006 and were recently resumed. Currently a small prototype with 420 kg target mass is under construction / operating (was scheduled to commence operation in 2020) and construction of a 5 t second prototype is planned for 2021.
The Baksan Neutrino Observatory (BNO) is located near Mount Elbrus in the Caucasus Mountains and has a large depth combined with a low reactor neutrino background (almost as low as in CJPL), which is favorable for the detection of geoneutrinos. For BLVST an overburden of 4760 mwe is planned.

Detector layout

The detector layout (fig. 1.5) currently foresees a spherical inner vessel of about 22.5 m diameter filled with a liquid scintillator as $\overline{\nu}_e$ target. This is enclosed by a

[50] Content based on [121, 151, 152].

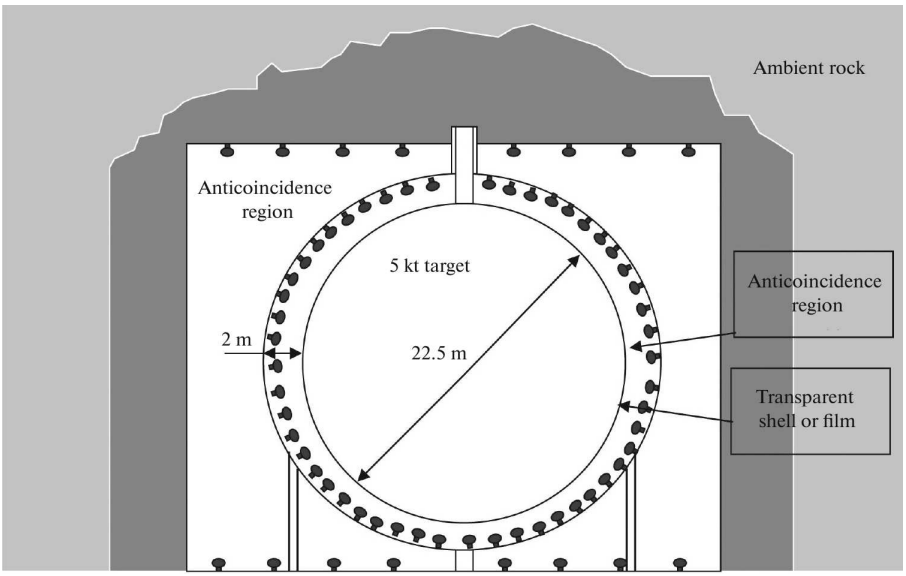


Figure 1.5: Schematic layout of the BLVST detector, from [152].

thin balloon (e.g. nylon) or a rigid casing (e.g. acrylic), which is surrounded by a buffer of 2 m thickness in form of a non-scintillating organic liquid. The buffer is contained in a metal tank at whose inner surface PMTs or other photosensors are mounted, and the tank is inside an active Cherenkov veto (water or mineral oil) instrumented with PMTs with a total size of about 33 m.
The liquid scintillator will be based on LAB (due to the abovementioned advantages) and the WLS and its concentration are chosen to minimize the temporal and thus spatial resolution. Optionally, 0.5–1 g/l Gd could be added. The BLVST group researches the maximization of the attenuation length and the minimization of ^{14}C content. For refined LAB samples the attenuation length at 420 nm could be increased from 14 m to 39.5 m, but deteriorated to 14.5 m within 1 a due to oxidation. The authors of [121] note that this could be avoided by storage in an inert gas blanket or the addition of antioxidants. ^{14}C impedes the detection of solar pp neutrinos and neutrinos from K decay in the Earth (see below), which is why it is desirable to minimize the $^{14}C/^{12}C$ ratio. So far the record set by Borexino ($2.7 \cdot 10^{-18}$) could not be surpassed.
As an alternative to PMTs, SiPM matrices are studied to improve tracking and discrimination capabilities. The high DCR of the SiPMs is compensated by using two cascaded stages of LCs [152]: The exit aperture of a truncated cone LC with a Fresnel lens in its entrance aperture is coupled to a fiber. Multiple fibers are bundled in a second set of truncated LCs, whose output is again collected in fibers, which are finally connected to an SiPM array [51].

Physics agenda

Regarding the physics potential, BLVST plans to measure geoneutrinos from U and Th (mostly from the crust) to determine the radiogenic component of the heat flux

[51] The present author would like to note that the introduction of materials with high refractive index into a liquid scintillator will likely lead to problems (see 5.3) and that with regard to performance far better LC designs are possible (also see 5.3) which would allow the omission of the Fresnel lenses.

of the Earth, estimate the potassium content in Earth's interior via elastic scattering on electrons [52], and test whether a chain fission reaction occurs at the center of the Earth.
Furthermore, the solar neutrino spectrum and the CNO rate can be measured.
For a SN burst, the analysis of the neutrino flux and spectrum can allow to explore the dynamics of the SN explosion. Furthermore, the measurement of the DSNB is targeted.
Finally, $\overline{\nu}_e$ oscillations could be studied from reactors worldwide.

1.4 Liquid scintillator detector requirements

The physics goals of an experiment impose various technical specifications on its detector(s) and in many cases the state of the art determines the physics potential. For ambitious goals thus typically additional technical R&D work is required.
Photosensor technology is of particular importance for all large LSNDs as well as many other neutrino detector types. Therefore, this work deals with the photodetection requirements of future large LSNDs and how to achieve them. Since these demands are similar for other large low energy neutrino detectors – i.e. WCDs and LAr-TPCs – the scope of the findings presented here extends beyond LSNDs, however.
The exact requirements can be quite different depending on the physics case of individual projects. For LENA and THEIA a small time resolution is crucial, whereas for JUNO a minimal energy resolution is of the utmost importance. In this work the focus is placed on timing and the more mature LENA design is used as a reference.

In the following sections, first the general requirements of a liquid scintillator neutrino detector will be explained. After this, the photodetection requirements of LENA will be discussed.

Physics level

At the physics level, there are two basic requirements: First, the properties of the detected particles need to be determined: particle identity and charge, kinetic energy, direction of flight and the particle's source. Frequently, the characteristics of a particle must be deduced from the properties of the secondary particles it produces in the detector, as is the case for LSNDs and neutrinos. These particle interactions alone already introduce inaccuracies for the obtainable primary particle properties; for example for neutrino–electron scattering one cannot deduce the exact neutrino energy from the electron energy on an event-by-event basis.
Secondly, events – interactions of the studied particles with the detector – must be distinguished from background.

Detector level

The implementation of these physics requirements sets demands at the detector level. Here the crucial questions are which information can be made available and to which

[52] $\overline{\nu}_e$ from the decay of ^{40}K cannot be detected via IBD, since their energy is too small.

extent, and how to minimize the occurrence of background. This governs the choice of detector type, layout and location, and the state of detector technology limits reachable physics results. The demands for event reconstruction and background suppression are often identical: If particle interactions can be reconstructed reliably, their signature can be used to distinguish them and then the various background types are just different event classes.

LSNDs can reconstruct the kinetic energy, vertex and interaction time of particles, and to a limited extent the track (for high energies), particle identity and charge (the latter two through shielding, event signatures and PSD).
Background stems from:

- Minute amounts of radioactive impurities in all materials inside and near the detection volume [53]
- Atmospheric muons and the spallation products they produce in or near the detector
- Fake signals produced by the photosensors (dark counts, light emission and afterpulses)
- Real signals: The signal of one analysis can constitute a background for another, for example reactor and atmospheric antineutrinos for a DSNB analysis. For accelerator neutrinos backgrounds results from beam impurities as well as NC and CC events.

Since the event signature of many background types cannot be reliably distinguished from true events, one has to employ spatial, temporal and topology cuts, which reduces the usable active detector volume to the fiducial volume (size depends on analysis) and introduces dead time, e.g. after the transition of a muon.

Thus, for LSNDs the relevant detector properties are the threshold, range, resolution and linearity of the energy, the spatial and time resolution, the tracking and PSD capabilities, the background contamination and – connected to the others – the usable active volume. In addition, all of these properties depend on the position in the detector, so their homogeneity limits the usable volume and affects the attainable average value.
The individual detector properties contribute to the physics level as follows:

- The energy information can be used to identify the source of the primary particle and to discriminate background by comparison with theoretical models and Monte Carlo (MC) data. The energy is the primary information for most physics analyses, and many background types can not be completely excluded on an event-by-event basis but only statistically from the shape of the energy spectrum.
- Time resolution allows to reconstruct the temporal structure and is critical for both event reconstruction (proton decay; SN and accelerator neutrinos) and background suppression.

[53] Among others, gamma rays from the photosensors/optical modules, the tank and the surrounding rock, alpha particles from ^{40}K and the U/Th chains from the scintillator, beta decays of ^{14}C in the scintillator, Rn emanating from the surfaces of detector structures whose decay chain produces alpha, beta and gamma rays (see e.g. [80] for an overview).

- Spatial resolution helps to reproduce the three-dimensional structure of events, which makes it important primarily for background rejection for point-like low energy events through spatial cuts.
- PSD allows to distinguish the type and charge of particles and to further suppress background from radioactive decays.
- Tracking [54] in turn enables event reconstruction and background discrimination for high energies, e.g. for accelerator neutrinos or by vetoing cosmogenic radionuclides in the volume around a muon track.

As can be seen, the detector requirements are governed to a large extent by the need to suppress the various types of strong, if not dominating, background sources. Moreover, requirements are frequently connected. For example, spatial resolution is linked to temporal resolution through the vertex-reconstruction algorithms and also affects track reconstruction capabilities.

Component level

The targeted detector capabilities in turn impose requirements on the detector components. An LSND essentially consists of scintillator, buffer liquid, photosensors, electronics, tank, and muon veto. All materials of these components which are in or near the detector have to fulfill demanding radiopurity requirements to varying degrees. Further critical demands comprise:

- The scintillator has to have a high light yield, absorption length and scattering lengths, fast emission (short first decay time and sufficient population of this decay component), and good PSD capabilities (large enough populations of longer decay times). Furthermore, its quenching properties need to be known, the fluid handling should be manageable (purification, flash point and toxicity), and a sufficient yearly production volume of the scintillator must be available.
- The electronics have to read out the photosensor signals, discriminate background, recognize and record events, as well as control and power the detector.
- By far the most requirements, however, have to be met by the photosensors:
  - The first group of requirements comprises photodetection properties. The sensors need to possess:
    - ▷ High sensitivity: high photodetection efficiency (PDE) through large values of its subcomponents; for PMTs these are quantum efficiency (QE), collection efficiency (CE) and multiplication efficiency (ME)
    - ▷ Good photon counting capabilities: low charge resolution or high peak-to-valley ratio (P/V), high gain (g), large linear and dynamic ranges
    - ▷ Good time resolution: small transit time spread (TTS), low amounts of early pulses (EPs) and late pulses (LPs), short dead time
    - ▷ Fast pulse shape: short rise time, fall time, and pulse width.

[54] This generic term subsumes various event reconstruction methods, among which also PSD algorithms may be used as part of e.g. a boosted decision tree.

Also, a sufficient fraction of the detector surface needs to be instrumented with sensitive area (optical coverage, OC).

– Additionally, the generation of background must be minimized. This includes intrinsic (dark count rate (DCR), afterpulses (APs), crosstalk) as well as exterior background sources (light emission and radioactive impurities).

– Furthermore, the sensors need to cope with environmental conditions such as the geomagnetic field (GMF), pressure and ambient temperature[55], and exhibit a long lifetime in the detector.

– Last but not least, organizational aspects are of importance. The sensor type must be available with a sufficient yearly production quantity and possess a good cost-benefit-ratio.

In the case of LENA, several requirements cannot be satisfied by available sensors themselves, as will be detailed later, and extra equipment is required to achieve compliance with specifications: Placing LCs (see 5.3) in front of the photosensors greatly improves energy resolution, vertex reconstruction, PSD and background reduction while lowering the total cost and construction time. In addition, the pressure in the tank of up to around 1.3 MPa requires the use of pressure housings (see 5.4) for the selected photosensor types (PMTs, see ch. 3 and 4). The buffer liquid can then be incorporated into the OM, which allows to dispense with an inner vessel, which would be technologically demanding and would introduce a radioactive load.

Analysis level

Finally, there is the analysis level: Based on all the demands imposed by physics, the detector and its components, the physics analyses have to be carried out on the obtained data to extract meaningful results. This requires sophisticated algorithms and computational facilities, which must be considered a part of the detector. Furthermore, the development of analysis algorithms and procedures has to occur in parallel to the definition of the technical specifications, since the physics potential and the choice of detector components depend on their efficiency.

1.4.1 Dependence on component properties

For all detector requirements listed above, their dependence on component requirements will be summarized in the following to provide an overview. Due to the complex multi-step detection and reconstruction process, there is a multitude of effects and dependencies, so these lists can only attempt at completeness. The photosensor effects mentioned in the lists will be explained shortly along with the LENA photosensor requirements in 1.4.2 and in detail in ch. 2.

[55] The moderate detector temperatures of e.g. about 23 °C in the case of LENA are not problematic for most sensor types, but exclude e.g. a blanket coverage with SiPMs due to a too high DCR.

A central quantity to characterize detector performance is the detector level variable here called detected photon yield (dpy, unit 1/MeV). This is simply the number of detected photons per MeV of energy deposited in the detector by a particle, meaning photons which produce a measurable pulse at the sensor output which is also registered by the electronics. The dpy of an LSND typically has a value of several 100/MeV.
A related parameter is the light yield (ly), which is the amount of photons produced per MeV in the scintillator and has far higher values (about 10^4 for LAB [123; p. 700]). When transport processes and the conversion of photons to photoelectrons are taken into account, the ly reduces to the photoelectron yield (pey), which is the number of photoelectrons (unit pe) produced in the sensors (e.g. PMTs) per MeV. This is not the statistical bottleneck, however, because the collection and multiplication efficiencies also have to be considered, as well as the losses incurred by discriminator thresholds and possibly readout dead time. If these factors are included, one arrives at the dpy, which truly marks the smallest amount of information carriers during event procession and therefore determines the statistics of event reconstruction. In publications often misleadingly the terms photoelectron yield (pey) or simply ly are used when dpy is meant. Since in literature detected photons are exclusively termed "pe", this notation is adopted here, although strictly speaking it would be more accurate to introduce the unit detected photons (dp) instead.

With rising dpy the relative statistical error of all reconstructed particle properties diminishes, which makes it imperative to maximize this quantity in *all* light-based detectors [56]. This is especially true for low energy events, such as the search for non-standard interactions via distortions in the spectrum of ^{8}B neutrinos. Naturally, the dpy is strongly connected to the pivotal energy information and depends on all detector components involved in photon production, propagation and detection, as well as a variety of their respective requirements.
The dpy influences the detector properties energy, spatial and temporal resolution, PSD and tracking [57].
On the other hand, the dpy depends on scintillator properties (light yield, quenching, absorption length and scattering lengths), optical coverage, transmission losses in the optical module, the use and shape of light concentrators, and sensor properties (photodetection efficiency, susceptibility to the GMF, gain (single photons must be detectable), affordability, availability and lifetime).
For LENA the baseline design foresees a mean dpy of at least 200/MeV throughout 30 years of detector operation [81].
Besides the mean dpy, the dpy variation throughout the detector is of equal importance, since it influences the same detector parameters, and a broad dpy distribution can deteriorate an otherwise good energy, spatial, and temporal resolution. The position dependence of the dpy can be rectified in part by calibrations with radioactive sources, but only to a certain degree, especially for low energy events [77]. Therefore, it is important to also optimize the dpy homogeneity. This can be achieved by

[56] The reciprocal of the dpy is the average energy loss required to create one pulse / basic information carrier. This quantity is much larger for scintillation detectors than for gas-filled or semiconductor detectors [153].

[57] In a detector which is nearly free of background (^{14}C, DCR) it would also introduce an energy threshold in the single-digit to sub-keV region due to the minimum energy required to produce and detect a photon.

large absorption and scattering lengths and particularly by using customized light concentrators (see 5.3.1).

In the following, the dependencies and limiting factors of the individual detector requirements are listed:

- **Energy resolution**
  - Mean and variance of dpy
  - Spatial resolution – due to positional variation of dpy
  - Sensor PDE spectrum and variance over sensitive area
  - Charge resolution of sensors and electronics (for the latter due to vertical resolution and sampling rate)
  - Sensor linearity – most critical for high energies
  - DCR – strong temperature dependence; since dark count pulses also create APs and photons, a large AP rate or the presence of light emission further increase the DCR (see 2.5.1 and 2.6.1)
  - Radioactive background from sensors
  - AP rate
  - Light emission from photosensors
  - Use and shape of LCs

- **Energy threshold**
  - Radioactive backgrounds
  - DCR
  - Measurable rate of data acquisition
  - Event reconstruction algorithm

- **Energy limit**
  Refers to the upper limit of the reconstructible particle energy.
  - Detector dimensions – events have to be fully contained for accurate calorimetry
  - Dynamic range of photosensors and electronics – at some point the signal saturates

- **Energy linearity**
  - Quenching factors are energy dependent
  - Sensor linearity
  - LC use and shape – LCs can focus more light onto specific sensors or shield them from peripheral photons

- **Spatial resolution**
  - Mean and variance of dpy
  - Emission time needs to be fast
  - Time resolution of sensors and electronics – the position reconstruction algorithms are based on arrival times
  - EP and LP rates must be low
  - Early afterpulses (EAPs) and ionic afterpulses (iAPs) are detrimental
  - DCR has to be low
  - LC use and shape

- **Tracking capabilities**
  - Time resolution of sensors and electronics
  - Emission time profile
  - Sensor pulse shape
  - Sensor linearity
  - Afterpulses
  - LC use and shape
  - To a lesser degree: mean and variance of dpy – at the energies allowing tracking a sufficient number of photons is produced

- **Temporal resolution**
  The requirements are similar to tracking. Since the most reliable information can be gained from the first hits, this leads to additional/changed dependencies:
  - dpy needs to be high enough to provide sufficient statistics
  - DCR
  - Linearity is not as crucial

- **PSD**
  - Mean and variance of dpy both need to be good
  - Emission time distribution
  - Quenching
  - Scattering lengths
  - Time resolution of photosensors and electronics
  - Sensor pulse shape should be narrow
  - Disrupting effects such as LPs, APs, DCR and light emission should be minimized
  - Use of LCs can greatly improve performance

- **Background suppression**
  - Radioactive impurities in all components in or near the active volume
  - Energy, spatial and temporal resolution, as well as PSD and tracking capabilities to allow efficient discrimination
  - Light concentrators (see 5.3.1)
    - ▷ Increased effective photon collection area – allows to economize sensors and move them further away from the FV, which reduces background
    - ▷ Limited field of view suppresses light from peripheral background
    - ▷ Improved PSD of alpha-decay background types
  - Pressure encapsulations – allow to construct the detector without an inner vessel, which could increase background

- **Fiducial volume**
  The usable active volume varies between analyses. In most evaluations the FV is only a fraction of the active volume due to the many background types, especially the ones introduced by the sensors.
  - Same dependencies as background suppression
  - Detection efficiency of events can be position dependent
  - The omission of a buffer volume by using pressure encapsulations allows to use a larger FV for some analyses

1.4.2 Photodetection requirements for LENA

As shown above, the photodetection requirements play a central role for detector performance and thus the accomplishment of the physics program. To be able to select suitable sensors and set exclusion parameters for production, limits for these properties have to be specified for each experiment.
LENA, JUNO and THEIA are the largest LSNDs ever planned, and the photodetection requirements for detectors with tens of kt target mass differ from previous experiments. This requires additional developments regarding photosensors or the optical modules housing them.
For the LENA detector the requirements were chosen based on experience with previous experiments (Borexino, CTF, Double Chooz), MC simulations assessing the physics potential of LENA given the expectable range of detector component properties [77, 81, 122, 123], and detector construction R&D performed in the framework of the LAGUNA and LAGUNA-LBNO collaborations. The LENA photodetection specifications are summarized in table 1.1.
In general, the requirements are designed to be sensor-independent. However, apart from common properties, each photosensor type shows unique effects which can affect the performance. These were only considered for the most relevant types. PMTs were identified as the most promising sensor type at an early stage, and this still applies as will be shown in chs. 3 and 4. Therefore several specifications are adapted to effects occurring in PMTs, such as the AP limits.
The individual specifications are discussed in the following.

Table 1.1: Photodetection requirements for existing and planned liquid scintillator neutrino detectors (LSNDs)

Property	**Borexino**	**Double Chooz**	**SNO(+)**	**JUNO** [a]	**LENA**
Active mass [t]	280	26.4	780	20k	50.8k
dpy [1/MeV]	438	200	390	1200	200
OC [%] [b]	30	13.5	59	78	30
PDE [%] [b] at	QE 21	QE 22		27 / 27 / QE 24	20
wavelength [nm]	420	400		420	420
P/V [b]	1.5	2.5	1.25	3 / 3.5 / 3	2
Gain [10^7] [b]	1	1	1	1 / 1 / 0.3	1
Linear range [pe],		[c]		1000 / 1000 / -	
max. dev. [%]				10 / 10 / -	
Dynamic range [pe]					0.3 pe/cm^2 [d]
TTS (FWHM) [ns]	3.1	4.4	4	2.7 / 12 / 5	3
Early pulse rate [%]			1.5	0.8 / - / 4.5	1
Late pulse rate [%]	4	3	1.5		7.3
Rise time [ns]		4		8.5 / - / -	
Fall time [ns]				12 / - / -	
Pulse FWHM [ns]		7			
DCR [Hz/cm^2] [d]	62	16	25	10 / 49 / 22	2.16
Afterpulse rate [%]	5	10	1.5	10 / 1 / 10	5 (0.1–200),
(delay [µs])	(0.05–12.4)	(0.1–16)		(0.5/0.5/0.05 – 20)	5 (5–100 ns)
^{232}Th [ppb]	10	90	90	400 / 50 / 400	10
^{238}U [ppb]	30	120	119	400 / 50 / 400	30
^{40}K [ppb]	2.4	24	24	40 / 20 / 200	2.4
Compressive			0.2	0.41	1.3
strength [MPa] [b]					
Magnetic field	GMF	GMF	GMF		GMF
Temperature [°C]			10	22	23
Environment	PC	Scint.	Water /	LAB / Water	LAB
	+ DMP [e]	mixture	LAB		
Lifetime [a] [b]		10	10	20	30

Average specifications from [39, 49, 77, 80, 81, 122, 123, 131, 134, 154–164], only quoted if defined, values converted to same unit. Limits must be undercut unless specified. Properties and requirements are discussed in text. SNO+ uses the same PMTs as SNO except for repair/replacement of faulty ones [49]. The first three rows show related detector properties.

[a] Photosensors: 20" PMTs / 20" MCP-PMTs / 3" PMTs

[b] Value must be exceeded

[c] Peak anode current 60 mA – not convertible without knowledge of pulse shape or height

[d] Per sensor cross-section

[e] Pseudocumene (1,2,4-trimethylbenzene) + dimethylphthalate

Detector properties

The mean dpy is required to exceed 200/MeV throughout 30 years of detector operation, taking into account photosensor failure rate and losses in both light guide reflectance and optical module transparency. At the end of the operation time a minimum dpy of 100/MeV might be sufficient. No requirement is specified for the maximum transmission loss in the optical module, since this is part of the LC design and already contained in the simulated dpy (see 5.3.1) and the degradation is analyzed separately in 5.3.2.

Optical coverage (OC) denotes the fraction of detector surface effectively covered with active photosensor area. Light-collecting reflective concentrators can increase the effective area per sensor. To fulfill the dpy requirements an OC of 30% is foreseen.

Sensitivity

The photodetection efficiency (PDE) is the probability of the successful detection of a photon hitting the sensor. Thu, the PDE is the product of the probabilities of all processes involved in the detection, such as the fraction of sensitive surface (fill factor, FF), conversion to charge carriers (QE), transport to amplification region (CE) and amplification (ME), and is wavelength dependent.
As minimum required PDE a conservative baseline value of 20% at 420 nm, the peak emission wavelength of the LENA scintillator, is chosen. In addition, the spatial non-uniformity of the PDE over the sensitive surface of the sensor should not be too large, as this increases dpy variation.

Photon counting

The charge resolution (or simply resolution) is a measure for how well photons can be counted. It is defined as the ratio of either the standard deviation or FWHM to the peak position of the single photon (spe) peak in the charge distribution and usually given in percent (see p. 119). The smaller the charge resolution, the better the peaks in the charge distribution – which correspond to the number of detected photons in the integration window – can be separated.
The peak-to-valley ratio (P/V), the ratio of the spe peak to the valley between peak and noise pedestal, is a related quantity. A small charge resolution typically corresponds to a large P/V.
For LENA the charge resolution needs to be sufficiently small, and a minimum P/V of two is required.

Gain designates the amplification factor of the sensor. A large gain allows to register spe pulses and distinguish them from electronic noise. A gain sufficient for spe detection is required; for PMTs at least a gain of 10^7 is deemed necessary.

For the reconstruction of high energy events the linear and dynamic range per sensor are key quantities. Linear range denotes the amount of simultaneously detected photons for which the charge output is still linear or has a given maximum deviation. Dynamic range is defined as the maximum number of detected photons before saturation.
For LENA a dynamic range of 0.3 pe/cm^2 sensor cross-section is seen as sufficient. For e.g. a 12”-PMT this amounts to 219 pe.

Time response

Transit time (T) denotes the time from photon incidence to pulse output and varies due to statistical and systematic effects. The time response of a sensor is mainly characterized by the spe transit time spread (TTS), which is the variation for pulses in the main peak in the transit time distribution (see 2.3.2) when using a single photon light source. It is usually given as FWHM of the peak or the standard deviation of a Gaussian fitted to it – which is only accurate close to the peak center. The FWHM therefore is more realistic.
In the LENA white paper [123] a value of 3 ns was specified for the TTS (FWHM). Most subsequent MC studies for LENA assumed a more optimistic value of 2.35 ns ($\sigma = 1$ ns) to be reachable. However, literature research and photosensor measurements (chs. 3 and 4) revealed that the required limit cannot be lowered to this value. Only two photosensor types fulfill all other requirements, namely PMTs and HAPDs (hybrid avalanche photodiodes), and a TTS of 3 ns would exclude all relevant PMT series except the Hamamatsu R6594 (5") and ETE[58] 9354 (8") – larger PMTs are more economical, however [165] – and also would eliminate the otherwise excellent HAPD series Hamamatsu R12850. Hence, the limit is maintained at 3 ns, which is satisfiable by several prime sensor candidates. Simulations show that an increased TTS affects position and temporal resolution negatively [81], so additional MC studies using the more conservative limit would be required.

Apart from the TTS, depending on the sensor type additional effects can occur which cause a systematic shift in the transit time to earlier (early pulses, EPs) or later times (late pulses, LPs). These effects are mainly known from PMTs, but are present in many photosensors (see ch. 4).
According to simulations, the event reconstruction was stable for EP/LP rates of 0.69% / 7.27% [81] (using the transit time distribution of the ET[59] 9351 (8") employed in Borexino). The PSD efficiency was unaffected by EPs, and even for an LP probability and EAP rate of 20% each the PSD efficiency was only reduced by 1% [77][60]. Therefore the limit from the white paper was kept for EPs (1%) and loosened for LPs (4% → 7.3%).
Prepulses – a very early type of EPs which in this work is denoted as $\gamma 1$ (see 2.3.5.1 and p. 282) – typically are discriminated effectively by common threshold settings, and thus no limit is set.

Pulse shape

A fast output pulse shape with short rise time, fall time and pulse width is advantageous for position reconstruction and tracking.

Background generation

Intrinsic background

In addition to pulses induced by photons, photosensors also produce noise pulses. These occur with either random times (dark count rate, DCR) – mostly caused by

[58] Electron Tubes Enterprises Ltd.
[59] Electron Tubes Ltd.
[60] That LPs and EAPs only have such a small influence is due to their correlation to the main transit time peak. They therefore cause the same changes to the pulse shape for α and β events and the difference between pulse shapes remains almost unchanged.

thermal excitation of charge carriers (see 2.5.1 for a complete list of contributions) – or correlated with and delayed to photon signals (afterpulses, APs).
The DCR limit was estimated at 15 Hz/cm^2 in the white paper. However, simulations have shown that this is not sufficient for effective PSD of ^{210}Po (important for the measurement of solar CNO and ^{7}Be neutrinos) due to the low visible energy, and a lower limit of 2.16 Hz/cm^2 is imposed [77]. The DCR has less strong effects on PSD for higher visible energies. The new limit corresponds to a mean DCR of 700 Hz for an 8" PMT, as was fulfilled in Borexino for the ET 9351 after long times in dark. The new limit could be difficult to achieve for many relevant sensors, especially high quantum efficiency (HQE) PMTs which exhibit higher DCRs. On the other hand, a higher QE might allow to reduce the OC at unchanged dpy which would lower the overall DCR.
For reliable position reconstruction in the presence of dark noise, an electronics configuration using FADCs is recommended over a conservative approach which only registers charge and first hit time [81].

Afterpulses can be caused by various effects, which are discussed in 2.5.2. The AP requirements were adapted to the iAPs (see 2.5.2.3) and EAPs (see 2.5.2.2) present in PMTs, but they set limits for every sensor type.
As mentioned above, EAPs were found to be less critical for PSD than expected [77]; however, their influence on position reconstruction has not been studied yet, so the limit for the AP rate (defined in 2.5.2.1) within 5–100 ns delay is left at 5%.
The AP rate limit in the window from 0.1 to 200 µs is also maintained at 5%. Borexino has observed correlated spurious pulses with delays peaking around 98 µs [154; p. 88], where as possible origins late afterpulses (LAPs, see 2.5.2.4) or luminescence of detector materials were suspected. To cover such effects, a large iAP/LAP window of 200 µs is chosen. The JUNO requirements originally also distinguished between fast and slow APs and in [39] the window for the latter was set at 200 µs as well. Later, however, windows of 0.5–20 µs (20" sensors) and 0.05–20 µs (3" PMTs) were seen as sufficient [131, 134].

Exterior background

Besides internal noise, photosensors also cause exterior background through gamma rays emitted by radioactive impurities in their materials (see 2.6.2) and light emission (see 2.6.1).

The radiopurity requirements were based on the Borexino experiment [80] and found to be acceptable if OMs and LCs are used [77]. Larger limits would reduce the FV for ^{8}B detection between 2–3.5 MeV too much due to ^{208}Tl.

Light emission must be prevented from causing artificial triggers and photon hits, but as of yet almost no simulations estimating its influence on performance have been carried out, so no selection limits can be quantified.
Correlated light emission through the effects described in 2.6.1 has to be investigated further. The occurrence of flashing has to be avoided at all costs by use of the precautionary measures described on p. 417.

Compatibility with environment

The peak pressure at the bottom of the tank can reach about 1.2 MPa (see p. 742). For the inner detector this is only the case during the cleaning with water, while the

outer detector is permanently filled with water. Including an extra safety margin, a requirement of 1.3 MPa is set for the compressive strength. None of the photosensors on the short list is rated for such high pressures (see 2.7.2 and table 3.7). It is therefore necessary to house the sensors in a pressure encapsulation (see 5.4).

No magnetization of the detector is foreseen, so the susceptibility to magnetic fields is limited to the GMF, which in Pyhäsalmi amounts to about 52–53 µT total over the intended lifetime of 30 a (calculated with the field model of [166]). This is problematic for many large sensor types, including PMTs, because the electron trajectories are very sensitive to the Lorentz force induced by magnetic fields, especially if they are perpendicular to the rotational axis. For PMTs even small fields such as the GMF can significantly deteriorate the CE, gain, charge resolution, P/V and TTS (see 2.7.1). As a consequence, a passive magnetic shielding is foreseen.

The detector temperature will be around 22–23 °C [122], the sensors will be contained in buffer liquid (ultrapure LAB) in the OM, and a lifetime of over 30 a is required for long-term SN observation [123].

Organizational prerequisites

Not listed in the table, but important nonetheless, is the availability of sensors in sufficient yearly production capacities to allow full detector instrumentation without delaying the construction schedule. This excludes sensors, which are currently only in the prototype stage.

Finally, the matter of affordability plays a role. While it is clear that costs need to be minimized wherever possible, the fulfillment of the physics program clearly must take precedence. It is difficult to assign a value to each sensor since this requires to weight every physics goal and evaluate its dependence on each photosensor property. To the author's knowledge, only the JUNO collaboration has undertaken this effort [167] and the used weighting factors unfortunately are not transferable to other detectors.

Other experiments

For comparison, table 1.1 also lists the requirements set by existing LSNDs (Borexino, Double Chooz) and LSNDs currently under construction (SNO+ and JUNO), as well as their active scintillator volume, dpy and OC.

Borexino is designed for solar neutrino detection, which sets stringent limits for the radiopurity, and requires a good TTS and a rather high mean dpy for improved energy resolution.

Double Chooz was constructed for the measurement of reactor antineutrinos, which allows for less strict limits on dpy, radiopurity, TTS and APs, but requires a good P/V, short pulse shape and a not too high DCR.

SNO+ is a special case, as here a previous WCD (SNO) is converted to a Gd-loaded scintillator detector for double beta decay measurements and the existing photosensors are reused [49]. Therefore the SNO requirements are given instead,

while dpy and active mass are quoted for SNO+. The limits are moderate with the exception of low EP, LP and AP rates.

The JUNO experiment aims to determine neutrino mass ordering among other goals (see 1.3.2). To achieve this, an extremely low energy resolution of only 3% at 1 MeV is needed, which in turn requires an ambitious dpy of about 1200/MeV, setting a new record. The detector requirements are dominated by this goal and so the OC, PDE and P/V have to be very high and the detector dimensions impose moderate compressive strengths, which are planned to be met by housing the PMT stems. The other requirements are somewhat mixed, since three different sensor types will be used, which can compensate each others' drawbacks (TTS, DCR, APs, radiopurity), so the limits represent the best achievable values for each type. All in all, a high linear range, for the 20" PMTs very good TTS and very low EP rates, but only moderate radiopurity are required. The GMF will be compensated by coils around the ID.

Comparing LENA to the other detectors, most requirements are modest to moderate and easily satisfiable using modern PMTs with even standard QE (SQE). The required dpy is comparatively low, allowing for extra margins, which is reasonable considering the long foreseen operation time. Accordingly, an average OC is sufficient. PDE, P/V and LP rate are very conservative, the gain and iAP rate are at typical values, and the dynamic range is lower than the linear range that is required for JUNO. The EP limit is rather strict, but might be extensible following further studies.
Several requirements might pose challenging, however: The DCR has to be extremely low for sufficient background suppression for CNO and ^{7}Be solar neutrino measurements; the TTS has to be as good as possible for position reconstruction; the radiopurity requirements are at the same level as Borexino but for an 180 times larger detector; the lifetime is the longest of all listed LSNDs in order to serve as an SN observatory; and the pressure weighing on the sensors is by far the highest due to the large detector mass and the upright cylindrical shape resulting from the attenuation length and rock properties. Also, the novel limit on the EAP rate requires independent sensor measurements, since this is typically not checked.
While several properties are demanding, the LENA photosensor requirements are considered obtainable with existing devices by careful sensor selection and the use of supportive equipment in form of OMs and LCs. It will be shown in this work that this is in fact the case.

1.5 Thesis overview

Because the physics case of future large liquid scintillator neutrino detectors (LSNDs) is in large part determined by photosensor performance, the purpose of this work is to study how the demanding photodetection requirements of these and similar detectors can best be fulfilled and to aid the progress of photosensor technology. Thus far photomultipliers (PMTs) are considered the baseline photosensors for all experiments, since they offer a unique combination of favorable properties, so a focus will be put on this sensor type.

This research is centered on the formulation of a theory of PMT behavior based on fundamental particle interactions, which allows a better understanding of internal processes and is directly applicable to all photosensors based on vacuum tubes. With this tool possible sensor improvements can be studied analytically, i.e. without simulations or prototyping, and the definition of detector requirements can be refined. Based on this theory, existing PMT series and other current and future photosensor types are analyzed with regard to their suitability for large LSNDs using the example of LENA, but the results should be mostly applicable to water Cherenkov detectors (WCDs) and liquid argon time projection chambers (LAr-TPCs) as well. As another means to improve and enable photodetection performance, auxiliary equipment for PMTs was developed. This comprises light concentrators optimizing the detector homogeneity and minimum weight pressure housings.

In the current chapter, first a short summary of the state and future potential of neutrino physics was given and existing techniques of neutrino detection were discussed. For the measurement of small neutrino fluxes at low energies, LSNDs were found to be best suited. Planned future LSNDs were presented and their detector requirements were addressed, which showed the importance of photosensors for these detectors. The sensor requirements of the LENA project – a far advanced detector design – were listed together with a brief explanation of the respective photosensor properties.

The following chapters will examine which sensor types and series fulfill these requirements based on measurements and literature research.

PMTs have been the sensor of choice for previous neutrino detectors and still are – among emerging new sensor types – as will be seen. Therefore, in chapter 2 first an overview of relevant PMT properties is provided. The characteristics of other shortlisted sensor types are in large part identical, since all are based on vacuum tubes. This chapter is intended as a compilation and critical review of the current theoretical and practical knowledge of photomultiplier properties. It is far more than just a collection of previous knowledge, however. In this work a focus was put on the analysis of timing (see 2.3), early afterpulses (see 2.5.2.2) and light emission from PMTs (see 2.6.1). These studies culminated in the formulation of a comprehensive theory of effects causing a) shifts in the transit time (time shift effects, TSEs) and b) early afterpulses (EAPs), which was called time shift theory (TST) and is presented and discussed in 2.3.5. This theory allows the prediction of expected effects and the calculation of all relevant properties with remarkable agreement with measurements. Previous knowledge is reappraised and extended in light of these findings throughout ch. 2 and all results from timing, EAP and light emission measurements are summarized here, since they further the understanding of PMT theory. Thus, this chapter contains many of the main results of this work.

After establishing the background knowledge necessary to assess PMT properties, the compliance of available PMT series with LSND requirements is studied in ch. 3. First existing series are evaluated in 3.1 based on an in-depth literature research to preselect candidate series.
Several requirements were newly identified for LENA or had not been studied in detail for the selected PMT series before, especially EAPs having delays from only a few ns to several tens of ns, light emission from PMTs, and timing effects. For this

reason, specimens of these series were measured using first an existing setup from the Borexino PMT characterization and then a specialized improved setup developed at the Technische Universität München (TUM) for this work (see 3.2). The setups and data analyses are described in 3.2.1 and the obtained results are presented in 3.2.2. Based on the measurements and literature research, the compliance of all considered series with the requirements for LENA is discussed in 3.3.

Following this, alternative photosensor types are examined for their suitability in chapter 4.
First an overview over existing photosensor types and their detection mechanisms is given in 4.1 before the properties of visible/UV-sensitive sensors are compiled in 4.2–4.4 by means of an extensive literature research of existing measurements. Finally, in 4.5 these sensors are analyzed for compliance with the requirements of LSNDs using the collected data and the most suited series are identified.
Of all sensors, PMTs still best fulfill the requirements, together with some hybrid avalanche photodiode (HAPD) series and possibly the IEBCMOS[61], and there is a number of promising sensors which are expected to become competitive in the near future. In addition, some ideas for novel photosensor classes and amplification structures are introduced.

The final chapter (ch. 5) analyzes the benefits for photosensor performance and compliance obtainable through additional equipment. This is done by developing an optical module (OM) for LENA and a specific PMT series which consists of a PMT, a light concentrator (LC), and shielding buffer liquid inside a pressure housing. To provide an overview, first the resulting layout of the OM is presented in 5.2. This is followed in 5.3.1 by a description of the development of a reflective non-imaging LC which minimizes the variation of the reconstructed energy in the detector and increases the effective photosensitive area. Optimized arrangement of the OMs on the detector walls was found to further significantly improve detector homogeneity. For suitable reflective LC materials the reflectance was measured and the chemical compatibility with the buffer liquid was ascertained via accelerated aging (detailed in 5.3.2). Finally, the development of the pressure housing through a finite element analysis (FEA) is described in 5.4.

The work closes with a summary of all findings and results.

For the reader's convenience, a list of abbreviations can be found following the conclusion and before the list of figures, list of tables, and the bibliography.
Each chapter can be read separately and contents of other chapters are referred to where needed. In ch. 2 – the main chapter – also every *section* (i.e. 2.1, 2.2, ...) can be read independently from the rest of the work.

[61] Intensified electron bombarded CMOS

2 Photomultiplier properties

The basic layout and operation principle of a photomultiplier (PMT) can be seen in fig. 2.1. The device is based on a glass vacuum tube.
An incident photon is converted into a photoelectron (pe) through external photoelectric effect in the photocathode (short cathode or k), which consists of a thin semiconductor film with a low work function evaporated onto the inside of the entrance window [168; ch. 12] [169].
The photoelectrons are then accelerated and guided by electric fields into an electron multiplier, in which they are amplified through secondary emission by repeated acceleration onto dynodes (dn), before collecting the electron avalanche at the anode (a), typically a fine mesh about 1 mm in front of the last dynode (reflex anode) [170; pp. 339–340].
This yields a current resp. voltage spike at the output, which for high gains is detectable without further amplification.
A voltage divider outside the vacuum tube (not shown) supplies the voltages to the different components to control the accelerating and focusing fields.
An extensive treatment on the function and properties of photomultipliers and their various subtypes is given in the various, excellent photomultiplier handbooks [169–173].

Photomultipliers are capable of detecting single photons with low time uncertainty, large sensitive areas, high gain, low noise [1], and large dynamic range at comparatively low cost. With the exception of multianode PMTs they are not image sensors, meaning each sensor only constitutes one pixel of detection. Major drawbacks are that they must be shielded from even low magnetic fields, such as the geomagnetic field (GMF), and ambient light.

Invented around 1934 [170; p. 3], photomultipliers can be considered very mature devices. They continue to evolve, however, and even during the past few years there have been large improvements in nearly all relevant parameters such as quantum efficiency [174–176], collection efficiency [177], transit time spread [178, 179], ionic afterpulses [177, 180], light emission [181, 182], radioactive background [80, 162, 183] [170; pp. 274–276], dynode structures and susceptibility to magnetic fields [184; chs. 4.2.1 and 9], pressure capability [185; pp. 3-101–3-125] [165, 186] [41; pp. 113–124], and operation at cryogenic temperatures [187, 188].
All in all, their properties make them favorable for a large number of applications, which explains their use in areas as diverse as high energy physics and astronomy, medical imaging and diagnostics, biotechnology, spectroscopy, LIDAR [2], surface analysis, thickness and level measurement, image scanners, radiation monitoring, pollution monitoring, as well as oil and gas production [171, 189].

[1] [170; pp. 5–6] remarks that the principal attribute of an electron multiplier is that it amplifies currents *without* a resistor – Johnson noise is only introduced by the external load resistor and the PMT otherwise is merely subject to shot noise from the quantized nature of light and charge.
[2] Light detection and ranging

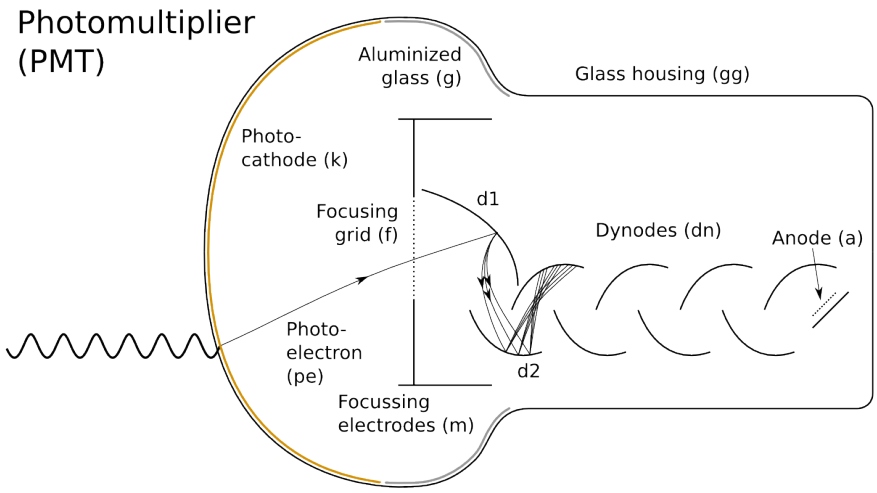


Figure 2.1: Photomultiplier (PMT) layout (cross section): An incident photon is converted in the photocathode (k) to a photoelectron (pe), which is amplified to a measurable signal by multiple stages of secondary emission through acceleration by electric fields onto dynodes (dn, numbered in order of progression), before the electron cascade is collected at the anode (a). Not all series include a focusing grid (f) in the aperture to d1 or aluminized glass (g) on the rear of the glass bulb. The circular platform part of the dynode mount (m) facing the cathode is called deck. Cascade drawn up to d3, dynode structure resembles a linear-focused design (see fig. 2.15).

In order to provide a basis for discussion for the subsequent measurements and studies, in the following the properties which are most relevant for large liquid scintillator detectors will be reviewed and analyzed by means of a novel theory describing interactions in the PMT which is introduced in 2.3.5. These characteristics can be divided into the following groups: sensitivity (section 2.1), photon counting (sec. 2.2), time response (sec. 2.3), pulse shape (sec. 2.4), intrinsic background (generated by the PMT inside itself, sec. 2.5), exterior background (produced by the PMT in itself, other sensors, or a detector, sec. 2.6), and environmental factors (sec. 2.7).
For many properties not only the mean value, but also the uniformity across the sensitive area is of importance and will be discussed where appropriate.

2.1 Sensitivity

2.1.1 Photodetection efficiency

The photodetection efficiency (PDE)[3] of a PMT denotes the probability that a single photon (spe) hitting the sensitive area of the device produces a detectable output pulse. It depends on several properties associated with the processes of

[3] Also called detective/effective quantum efficiency or anode detection efficiency.

photon-to-pulse conversion: The quantum efficiency (QE) describes the likelihood that the incident photon causes emission of a photoelectron from the cathode.
The collection efficiency (CE) is the probability that a photoelectron is successfully transported into the electron multiplier, meaning that it hits the first dynode.
And the multiplication efficiency (ME) [4] denotes the fraction of photoelectrons hitting d1, which are amplified into detectable pulses [5].
The PDE is related to these factors through [6]

$$\mathrm{PDE} = \mathrm{QE} \cdot \mathrm{CE} \cdot \mathrm{ME}. \tag{2.1}$$

Since in (2.1) the right-hand components depend on wavelength, position on the cathode, incident angle, polarization and voltage, so does the PDE.
It *also* depends on the electronics threshold, since a lower threshold allows to detect more underamplified pulses. It is therefore important to properly eliminate electronic noise in the setup for optimal PMT performance, and to specify the threshold used in the measurement. This is usually quoted as fraction of the mean or peak of the single photon pulse charge (or height), using the unit pe. Typical values lie at or below the valley in the spe charge distribution, in the range of 0.25 pe [190], sometimes as low as 0.1 pe [185].
As will be shown in 2.1.3 (based on 2.3.5) eq. (2.1) is only an approximation and the terms QE, CE and ME have to be redefined.

2.1.2 Quantum efficiency [7]

The quantum efficiency of a structure is simply defined as the ratio of emitted photoelectrons $\mathrm{N_{pe}}$ to incident photons N_γ [8]:

$$\mathrm{QE} = \frac{\mathrm{N_{pe}}}{\mathrm{N}_\gamma} \tag{2.2}$$

On the particle level, QE is the probability of photoelectron creation in the cathode and subsequent emission from it, as detailed in the following. It depends on wavelength and the composition, materials and thickness of the photocathode and is always less than unity due to the necessary trade-off between photon absorption and electron emission (see eq. (2.4) and p. 49).

Manufacturers often list the cathode radiant sensitivity S_k [mA/W], which is the ratio of photocathode current to incoming light power, in lieu of the QE [194; p. 299]. The QE can be calculated from S_k if the photon energy (i.e. wavelength λ) is known through [169, 194, 195]:

$$\mathrm{QE} = \frac{h\,c}{e} \cdot \frac{1}{\lambda} \cdot S_k \ [1] \approx \frac{1240}{\lambda[\mathrm{nm}]} \cdot S_k \cdot 100 \ [\%]\,. \tag{2.3}$$

[4] The ME is related to the backscattering losses at d1 by ME = 1 − backscattering losses.
[5] Many authors subsume CE and ME as CE.
[6] The definition depends on the device, for cell-based sensors such as the silicon photomultiplier (SiPM) in (2.1) the right side has to be multiplied with the fill factor (FF) as well.
[7] Content based on [153, 172, 184, 191].
[8] This definition also accounts for the emission of more than one pe per photon, as is the case for double (or even triple) photoelectron emission in vacuum ultraviolet (VUV) PMTs [192, 193].

The cathode luminous sensitivity S is defined as the photocurrent per lumen of incident light flux [µA/lm]. Since the lumen is defined physiologically over the response of the human eye, it is not very useful for the purposes of this work since neither maximum nor mean QE can be derived from S, as it is obtained from the integrated weighted response curve and easily influenced by tails in the QE spectrum [170; p. 32] [195; p. 178]. A bialkali cathode with 30% peak QE can have a low $S = 50\,\mu A/lm$, while for an S20 cathode with only 15% S can be as high as $250\,\mu A/lm$ due to its tail to long wavelengths.

Photoemission process

The process of photon-to-electron conversion is based on the external photoelectric effect, takes place in the semiconductor photocathode evaporated on the inside of the transparent glass window, and can be divided into three steps: photon absorption, photoelectron diffusion and photoelectron emission (compare fig. 2.3) [153, 172, 184, 191].

- **Photon absorption:** After the photon passes the window, it is absorbed in the photocathode by photoelectric effect, transferring its energy to an electron from the valence band (VB) which is excited to the conduction band (CB) and leaves a hole in the VB.

- **Diffusion:** The photoelectron then diffuses to the inner surface of the photocathode during which it loses energy through scattering on the lattice (phonons) and on free electrons and holes (far smaller contribution than phonons since at room temperature only few free carriers are in the CB).

- **Emission:** When the pe reaches the surface with enough energy left to overcome the surface barrier, it escapes into vacuum where it is accelerated into the electron multiplier and amplified. The emission of *free* photoelectrons *from* a solid is called external photoelectric effect and used in all photosensors based on photocathodes (e.g. PMT, micro-channel plate PMT (MCP-PMT), all hybrid sensors) – that is all sensors using vacuum tubes and the majority of gas filled sensors.

The first step, the emission of a photoelectron *into* a solid (*excited* into its CB), is called internal photoelectric effect and sensors based on it (all diode sensors, such as photodiode (PD), avalanche photodiode (APD), SiPM and charge coupled device (CCD)) either collect and possibly amplify the excited charge carriers produced at this point (photovoltaic effect), or use the change in conductivity caused by the excess of carriers (photoconductive effect).
In gases the photoelectric effect ionizes gas atoms or molecules, directly producing free photoelectrons, and is called photoionization. Sensors using this effect also collect the emitted electrons.
For liquids, the process lies in between external photoeffect and photoionization: Photoelectrons are produces by ionization, but with a *reduced* ionization potential compared to the gas due to a thin CB and VB already being formed, and transport and extraction processes also must be considered [196; p. 2].

Limiting factors

In place of eq. (2.2), the QE can also be expressed in view of the processes required for – or impeding – successful pe emission (eq. based on [172, 184, 197]) [9]:

$$\begin{aligned} \mathrm{QE} \approx (1 - R_{ew}) \cdot \exp\left(-\mu_w d_w\right) \cdot (1 - R_{wk}) \\ \cdot \left(1 - \exp\left(\frac{-d_k}{2 l_p \cos\alpha}\right)\right) \cdot \exp\left(\frac{-d_k}{2 l_e \cos\beta}\right) \cdot p_e \end{aligned} \tag{2.4}$$

The terms describe reflection losses at the window and photocathode (first and third), absorption losses in the window assuming a Beer-Lambert law (second), photon absorption leading to pe production (fourth), pe transport losses to the surface (fifth), and the probability to overcome the surface barriers with the remaining energy (last). R_{ew} and R_{wk} are the reflection coefficients at the interfaces environment–window and window–cathode, μ_w is the attenuation coefficient of the window, d_k and d_w are the cathode and window thickness, l_p and l_e the photon absorption length resp. the electron scattering length, α and β are the photon incidence angle and the emission angle of the pe.

Optical properties [10]

Glass window

The glass window of typical PMTs has a refractive index of $n_w \approx 1.49$ and a thickness of 1–10 mm [170; p. 62], around 2–3 mm for large hemispherical PMTs [170; pp. 59, 279]. The Hamamatsu [11] 8" R1408 had an average thickness of ≈2 mm(Schott 8246 glass), the 20" Hamamatsu R3600 of ≈4 mm [185; p. 3-103]). For small PMTs the window often is thicker towards larger radii, for large PMTs which are designed to have high compressive strengths it is thicker in general.

Photocathode

The photocathode is evaporated onto the inside of the window with a thickness d_k in the region of 20 nm, and a refractive index of $n+ik \approx 2.7+1.5i$ at 442 nm [170,199,200]. In [199] also a calculated wavelength dependence of n and k can be found).
The imaginary component k is responsible for the absorption in the photocathode and therefore related to the QE. Due to the small thickness, which is a trade-off between absorption and electron emission, the cathode is semitransparent and a large fraction of photons pass the cathode without being converted to photoelectrons – about 23% at 442 nm in air [200]. This explains why one can see the dynode structure and also the orange-brownish color of bialkali cathodes for which absorption is stronger for shorter wavelengths.
Absorption as a function of distance follows an exponential law as in (2.4) only for thick layers. For real photocathodes the behavior is more complicated and depends on the internal reflections inside the cathode layer and interference effects which enhance absorption, especially close to the emissive surface, and make thicknesses of a few tens of nm enough to absorb the majority of incident photons [172].

[9] This formula is only a rough approximation of a highly complex process and makes simplifications for all processes involved. It is intended to merely convey an idea of the effects. A more exact treatment can be found e.g. in [198].
[10] Content based on [170, 199, 200].
[11] Hamamatsu Photonics K. K.

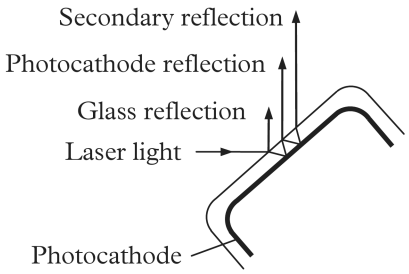


Figure 2.2: Reflections on window and cathode, from [170; p. 44]

Reflection losses and gains

Light entering the glass window (w) can incur Fresnel reflection losses ("Glass reflection" in fig. 2.2) depending on the outer material coupled to the window, the incident angle, and polarization. In air for normal incidence this amounts to about 4%. Linear alkylbenzene (LAB), however, has roughly the same refractive index as PMT glass, so almost no losses at this interface are expected in LENA (Low Energy Neutrino Astronomy).
At the k–w interface inevitably reflection losses (8.3% for normal incidence) occur due to the difference in refractive index (see fig. 2.2). At the cathode surface to the internal vacuum already for incidence angles above 42.2° in LAB total internal reflection occurs, which allows a second pass through the cathode – possibly more if for large angles total internal reflection also happens at the k–w interface – which increases the effective QE.
The exact behavior of reflections and absorption probability in the cathode depends on wavelength, polarization and incidence angle in a rather complicated manner. A summary is given in [170], which takes into account the findings of [199, 200].

Absorption in the window

Absorption can also occur in the glass window, reducing the achievable QE. Window thickness affects the transmission, especially in the UV range [170]. Since the band gap of all eligible window materials has finite values, this leads to a cut-off at low wavelengths.
For the usually employed borosilicate glass[12] transparency drops to 50% around 300 nm (full transmission above 350 nm) [170, 189, 201], see fig. 2.5c.
The photocathode itself is still sensitive down to below 120 nm and special window materials can extend the sensitive spectrum (with UV glass to ≈ 200 nm, fused silica 160 nm, MgF_2 115 nm, LiF 105 nm, see fig. 2.5b) [170, 171]; however, this increases the price significantly. Furthermore, MgF_2 and LiF require careful storage and handling due to the mismatch of the expansion coefficient between window and envelope, or the window might "pop off". [170] LiF also is hygroscopic and can suffer from color centers. Thus, this material is not listed in manufacturer catalogs anymore.

[12] Borosilicate glass contains about 70% SiO_2, 20% B_2O_3 and various metal oxides, where the exact chemical composition varies between glass suppliers and a PMT manufacturer can use several companies [185; pp. 3-104–3-105]. The annealing temperature and cooling process during PMT production also affect glass properties.

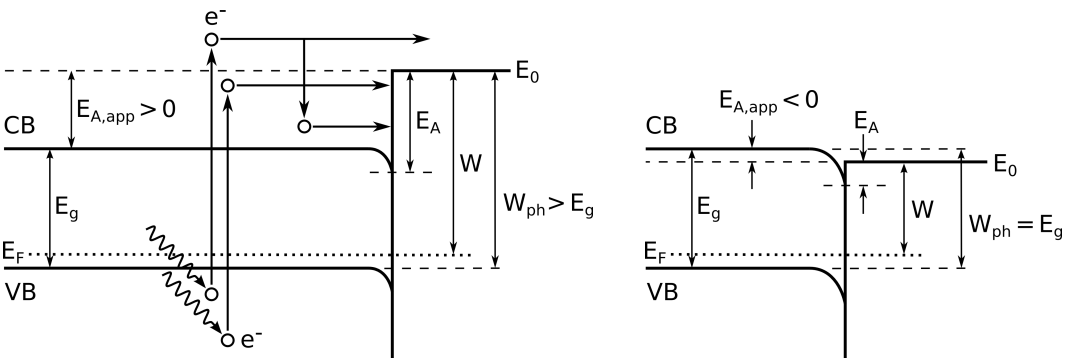


Figure 2.3: Band structure of alkali photocathodes. Terms and discussion see text. Left: Positive electron affinity and photoemission process. Right: Negative electron affinity.

Low background glasses

The usually employed borosilicate glass contains radioactive impurities of ^{40}K, ^{232}Th and ^{238}U which constitute a main background source for sensitive experiments such as neutrino detectors and dark matter search. In these cases special low-background glasses (e.g. Schott 8246 (Borexino, SNO [13]) [80, 162], ETE [14] B53 [170; p. 61]) are used for the PMT front windows and often, but not always [189], for the stem as well (see also 2.6.2). These glasses are potassium-free (naturally occurring potassium contains 0.0118% ^{40}K), which makes them more difficult to work and can cause optic anisotropies such as striations, cords and bubbles [170; pp. 61, 274]. These do not impact PMT performance seriously, but can be an issue regarding compressive strength.

Cathode material [15]

Band structure

Photoelectrons raised across the band gap E_g to the CB must overcome a potential barrier to the vacuum level E_0 before emission from the cathode (fig. 2.3). Several material properties play a role in this and excitation can occur by photon absorption, impact of electrons or ions, ionizing radiation, or thermally.
The electron affinity E_A denotes the potential difference between the lowest level on the CB and the vacuum level at the surface.
In contrast the work function $W = E_0 - E_F$ is the energy difference between vacuum level and Fermi level E_F, which depends on doping.
Finally, the photoemission threshold W_{ph} is the minimum required photon energy for photoelectron emission.

Semiconductors

For semiconductors, E_F lies in the band gap, at room temperature the VB is almost full and the CB almost empty, so no notable amount of photoelectrons can be excited from the CB. Therefore, the energy required for pe emission is the sum of E_g and E_A, barring surface effects, and a material with small values for both should be chosen

[13] Sudbury Neutrino Observatory
[14] Electron Tubes Enterprises Ltd.
[15] Content based on [153, 170, 172, 184, 202, 203]

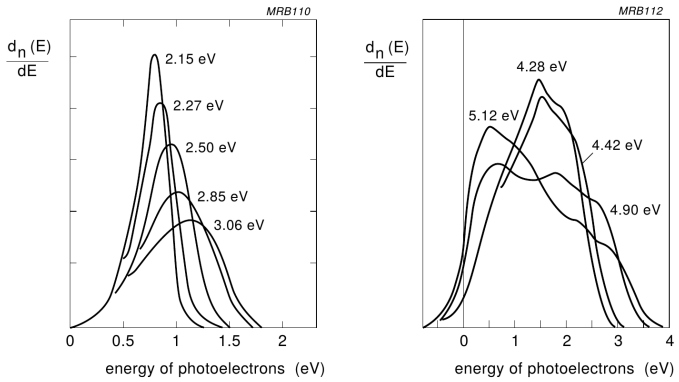


Figure 2.4: Photoelectron emission energies for K_2CsSb at 290 K and varying photon energies, from [172; p. 1-32].

to enable sensitivity in the visible range. The threshold can be further lowered by heavy p-doping with appropriate surface treatment: The adsorption of alkali metals such as Cs produces a bipolar layer due to ionic bonds which reduces the work function, and electron transfer of the adsorbed atoms to the semiconductor bends the bands to lower energies (fig. 2.3). So photoelectrons created further inside than the curvature region only have to overcome an apparent electron affinity $E_{\mathrm{A,app}} < E_{\mathrm{A}}$ for emission, ignoring energy losses through scattering, and then $W_{\mathrm{ph}} = E_{\mathrm{g}} + E_{\mathrm{A,app}} > W$. With Cs adsorption W can be lowered to ≈1.4 eV, with co-adsorption of oxygen even to < 1 eV [16]. W_{ph} imposes a long wavelength cut-off, typically lying in the red or near-infrared (NIR) spectrum [153]. Apart from that, W_{ph} should be minimized even for photons with higher energies to maximize the fraction of escaping photoelectrons.

Energy losses and emission energy

Photoelectrons lose energy on the way to the surface due to multiple inelastic collisions, which results in rather complicated energy and angular distributions of emitted photoelectrons, which depend on crystalline structure and initial photon energy. [172] published data on energy distributions (fig. 2.4) showing one peak shifting with $h\nu$ for photon energies below ≈3 eV leading to pe energies of several tenths of eV, and two broad peaks for larger photon energies, one for slow photoelectrons with energy independent of $h\nu$ and one for fast emitted electrons increasing with initial energy.
As a rule of thumb, between 300 and 800 nm the maximum pe energy is about 1 eV less than the photon energy E_{ph} [170; p. 323], where

$$E_{\mathrm{ph}}\,[\mathrm{eV}] = \frac{h\,c}{e\,\lambda} \tag{2.5}$$

For violet light of 400 nm, close to the typical QE peak of PMTs, the emission energy ranges from 0 to about 2 eV.

[16] [202] assumes this is due to formation of cesium oxide at the surface, lowering the surface barrier by formation of dipoles or heterojunctions; additional oxygen damages the cathode, however. [203] suggests that a pe associates with Cs to form a negative Cs surface ion, and is then transferred by a resonance energy process to the surface bonded oxygen where orbital rearrangement might favor electron extraction into vacuum.

According to [170] the very irregular shapes of measured pe energy distributions might be due to the difficulties of the measurements instead of physics, and no information on angular distributions is available for the same reason. Both distributions are important, however, since both pe emission energy and angle affect the spot size on d1 (see p. 170 and 179), which influences timing and CE. Due to lack of knowledge, manufacturers often assume launch energies of ≈1 eV and a Lambertian angular distribution for simulations.

Metals

Besides semiconductors, metals also can be employed as photocathodes. However, this yields several important drawbacks [196; p. 5]:

- **High reflectivity:** Metallic layers are highly reflective in the visible spectrum (380–740 nm [204; p. 92]) resulting in high reflection losses. This is less severe for semiconductors.
- **Large work function:** For metals E_F lies in the CB so W_{ph} is merely W, but for most metals $W > 3$ eV and only in alkali metals it is low enough to allow photoemission from visible light [172].
- **Small escape depth:** In addition, since the CB is partially filled, there are many free electrons on which scattering occurs, so excited photoelectrons quickly reach thermal equilibrium and only electrons excited near the surface can overcome the vacuum barrier and escape. The escape depth – the depth in the material at which photoelectrons can be created and still reach the surface with sufficient energy to overcome the potential barrier – amounts to only a few atomic radii.

For semiconductors, the probability of collisions with free electrons is very low, and energy losses are mainly from phonon scattering and very small (≈0.05 eV on average [172]). Therefore excited photoelectrons can travel long distances before reaching thermal equilibrium (mean free path between collisions 2.5–5 nm [172]), yielding escape depths of some tens of nm [153, 172]. Thermalization through phonon interactions occurs within about 1 ps [153], which limits the escape depth. After thermalization, the photoelectrons remain for about 100 ps [153] at the bottom of the CB before recombination with a hole, during which they can diffuse up to a few µm [172]. This is important for negative electron affinity (NEA) materials (see p. 65) since in these electrons at the bottom of the CB can escape as well, which greatly increases the escape depth but also causes a non-negligible contribution to the transit time spread (TTS) in the order of 100–150 ps [203, 205].
For these reasons, semiconductors usually are preferred as photocathodes for light sensors.

It should be mentioned that metals can reach notable QE values of 1–10% in the VUV range above 12 eV (Ni, Cu, Pt, Au, W, Mo, Ag, Pd, Al) [206], since in the UV their reflectivity is lower and the pe energy is higher, so electrons can overcome the vacuum barrier even after incurring losses through scattering (mainly on electrons), which increases the escape depth [196; p. 5].

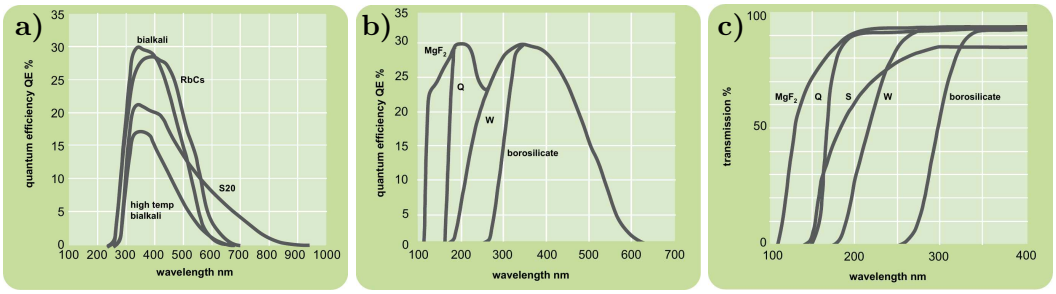


Figure 2.5: Quantum efficiency (QE) spectrum and window transparency:
a) QE spectrum (spectral response) for common alkali-based cathode materials deposited on borosilicate windows.
b) Bialkali QE spectrum using different window materials.
c) UV transmission curves of window materials.
Discussion see text. Abbreviations: W = UV glass, Q = quartz (fused silica), S = sapphire (Al_2O_3). All figures from [171].

Materials

Many different photocathode materials have been developed throughout the years. Since a low work function is required for sensitivity in the visible/near-UV spectrum and increases the escape probabilities at wavelengths below the cut-off, usually alkali metals (especially Cs) are employed because they have the smallest work functions of all elements [207; p. 12-123]. The QE spectrum for some important cathode materials is shown is fig. 2.5a which can be compared to the transmission curves of different window materials in fig. 2.5c. The combination of a bialkali cathode with varying window types results in the effective QE spectra shown in fig. 2.5b.

The most relevant material for neutrino detectors is **bialkali** (K_2CsSb, K bialkali), since its spectral response matches common scintillators (possibly with wavelength-shifters) and Cherenkov emission in water/ice and it has a higher sensitivity resp. a lower dark current than older materials (S1 or monoalkali). E_g was measured to be 1.2 eV [202, 208] to 1.0 eV [209](optical absorption). E_A lies between 1.1 eV [210; ch. 7] and 0.7 eV [202]. These findings agree well with the observed high-wavelength cut-off at 640–690 nm [189, 201] when taking into account band bending at the surface and low concentrations of impurity states possibly extending the sensitivity beyond the normal cutoff [153; p. 267]. The density of states at the lower end of the CB is very low [211], which likely contributes to the low-QE tail at long wavelengths (see fig. 2.5a). The peak QE lies in the region of 340–420 nm [189, 201, 212, 213] with peak values of modern high-QE (HQE) PMTs reaching 30–35% [same sources] on average, occasionally up to 45% [170, 175, 176].

Other alkali-based materials are rubidium bialkali, multialkali, high temperature bialkali, and monoalkali [210; ch. 7] [170, 202]. Further notable materials comprise AgOCs and the III-V semiconductors GaAs and GaAsP. Solar blind materials such as Cs_2Te and CsI have no sensitivity in the wavelength range relevant here and are excluded from the start. The properties of the other materials are summarized in the following list.

- **Rb bialkali** (Rb_2CsSb, green enhanced / extended green bialkali) has a QE spectrum shifted towards higher wavelengths. At 420 nm the QE is comparable or slightly better than for K bialkali with standard quantum efficiency (SQE) [170]. However, the dark count rate (DCR) is four times higher.

- **Multialkali** (Na_2KSb:Cs, S20, trialkali) has a broad sensitivity spectrum extending up to 870 nm, but the DCR is 40 times higher than for K bialkali, and the cathodes are more expensive [170].

- **High-temperature bialkali** (Na_2KSb, low noise bialkali) cathodes are usable at elevated temperatures up to 200 °C at the cost of strongly reduced performance and lifetime. The QE spectrum is similar to K bialkali but the peak QE is 30% lower [170, 184]. When used at room temperature they have very low DCR.

- **Monoalkali** (Cs_3Sb, S11) shows higher dark current, lower QE and a lower resistance than bialkali [184, 202]. Today it is chiefly used for reflection type cathodes and *dynodes*, which then also are photosensitive and can produce photoelectrons which start pulses. This introduces disruptive effects regarding timing and early afterpulses (EAPs) (see 2.3.5, 2.3.6 and 2.5.2.2). The QE(λ) of Cs_3Sb is given in [169; pp. 34, 36] [17]; the peak QE is 20% at 280 nm [18]. At 400 nm the QE amounts to 14.6%, which is smaller than for bialkali cathodes (25–35%), but still sizeable. For high-gain dynodes such as employed at d1, better crystallinity is expected, which should raise the QE even further.

- **AgO:Cs** (S1 [19]) is known for its instability and extreme levels of DCR, which lies over three orders of magnitude higher than the already high S20 [170]. Today it is mainly used in the infrared with cooling [169, 215].

- **III-V semiconductors:** GaAs and GaAsP have higher costs and need greater care in use and storage than bialkali [170]. It should be noted that GaAsP and GaAs have much slower pe emission times, which can introduce a TTS in the order of 100–150 ps [205], although this would not be problematic for the applications discussed here.

  - **GaAs:Cs** (activated with Cs) has a spectral response ranging from <160 nm (reflective mode, nearly flat curve between 300 and 850 nm) or ≈ 320 nm (transmission mode; spectrum in fig. 2.8b) to 920 nm, the QE at 420 nm in both cases is lower than for bialkali [170, 184, 189]. Since it is an NEA material [210; ch. 7](see p. 65), the DCR reaches very high values of $\approx 10^4\,\mathrm{Hz/cm^2}$. Due to several damage mechanisms, mainly ion back bombardment, the QE degrades faster from aging and exposure to high light intensities than for bialkali [216].

  - **GaAsP:Cs** is sensitive from 280 to 720 nm, and shows exceptionally high QE in the visible range (fig. 2.8b): 40% from 480 to 530 nm are quoted by Hamamatsu [189], and ≈ 55% from 460 to 600 nm with ≈ 45% at 420 nm have been reported for the MAGIC-II [20] 1" hybrid avalanche photodiode (Hamamatsu R9792U-40) [217], which in addition used wavelength-shifting coating to enhance the QE in the range of 300–400 nm.

[17] Curve 350U; unfortunately the graph is optimized for readability of the cathode radiant sensitivity, but this can be converted to QE values using (2.3).

[18] This is in agreement with [214], who measured the pe emission energy distribution for Cs_3Sb for 4.45 eV photon energy (corresponding to 280 nm). If one integrates over fig. 10 of the paper, a total QE of about 20% is found.

[19] This was the first compound photocathode material.

[20] Major Atmospheric Gamma Imaging Cherenkov Telescopes

There are several drawbacks, however. As for GaAs, accelerated QE degradation is observed [184] and MAGIC found a QE drop of 10–20% after about 10 a at 400 MHz pe rate (accelerated aging) [218]. In addition, with $5 \cdot 10^5\,\mathrm{Hz/cm^2}$ the dark count is astronomically high [170]. Finally, the material is expensive and not available for large cathode areas [219].

- **Al:** Not commonly used as photocathode due to its low QE, aluminum is of relevance here, since frequently the inside of the PMT bulb opposing the cathode is covered with an aluminum coating to optimize electron optics. Photons which are not absorbed in the semitransparent cathode layer can thus produce photoelectrons on the Al coating, which creates pulses with transit times differing from the regular case which can affect the time response (see 2.3.5 and 2.3.6). Measured values of the QE spectrum are difficult to find, but in any case the QE is very small and only has notable values in the UV range [196, 206]. The QE amounts to 2% at 17 eV photon energy [206], but rapidly decreases and at 213 nm only is $8.4 \cdot 10^{-2}$%, at 266 nm $3.2 \cdot 10^{-3}$% and at 355 nm already merely $3.4 \cdot 10^{-5}$% [220]. Since the glass transparency has a sharp cutoff in the UV (fig. 2.5c) and the cathode absorptance rises with falling wavelength [221], the QE of the Al layer in PMTs is expected to be greatly reduced. Nevertheless, the coated area is very large. Furthermore, calculations for a transmissive Al photocathode show QE values of up to 0.2% (0.5%) at 275 nm (300 nm) for a small thickness of 10 nm [222], so photons which hit the PMT from the side – *without* passing through the cathode first – can very well produce photoelectrons in the Al coating. Wright mentions the QE of the Al coating as a problematic effect [170; pp. 45, 74], which shows that it is not negligible. Assuming a borosilicate window, QE values could be as high as 0.1% for an area of about the same size as the cathode in large PMTs.

- **BeO:** The oxidized BeO layer on BeCu *dynodes* also is photosensitive below about 300 nm with a peak QE of 17.3% at 69 nm [184; pp. 223–224] and thus can cause unwanted photoemission like the common dynode material Cs_3Sb and the Al coating.

Taking into account the strict requirements on DCR, QE and lifetime, bialkali remains the material of choice.
High-temperature bialkali might be a fallback solution if the DCR levels cannot be managed with K bialkali. However, this comes at a price (literally), since the lower QE would have to be compensated by a larger number of PMTs, likely offsetting the advantage through higher radioactive background.

Oxidation

Apart from the activation of the cathode with a small amount of oxygen, alkali photocathode are sensitive to oxidation [202]. They cannot be exposed to air even shortly (e.g. during production) and gases with which they come into contact may only have extraordinarily small amounts of oxygen [196].

Dark counts

Another matter intrinsically connected to the choice of material is thermionic emission of electrons. At room temperature the electrons in the CB have thermal kinetic

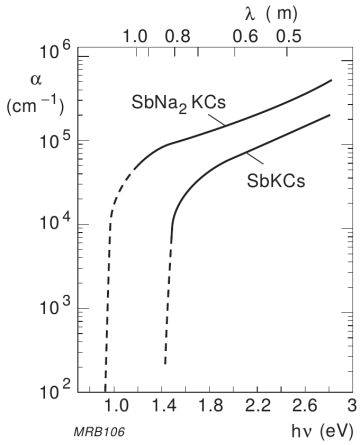


Figure 2.6: Absorption coefficient α vs. photon energy for bialkali and multialkali photocathodes, from [172; p. 1-27].

energies of about 25 meV on average, which is low compared to E_A; however, the electrons at the extreme upper end of the distribution can occasionally have energies exceeding the potential barrier. If this occurs close enough to the surface, the electrons can escape causing pulses in the absence of light, the so-called dark noise (DN) or dark counts.
This thermal emission increases with temperature, obeying Richardson's law [172,184]:

$$J = A\,T^{\gamma} \exp\left(-\frac{W}{k_B T}\right) \tag{2.6}$$

where J is the current density, and A and γ are constants. [184] gives a value of $\gamma = 5/4$, while [172] uses a value of 2.

The thermal emission rate is low in metals ($\approx$0.01 Hz/cm^2 due to the high potential barriers. In semiconductors, however, the (desirably) low work function leads to high rates of up to 10–10^4 Hz/cm^2 [153; pp. 267–268] [172; p. 3-3]. In general, if the photosensor is not cooled, lower potential barriers (yielding higher QE) will always be coupled to higher dark noise rates [203]. This is most extreme for a) NEA materials, where every electron thermally excited into the CB within diffusion length from the surface will contribute to thermionic emission, and b) single-photon sensors based on internal photoeffect, such as SiPMs, since all carriers thermally generated in or near the sensitive area produce dark counts. These sensors typically will be cooled or are used in cases where the signal-to-noise ratio is improved by short time windows or higher numbers of detected photons.

Cathode thickness [21]

Photon absorption in the cathode is a function of distance x from the surface. For thick layers the flux decreases exponentially with distance from surface: $I = I_0 \exp(-\alpha(\lambda)\,x)$, with α the absorption coefficient of the material, which depends on band structure and wavelength (fig. 2.6). For bialkali cathodes α has a steep onset at W_{ph} and increases with photon energy – the absorption is higher for blue light, so the QE is lower at larger wavelengths, since the cathode thickness stays the same.

[21] Content based on [172]

At 420 nm $\alpha \approx 0.03/\text{nm}$, and from the simple exponential law a thickness of over 150 nm would be required for 99% absorption, while for $x = 20\,\text{nm}$ only 45% are absorbed.
For realistic cathode layers the behavior is more complicated due to reflections and interference, [200] measured a total absorptance of around 57% for a 23 nm bialkali cathode with water as external medium for incident angles below 48°, while for larger angles first an increase up to 77% at 68° is observed, which is followed by a drop to 0 at 90°.
In general, the absorptance depends on wavelength, angle and polarization (see e.g. the simulations for a S20 cathode in [221]) and a larger photocathode thickness yields higher absorptance.

On the other hand, the escape depth amounts to only a few tens of nm for bialkali, so in order to maximize the fraction of produced photoelectrons which can escape into vacuum, the cathode should be as thin as possible.

The chosen thickness of the photocathode layer therefore is a trade-off between photon absorption and photoelectron emission, limiting attainable QE. The cathode cannot be made as thick as would be required for total absorption, so it is semitransparent.

Increasing the thickness also shifts the QE peak towards larger wavelengths [173; p. 18]. This is due to two effects: Blue light is preferably absorbed near the input window but the pe has to traverse the complete layer for emission; the QE drops for blue light for a thickness above ≈30 nm for S20 cathodes, which indicates an electron mean free path comparable to this value [203]. On the other hand, a larger thickness allows more photons to be absorbed in the red spectrum, where the absorption coefficient is smaller, which increases the QE for these wavelengths.

While transmissive cathodes generally are not thicker than the escape depth, reflective cathodes can have somewhat greater thicknesses since the photoelectrons are emitted from the same surface the light enters [153].

Cathode production [22]

Besides the choice of material and thickness, the photocathode manufacturing process imposes significant limitations. Basically speaking, the PMT is its own factory.
The production of the cathode is based on simple evaporation techniques of the 1930s – the base material sources are typically permanently installed *inside* the PMT and contained in beads fused to resistive heating wires facing the window (Sb) and metals foils or containers attached to the dynode chain (K, Cs); Cs can also be produced in an external generator. After placing the dynode chain inside the PMT and sealing it on a vacuum pump, the PMT is baked, plasma-cleaned, heated, and then K and Sb are evaporated repeatedly by heating the respective containers through controlled currents, after which a final layer of Cs is applied in the same manner. Following the evaporation, the PMT is vacuum baked and the getters are activated, before sealing off the glass tube.
The advantages of this method are that it is far cheaper, faster and less complicated in manufacturing than e.g. molecular-beam epitaxy (MBE).

[22] Content based on [203, 223, 224]

The problems are manifold, however. Overall, the procedure is less controlled and it is therefore difficult to obtain the optimum recipe [23]. The material deposition occurs in many steps with varying material composition, the source materials are not of semiconductor industry purity, and the vacuum during deposition is comparatively poor. The crystalline phases are therefore unlikely to be homogeneous and small grains tend to form. This causes electron scattering due to a high number of lattice defects and electrons traps. In addition, due to the stepwise deposition many layers with interfaces are assumed to form, which cause additional electron scattering and inhibited pe movement towards the surface by 2D electron gas formation.

Spatial uniformity

Besides the aforementioned problems, the spatial uniformity of the cathode thickness and its material composition varies over the sensitive area, which causes QE changes and limits detector resolution [153; p. 268].
QE uniformity has to be distinguished from PDE uniformity, in which CE variations also play a role, and is less straightforward to measure.
It is difficult to give typical QE uniformity values, since there are not many publications, but the QE fluctuations over the cathode lie in the order of ±10% for modern large PMTs [225–228], and most publications observe a slight drop of the QE towards the cathode brink.

Temperature dependence

The quantum efficiency also shows a temperature dependence, which varies with wavelength. Based on typical use cases three temperature regions can be distinguished.

Near room temperature

Around room temperature (−30 to 60 °C) the QE temperature coefficient α_{QE} for bialkali is negative (i.e. the QE decreases with rising temperature) below ≈ 620 nm, and positive above (fig. 2.7a); at 420 nm $\alpha_{QE} \approx -0.05\%/°C$ [170; p. 66].
The increased QE in the blue for lower temperature is due to less phonon scattering. For the corresponding decrease in the red several effects could contribute [229] [173; p. 38]: an increased band gap, reduced occupation of defect levels lying in the band gap, possibly unfavorable changes in band bending, or for large impurity levels the QE increase with higher temperatures could be due to thermally assisted photoemission. [173] assumes the latter to be the case due to the change of slope in the QE spectrum at high wavelengths when the temperature is raised.

Low temperatures

According to [170] most PMT types can be operated at liquid argon (LAr) cryogenic temperatures (87 K) without modifications, albeit at the cost of reduced performance. The electrical conductivity in the cathode is significantly reduced at low temperatures, which affects the recharging due to space charge buildup and reduces QE [187]. The rising resistivity causes a steep drop in QE below a certain temperature which depends on light flux (fig. 2.7b). To reduce cathode resistivity, a thin Pt substrate can be

[23] Townsend compares it to cooking [203]: "The ingredients are known, but variable, and the processing and final product is highly subjective."

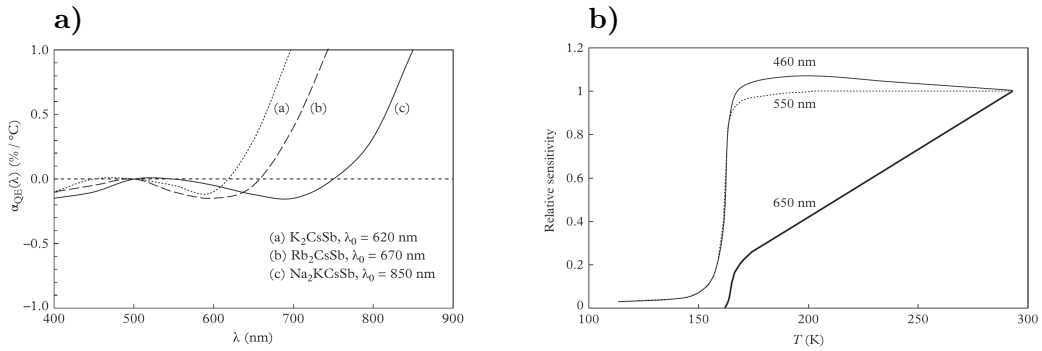


Figure 2.7: Temperature dependence of the quantum efficiency spectrum:
a) QE temperature coefficient α_{QE} vs. wavelength for alkali-based photocathodes for temperatures within $\pm 50\,°C$. λ_0 denotes the cut-off wavelength where the QE has fallen to 1% of its peak value. From [170; p. 66]
b) Relative QE vs. temperature for different wavelengths (EMI 9750B). The temperature where the drop occurs depends on illumination intensity and mode (pulsed or continuous). From [170; p. 68].

used, which extends stable operation down to below 77 K but reduces transmission by a factor of ≈ 0.7.
In addition, W_{ph} changes from about 1.81 eV at 300 K to 1.875 eV at 77 K [230], which reduces QE in the red.
Apart from the possible limit due to resistivity, for blue light (469 nm) the QE increases with temperatures falling from 300 K ($\alpha_{QE} = -0.15\%/K$) down to ≈200 K (fig. 2.7b), and then drops with $+0.2\%/K$ down to 77 K. The QE stays stable for green light (550 nm) at least down to 200 K, and decreases monotonically for red light (650 nm, $+0.6\%/K$). The initial rise in the blue is due to reduced phonon scattering, the decrease below 200 K could also be due to the increased W_{ph}.

High temperatures

Bialkali PMT can also be operated at elevated temperatures up to ≈100 °C but with reduced lifetime [170]. Beyond this temperature most alkali antimonides are unstable and start losing QE, likely due to Cs dissociation from the surface, since Cs has the highest vapor pressure of the alkali metals [202]. This is corroborated by the observation that sensitivity can often be restored by addition of further Cs.
High temperature bialkali photocathodes (Na_2KSb) for this reason do not contain cesium and can be operated up to 200 °C, but with serious loss of performance and reduced lifetime. Up to 100 °C $\alpha_{QE} = -0.15\%/K$ at 420 nm, and at 100 °C the QE is about 85% of its value at room temperature. Above 100 °C the coefficient rapidly increases and at 175 °C the QE is only ≈ 40% of the RT value [231].

Photocathode lifetime [24]

Given appropriate operating and environmental conditions, photocathodes have shown lifetimes of over 30 years. Factors affecting lifetime are:

- Helium permeating through the glass window causes damage by ion bombardment from He afterpulses (see 2.5.2.3), heightens DCR and lowers the breakdown voltage until operation no longer is possible. Therefore, the PMT

[24] Content based on [170, 184]

must be kept away from He gas at all costs even when inactive, [232] even recommends storage in a nitrogen atmosphere. Fused silica glass windows are especially permeable [184].

- Slow degassing of internal metal and ceramic elements poisons the cathode.
- High temperature decreases lifetime due to Cs evaporation from the surfaces.
- Uneven cooling of the PMT can damage the device through migration of Cs to the cathode [170; p. 463]
- A potential gradient between window and cathode (from e.g. applying negative high voltage (HV) and using a grounded metallic holder ring on the PMT) causes electrolytic decomposition of the material [202]: Na ions migrate from the glass into the cathode, poisoning it.
- Long term exposure of an *inactive* PMT to bright light such as the sun or fluorescent lighting will cause noticeable bleaching over the course of several months. The yellow-brown color will disappear, as will the sensitivity.
- Even short exposure of an *active* PMT to room lighting or daylight will lead to PMT failure due to the high photon flux [171]. In some cases recovery is observed with reduced performance.
- Exposure to neutron or gamma radiation will color the glass window and reduce transmittance, especially in the UV and blue spectrum. Synthetic silica is less strongly affected, followed by fused silica, UV glass and borosilicate glass. Storage will recover some transmittance, especially at elevated temperatures.

Improvements

Since quantum efficiency is limited by many factors, there are many possible approaches to increase it. Starting in the early 2000s, efforts to improve the photocathode QE were undertaken by all major manufacturers in collaboration with Mirzoyan and Teshima, which resulted in an increase of peak bialkali QE values from the level of 25–27%, which was already first achieved in the 1960s, to 30–38% [174, 175, 233]. In 2006 Hamamatsu reported even higher peak QE values of 43–45% around 400 nm [234], in 2007 a peak QE of as high as 55% from 370 to 390 nm was measured for a Photonis Super3 bialkali cathode on a 3" PMT [235]. However, in the QE map also shown in [235], such extreme values only occurred close to the PMT brink, which points to possible boosts through double cathode effect (see p. 64) or reflections on structures outside the PMT, which can raise the QE near the edge. To the present author's knowledge such high values have not been reproduced for bialkali cathodes since.
The resulting PMTs with improved QE are, independent of the used enhancement techniques, commonly dubbed high quantum efficiency (HQE) to distinguish them from "standard QE" (SQE) PMTs using older techniques. While at the beginning of the HQE era the production processes were not yet under good control and large diameter HQE PMTs had to be selected from the high-QE tail of production (marked

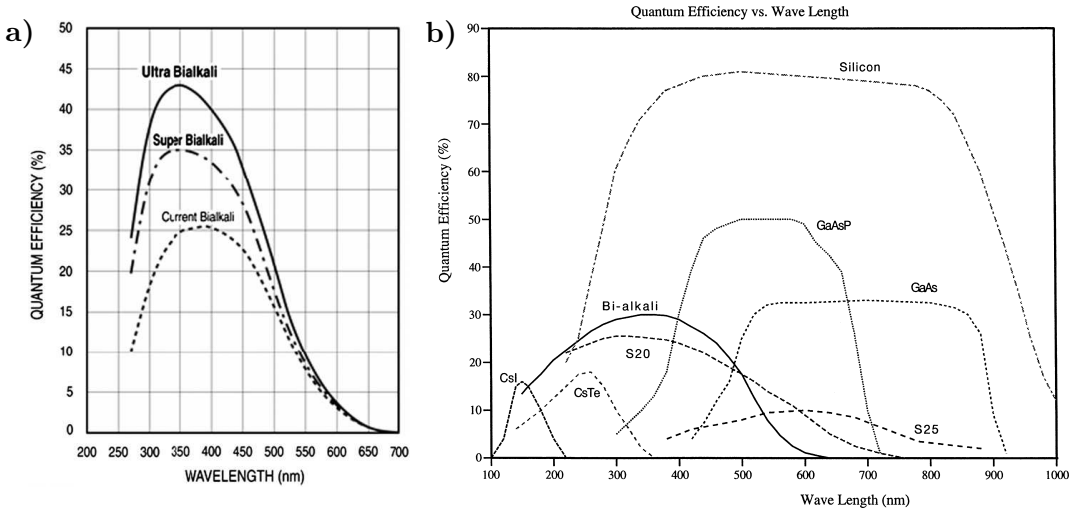


Figure 2.8: Quantum efficiency spectra for different materials:
a) HQE photocathodes (super bialkali (SBA), ultra bialkali (UBA)) compared to normal bialkali, from [176].
b) GaAsP, GaAs, Si (internal photoeffect), CsI, CsTe and others vs. bialkali, from [237].

by the "-SEL" appendix for Hamamatsu PMTs), the process now seems to be established. Meanwhile, Hamamatsu offers Super Bialkali (SBA, series appendix "-100") PMTs with a peak QE of typically 35% for up to 12" diameter (the 20" R12860 reaches 31% [236]) and Ultra Bialkali (UBA, appendix "-200") PMTs with typically 43% peak values for 30 mm square metal package PMTs [213].
In fig. 2.8 the QE spectra of SBA and UBA cathodes can be compared to those of other photocathode materials.

In the following an overview of possible ways to increase QE shall be given.

Window modifications [25]

Sandblasting

Sensitivity can be improved if the outer glass surface is roughened by sand blasting, chemical etching or mechanical abrasion through grinding or sandpapering. One reason for increased QE is from longer paths lengths through the cathode layer caused by refraction on the rough surface yielding a broadened angular distribution (fig. 2.9b). In addition, on the glass–cathode interface about 20% of light is reflected and the diffuse surface scatters up to half of this light back towards the cathode. The size and depth of surface features is in the order of several µm and does not depend much on the particle size used for ablation, hence diffraction plays no role. A QE increase of 20% at 420 nm was reported for an ETE 8" 9354 PMT. A pronounced effect is only obtained in air or vacuum, however; in water or liquid scintillator the similar refractive indices limit the increase.

Matt coating

[174, 206, 238] use a similar approach, by first sandblasting the window and then coating it with a milky layer of matt lacquer (Paraloid), optionally also containing

[25] Content based on [170, 203]

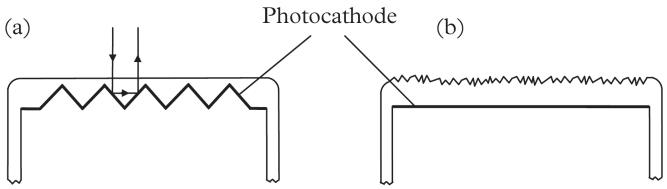


Figure 2.9: Modified PMT entrance windows: a) Prismatic window. b) Sandblasted window. From [170; p. 116].

a wavelength shifter improving the UV response. [238] reports a ≈ 15% relative QE increase (rising towards the red) in a wide spectral range above 330 nm. For Hamamatsu PMTs [206] observed only modest increases of the peak QE to 35% (340–390 nm), possibly due to their low photon back-reflection on the cathode (due to an antireflective (AR) layer, see p. 64), while for ET [26] PMTs strong improvements yielding a peak QE of ≈ 30% at 390 nm were found.

Prismatic window

Another possibility is the use of a prismatic window with four-sided pyramids on the inner glass surface serving as cathode substrate (see also [233]). This increases absorption probability through longer paths in the cathode due to the tilted facet normal and because a fraction of the non-absorbed light hitting a glass–cathode facet is reflected onto a neighboring facet where it can be absorbed again (fig. 2.9a). This structuring, however, only has notable advantages for light coming from the front with a narrow angular spread.

Curved cathodes

As it turns out, curved cathodes, which are used for all large diameter PMTs, are beneficial for sensitivity, since they offer higher absorption for light incident at non-normal angles due to photon-trapping. In addition, the CE is improved.

Antireflective coating on the glass–air surface

AR coatings on the outside of the PMT glass are not frequently used since they are expensive and the reflection loss at the air–glass interface is small (≈ 4% at normal incidence) compared to the glass–cathode interface.

Reuse of transmitted photons

Reflective coating on the bulb

In many tubes a reflective aluminum layer is evaporated onto the inside of the bulb opposing the cathode, which is connected to cathode potential to deflect photoelectrons away from the glass [172]. This aluminized layer is photosensitive and can produce photoelectrons with a small probability [170; p. 74].
The disk-shaped support flange (deck) in front of the dynode stack is also highly reflective.
Both can reflect photons passing the cathode back towards it, allowing a second pass [174, 219]. The delays in detection time due to the extra distance covered by

[26] Electron Tubes Ltd.

the photon is merely a fraction of a ns, but yields a non-negligible contribution to the trailing flank of the transit time peak (compare 2.3.6), which is why these enhancements are mostly seen as unwelcome [170]. Notable exceptions comprise the CLARITY XP series of Photonis where a special tulip-shaped reflective backside was used to obtain a HQE of $38 \pm 5\%$ [175].

Double cathode effect

Furthermore, transmitted light coming from the side can hit the cathode a second time if the angle and incident position are right [239]. This so-called double cathode effect is especially strong for sensor designs with near-spherical cathodes (X-HPD, SMART/XP2600, QUASAR-370; see p. 651. This improves QE for incidence at the side compared to the top by a factor of 2.1 for the X-HPD, which is practically wavelength-independent [240]. Comparative measurements done for the likewise spherical XP2600 revealed a maximum increase from pole to equator of about 44.5% at 405 nm, and for the mushroom-shaped Quasar-370 still an improvement by $\approx 22\%$ at 380 nm with a more pronounced wavelength dependence was found [239].
The enhanced sensitivity for non-normal light is beneficial for experiments requiring 4π field of view, e.g. underwater neutrino telescopes, and provides less benefits for sensors focused on mostly frontal detection such as in LENA.

Photocathode modifications

The main contributions to the development of HQE PMTs come from modifications to the cathode. Key factors are the reduction of optical losses, electron scattering and the potential barrier to the vacuum.

- **AR layer:** Using an anti-reflective layer (e.g. HfO_2 or Y_2O_3 [241]) to couple the cathode to the window reduces losses due to internal reflection [206, 242]. These range from $\approx 10\%$ (blue) to $\approx 30\%$ (red) [170, 200, 206]. Optimal tuning of composition and growth will further reduce remaining losses [219].

- **Thickness:** Tuning the cathode thickness optimizes the trade-off between photon absorption and pe emission [242]. The thickness also determines the peak position in the QE spectrum due to enhancement by interference [206]. For HQE PMTs often a shift of the peak position to the blue is observed (compare spectra of bialkali, SBA and UBA in fig. 2.8a).

- **Composition:** The exact composition of bialkali cathodes varies between manufacturers and optimizing it can yield lowered work functions [175] and improve spectral response [206].

- **Purified materials:** Using highly purified basic materials for photocathode production will reduce scattering and increase escape depth [242]. [206] considers it a dominant effect for QE improvement and suggests a purity of 6N or higher[27], [219] argues that according to some experts 3N might be sufficient.

[27] The nines notation denotes the number of nines after the decimal point of a quantity x which is very close to 1, using the abbreviation N in place of a unit [243; p. 387]. More formally, this can be defined as the logarithmic distance from 1 using the formula $-\log_{10}(1-x)$, which also allows to determine intermediate values. For example $0.999999 = 6\text{N}$ and $0.9997 = 3.5\text{N}$; the unabbreviated quantities would be very cumbersome to write and compare.

- **Growth:** A better understanding and control of the formation process allows to minimize growth defects and improve the crystallinity of the film, which will increase the escape length [175, 176, 242, 244]. [176] showed that a QE of 40% at 400 nm is related to the presence of a well-defined Sb crystal peak on X-ray diffraction spectra of the initial Sb layer [210; ch. 7]. Uniformity and band bending are also key growth parameters affecting QE [242].

- **Surface:** The surface structure and cleanliness have an impact on cathode growth and impurity diffusion [242].

- **Underlayer:** An additional La_2O_3 underlayer between the AR layer and the cathode suppresses alkali dispersion towards the substrate during heat treatment of the cathode, which would reduce the QE and could reflect photoelectrons diffusing towards the window, and yields a dramatic QE increase [241].

- **Negative electron affinity:** If a surface layer of electropositive materials such as Cs_2O is applied, the band structure is bent down, which can raise the minimum CB level *above* the vacuum level for certain doping and band gaps, causing an apparent negative electron affinity (NEA, right part of fig. 2.3) [172, 184] [210; ch. 7]. In this case even electrons which have reached thermal equilibrium through phonon scattering with a remaining energy of only a few $k_B T$ above the CB bottom can be emitted and the escape length becomes equal to the diffusion length of several µm. The depletion layer also decreases scattering. In addition, W_{ph} now is equal to E_g, which extends the sensitivity to longer wavelengths with a still high QE close to the threshold.
  Examples for NEA materials include III-V semiconductors such as GaAs, GaAsP and GaN, Si and multialkali [203]. In case of the latter, NEA is achieved through a thick cesiated K_2CsSb layer at the surface.
  As mentioned, the drawbacks are drastically increased DCRs and a small increase in TTS.

Although it is not publicly known for all of the above approaches if they are used by the manufacturers in the production of HQE photocathodes, some educated guesses can be made, for HQE PMTs have certain drawbacks, namely higher DCR and increased afterpulse (AP) probabilities [190, 219, 242, 245, 246], which points to a lowered work function and/or electron affinity.
HQE PMTs costs are not much higher than for SQE, since the employed technologies are similar to the previously existing [206], which makes them a cost-effective solution with regard to detected photon yield (dpy) while at the same time saving channels [190].

Reflective photocathode

So far only transmission mode photocathodes were discussed, which derive their name from the photoelectrons leaving the cathode on the side opposing light incidence. In reflective photocathodes, the pe leaves from the same side. This can increase QE and extend sensitivity to lower wavelengths [247, 248].
According to eq. (2.4) more photons are absorbed near the entrance surface; however, the emission probability of a pe decreases exponentially with traversed distance. In transmissive cathodes the photoelectrons thus have to traverse most of the cathode

layer, leading to a mismatch between absorption and survival probability. In reflective cathodes, however, the electrons have short paths and therefore increased emission probabilities, especially for shorter wavelengths - the electron scattering length l_e drops with decreasing wavelength [172]. If thin photocathodes and interference multilayer systems are used, the QE can be further increased [248]. In the literature peak QE enhancements up to a factor of 2 are expected, in practice an enhancement by $\approx 40\%$ can be reached [247].
According to [203] also the long wavelength performance, which suffers from low absorption coefficients, can be boosted by increasing the thickness. [202; pp. 92–93] reports an increased QE for multialkali cathode at higher wavelengths, and a $\approx 30\%$ higher peak QE if a reflective substrate is used, which effectively doubles the cathode thickness. The peak is shifted to higher wavelengths, while below 430 nm the QE remained unchanged or even dropped. This is because blue light does not profit from the reflection on the backside since it is already absorbed near the surface.
Another advantage of reflective photocathodes can be lower construction costs than transmission cathodes [203].

Practical use of reflective cathodes for large sensors is limited, however.

The ReFerence (see p. 679) photosensor prototypes incorporated reflective cathodes; however, they faced problems overcoming the large reflectance losses towards the vacuum side [249].
The JUNO [28] MCP-PMT (NNVT [29] GDB-6201, see p. 643) uses a reflective cathode at the backside of the bulb for photon recycling of transmitted light [250]. Combined with a CE of 98% this allows to obtain a high PDE of over 30% [251]. The assumed QE of the reflective cathode lies between 30 and 50%; however, no measured values could be found in literature. The time delay from the additional photon path to the rear cathode will increase the TTS, but this effect is negligible compared to the marked transit time differences resulting from the electron optics and backscattering reported e.g. in [252].
For PMTs reflective cathodes are only used for the side-window type, with sizes only up to about 1", peak QE values of not higher than 30% [189] – possibly due to reflectance losses – and extended sensitivity in the NIR region [169].
Another problem is that the amplification structure either blocks the way for incident photons (ReFerence, GDB-6201) or has to be placed sidewards (side-on PMT) with the consequence of large CE non-uniformities [189].

As a side note, dynodes are always used in reflective mode, but transmission dynodes, so-called tynodes, are being developed as well [253] [30].

Schottky effect [31]

An accelerating external electric field near the cathode surface can have a measurable impact on the photoemission efficiency due to the Schottky effect. This reduces the

[28] Jiangmen Underground Neutrino Observatory
[29] North Night Vision Technology Co., Ltd.
[30] Transmission dynodes are not a new concept and were studied already in the 1960s [170], though with much larger thicknesses for mechanical robustness, which negatively affected the electron yield.
[31] Content based on [172]

work function and therefore the potential barrier by $\sqrt{\frac{e^3 E}{4\pi\epsilon_0}}$, with ϵ_0 the vacuum permittivity and E the surface electric field strength [254], The emission threshold is shifted to longer wavelengths and the QE is improved mainly close to the threshold. Appreciable effects are, however, only attained at field strengths larger than several kV/cm, which can only be reached for proximity focused photosensors [32].

Conclusions

Summarizing, for the detectors discussed here, HQE bialkali photocathodes with ultra-low background borosilicate glass windows are the best choice. For LENA this applies under the premise that the DCR requirements can be fulfilled, otherwise SQE PMTS will have to be used.

2.1.3 Collection efficiency and multiplication efficiency [33]

Definitions

Similar to the process of photoelectron emission relevant for QE, which is divided into the three stages of photoelectron generation, transport to the surface and emission, the process of output pulse generation which is linked to PDE, can be described in three steps: Photon conversion, transport into the electron multiplier, and amplification. The first simply corresponds to the pe emission from the cathode, and the limits imposed on attainable QE values were discussed at length. The transport and amplification steps, however, also have losses which affect energy resolution in a detector.
The collection efficiency (CE) can be defined as the ratio of photoelectrons reaching d1 to photoelectrons leaving the cathode [34], the multiplication efficiency (ME) as the fraction of photoelectrons hitting d1 which produce an anode output pulse [35]. In the literature the term 'collection efficiency' most often comprises the effects of both CE and ME; this extended collection efficiency will be called pulse efficiency (PE) here, where $\mathrm{PE} = \mathrm{CE} \cdot \mathrm{ME}$. Pulse efficiency is then defined as the ratio of output anode pulses to emitted photoelectrons (resp. produced primary charge carriers) [169, 170] [36]

The results from time shift theory presented in 2.3.5, 2.3.6 and 2.5.2.2, however, show that these common definitions are not sufficient. When a photoelectron does not

[32] Fast small PMTs can have up to 1 kV/cm extraction field at the cathode (Hamamatsu R4998), with metal channel dynodes up to 1.5 kV/cm (Hamamatsu R7400); for comparison, a proximity-focused MCP-PMT (Photek PMT413LJ) has 12 kV/cm [170; p. 336]
[33] Content based on [169, 170]
[34] General definition: CE is the probability to transfer the primary charge carrier to the amplification stage or readout channel [244].
[35] General definition: ME is the probability that the amplification process produces a detectable signal [244].
[36] Other definitions, such as the fraction of photoelectrons landing on the effective area of d1 to emitted photoelectrons [169] – the effective area is designed for efficient multiplication on later dynodes – can be misleading, since they ignore backscattering losses.

hit d1 but another structure (such as the aluminized glass, deck, focusing grid, light shield or later dynodes) it can produce secondary electrons there, which can be focused into the electron multiplier and trigger a cascade nevertheless. In addition, photons passing the semitransparent cathode can hit a different photosensitive structure (Al, deck, grid, dynodes) and produce a photoelectron, which also can start a pulse.
While these additional pulse-generating effects are detrimental for timing, they raise the PDE above the value from pe production on k followed by collection and amplification on d1 alone. It is thus reasonable to define a QE for each photosensitive structure i (QE_i), and CE and ME values for each stage j which can be hit by the photoelectrons produced on i (CE_{ij}, ME_{ij}). The PDE of a single photosensitive structure (PDE_i) is its QE multiplied by the sum over all PE terms:

$$\mathrm{PDE}_i = \mathrm{QE}_i \cdot \sum_{j=\mathrm{dn,g,m,f,s}} (\mathrm{CE}_{ij} \cdot \mathrm{ME}_{ij})$$

where dn denotes the n-th dynode, g the aluminized glass, m the deck of the dynode mount, f the focusing grid, and s the light shield (nomenclature as in 2.3.5).
The complete PDE then follows from the sum over all PDE_i:

$$\mathrm{PDE} = \sum_{i=\mathrm{k,g,m,f,dn}} \left(\mathrm{QE}_i \cdot \sum_{j=\mathrm{dn,g,m,f,s}} (\mathrm{CE}_{ij} \cdot \mathrm{ME}_{ij}) \right)$$

Even this formula is still not complete. The different time shift effects which can occur on each stage (see 2.3.5) vary the ME. For example the threshold losses (see p. 84) are different for each effect, since the pulse charge varies. In addition, some effects can produce luminescence, which can cause a pulse even if the original one is lost. Therefore the ME_{ij} actually are effective terms, which result from the sum over all possible effects x weighted by their probability p_x:

$$\mathrm{PDE} = \sum_{i=\mathrm{k,g,m,f,dn}} \left(\mathrm{QE}_i \cdot \sum_{j=\mathrm{dn,g,m,f,s}} \left(\mathrm{CE}_{ij} \cdot \frac{\sum_x (p_x \cdot \mathrm{ME}_{ijx})}{\sum_x p_x} \right) \right)$$

In fact, also the CE and ME of amplification stages after the first also have to be taken into account, since with a finite probability all electrons produced on the first can get lost or skip amplification (see pp. 78, 83). These (small) losses, however, can be included in the effective values ME_{ijx}.
Since measurement of these additional QE, CE and ME terms is extremely difficult – even obtaining $\mathrm{QE_k}$ and $\mathrm{PE_{d1}}$ is non-trivial (see p. 86) – and they only yield small corrections compared to the PE from d1, in the following primarily the properties for d1 will be analyzed, and the terms CE, ME and PE without subscript shall refer to d1 and photoelectrons from k. In 2.3.5 and 2.3.6, the QE and probability of effects on the other structures is calculated and verified with measurements, which can be used to estimate the overall PDE.

Transport

Photoelectrons are accelerated onto d1 by a potential difference of typically several hundred volts. To ensure uniform collection from the whole cathode area, focusing

electrodes and grids between cathode and d1 are used (see fig. 2.1 and 2.16). For fast response sometimes accelerators with higher voltage than at d1 are placed in between k and d1 [170]; in this case the pe actually decelerates between accelerator and d1.

Secondary emission [37]

When a primary electron impinges on a dynode with a kinetic energy E of several hundred eV, gained through acceleration in the electric field, it typically knocks out multiple secondary electrons (SEs) from the entrance surface produced by ionization and excitation in the material, the so-called secondary emission. This is the amplification process upon which the electron multiplier is based, and the secondary electrons are then accelerated onto the next dynode where they again produce secondaries.
The process of secondary emission is similar to photoemission and again can be divided into electron generation, diffusion to the surface and escape. For a material to have high secondary emission yield (SEY, δ), it has to fulfill similar requirements as a good photocathode material – consequently many cathode materials are or have been employed as dynodes (e.g. Cs_3Sb, K_2CsSb, K_2CsSb:O, Rb_2CsSb, Na_2KSb:Cs, AgOCs). Metals are poor secondary emission materials, just as they are bad photocathodes. This is due to their low escape depths ($\approx$3 nm) leading to low SEYs (1–2) at 500 eV [172; p. 1-35]. NEA materials on the other hand can have very high SEY for the same reasons as why they have high QE.
In the following the SEY on dynode n is abbreviated as δn, since this is more suitable than δ_n for the notations describing PMT effects which will be introduced in 2.3.5.

Dynode materials in photomultipliers

Typical dynode surface materials are Cs_3Sb, BeO (oxidized CuBe substrate), MgO (oxidized AgMg substrate), Al_2O_3 (oxidized NiAl substrate), GaP (NEA) and GaAsP (NEA) [169, 172, 258–260], placed on a conductive electrode made of Ni, stainless steel or BeCu. Depending on substrate material this can be achieved by oxidation or coating. Typically Cs-activation is performed (not mentioned in the chemical formulas). Other materials comprise the ones mentioned in the last paragraph. Diamond dynodes are also being investigated and have shown excellent properties [258, 260].
According to [261–263], Cs_3Sb is employed as dynode surface material in the Hamamatsu R7081 and R11780 as well as in the ETE 9354 and D784, four short-list photosensors (see 3.1), whereas the ET 9351 used BeCu dynodes [80].

First dynode

On the first dynode, for newer PMTs typically a material with higher SEY is used, as δ1 determines the PMT charge properties to a large extent (see 2.2.2) and also affects timing (see 2.3) [170; p. 157]. Modern PMTs have either only high-gain dynodes or at least a high-gain first dynode, e.g. BeO dynodes with a high-gain K_2CsSb d1. BeO dynodes are beneficial for linearity (however, see p. 152), so combining a high-gain d1 with BeO at later stages provides good charge resolution R and large dynamic

[37] Content based on [169, 170, 172, 173, 184] and [255–257] – the fields of scanning electron microscopy (SEM) and transmission electron microscopy yield fruitful insights.

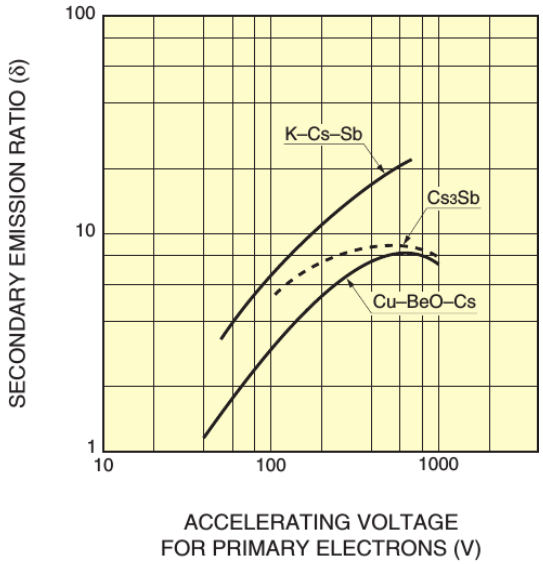


Figure 2.10: Secondary emission yield (SEY) vs. energy for common dynode materials, from [169; p. 19].

range. [171] shows a curve for high-gain Cs_3Sb employed on d1, possibly obtained by better process control or optimized growth parameters, which has an SEY of > 15 for voltages exceeding 370 V and reaches 20 at 600 V.
[172] states that in the charge spectrum a peak only is observed if $\delta 1 > 6$–8, and that the peak-to-valley ratio only exceeds 2 if $\delta 1 > 12$, which allows to determine the charge resolution $R_{F,p}$ defined as FWHM [38] of the peak divided by peak position. Since for all PMT series considered here (see 3.1) an FWHM charge resolution was obtained and the SEY for standard Cs_3Sb in fig. 2.10 does not reach this value [39] this implies that all use a high-gain material for d1. By contrast, older PMT types such as the Hamamatsu R1449 (BeCu) still employ the same material for d1 as for the rest of the dynodes [264].

Secondary emission yield

The secondary emission yield δ depends on the dynode material, the energy of the primary electron, and the incidence angle. It also is dependent on the activation (oxidation or heat) and vacuum, and can change with time.

Energy and material dependence

The dependence on energy E and material is depicted in fig. 2.10 (see also [169, 170, 172, 173, 184]), for PMTs electron energies from ≈10–1000 eV are relevant.
At a certain accelerating voltage a maximum is passed and for higher voltages the SEY decreases again, which occurs for three reasons: First, $\mathrm{d}E/\mathrm{d}x$ decreases approximately with E^{-1} (Waddington law), so primaries with higher energies deposit less energy which can be converted to SEs. Second, a primary electron with higher energy permeates deeper, so the distance the secondaries have to overcome to the surface increases and the probability of losses grows [255]. And third, the material

[38] Full width at half maximum
[39] Note that the shown Cs_3Sb curve only applies to the measurement of a single series using a specific voltage distribution; in other sources SEY ≈ 11.5 at 600 eV [171].

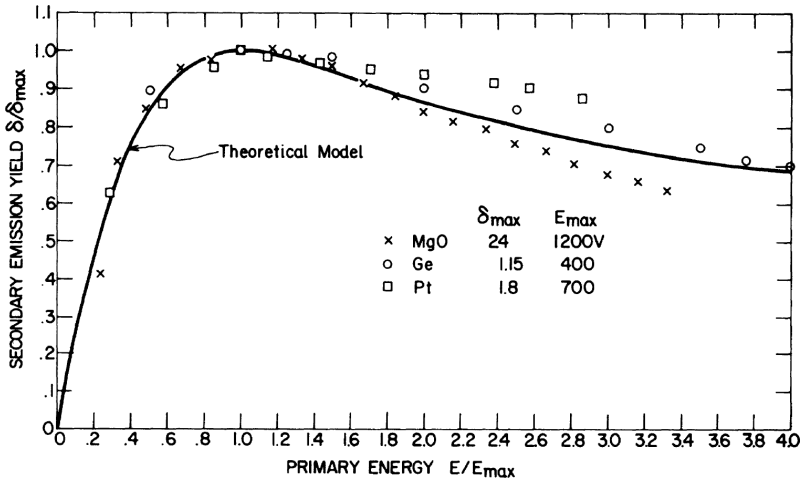


Figure 2.11: Reduced yield curve: Formula (line) compared to measured values for different materials: MgO (insulator, crosses), Ge (semiconductor, circles), Pt (metal, squares); the diversity of materials supports the notion of a universal curve according to [170; p. 215]. From [265].

with high secondary emission only is a thin layer on a conductive substrate, so for high energies the primary electron will reach the substrate where far fewer secondaries are produced.

All materials seem to follow a universal function, the reduced yield curve, if the axes are normalized to the values at the peak (fig. 2.11) [170, 261]:

$$\frac{\delta}{\delta_m} = \frac{1 - \exp\left(-\left(\frac{E}{E_m}\right)^{1.35}\right)}{\left(\frac{E}{E_m}\right)^{0.35}} \tag{2.7}$$

Angular dependence

For smooth surfaces the SEY increases with incident angle vs. surface normal ($\approx 1/\cos\theta$), as more secondaries are produced near the surface. For rough surfaces no angular dependence is observed since some electrons become trapped [170, 266]. The slanted angle of incidence is one of the reasons why the gain of the first dynode is much higher than for later dynodes.

Determination of the first dynode gain

Said gain at d1, $\delta 1$, can be deduced by illuminating the cathode such that the photosensitive d1 is also illuminated and comparing the charge of the $\gamma 1$ pulses (called prepulses in literature; see 2.3.5.1 and p. 282) to the normally amplified pulses. For strong illumination of d1 two peaks are visible in the charge spectrum, and $\delta 1$ can be obtained from the ratio of the peak positions (or the mean charges) [170].
An alternative is to use transit time cuts including only the $\gamma 1$, but then care must be taken that the mean charge is not polluted by dark noise pulses, which is only achievable for low DCRs and charge cuts.
For both methods one has to make certain to use an extremely low threshold and a noise-free setup, or else only a fraction of $\gamma 1$ will be visible due to the typically large $\delta 1$.

It is hard to verify that the center of the $\gamma 1$ charge distribution was contained in a measurement, since an observed drop to lower charges can also result from the height threshold, which smears the transition region in the charge distribution.

In 2.2.2 a different method is presented, which is based on fitting a compound Poisson distribution calculated from the estimated gains of d1, d2, ... to the spe charge distribution. By varying the fit parameter $\delta 1$ and calculating the other δn from the measured gain and the voltage divider resistance ratios, $\delta 1$ can be determined very reliably. Compared to the methods based on $\gamma 1$, this has the advantage that no very low thresholds are necessary and the DCR plays a far smaller role (as long as the charge spectrum is dominated by photon pulses – on p. 91 a criterion for this will be derived).

Typical values for dynode gain

Typical values for $\delta 1$ are 10 for Cs_3Sb, 20 for K_2CsSb:O and 40 for GaP:Cs (due to NEA [267]), for metals values lie close to 1, insulators[40] (BeO, MgO, ceramics, alkali halides) range from 6 to 20 [170].
The voltages between the dynodes are lower and typically lie around 100 V. δ therefore lies around 5, and ranges from about 3 to 10 [259]. The mean effective dynode gain δ_m (including losses) can be easily derived from the known total gain g of the PMT using $\delta_m^n = g$ with n the number of dynodes; for a gain $g = 1 \cdot 10^7$ and 10 dynodes $\delta_m = 5.01$. This also shows the large influence of even slight changes to the mean dynode gain, as a 1% increase in δ_m increases g by 10.5%.
To give an example, a Hamamatsu R5912 operated at $g = 1.3 \cdot 10^7$ (1425 V) with a front-tapered voltage divider (E7694-01) shows interstage voltages from 45 V (V_{d5-d6} and later stages) to 227 V (V_{d2-d3}) in the dynode chain, and has a (V_{k-d1}) of 514 V which is very close to the maximum in fig. 2.10.

Secondary electron energy distribution

In fig. 2.12a (see also [170, 172, 173]) a typical energy distribution of the emitted electrons is displayed. This is clearly separated in three contributions.

At low energies true secondary electrons are found, which were originally in bound states in the material. This constitutes 70–80% af all emitted electrons.

At the highest energies, a distinct peak is observed which originates from *primary* electrons, which are elastically backscattered on atomic nuclei (reflected primaries) nearly without energy loss. This accounts for a few percent of the spectrum.

The plateau extending over all energies originates from inelastically backscattered primary electrons, meaning that they are scattered after losing a part of their energy, e.g. by creating secondary electrons. For lower energies the distinction between inelastically scattered primaries and SEs blurs. The fraction of primaries undergoing inelastic backscattering ranges from 10 to 30%.

[40] Insulators are used throughout the PMT, which can affect the gain stability at light flux changes (light hysteresis), since they develop surface charges through electron bombardment if their δ is > 1.

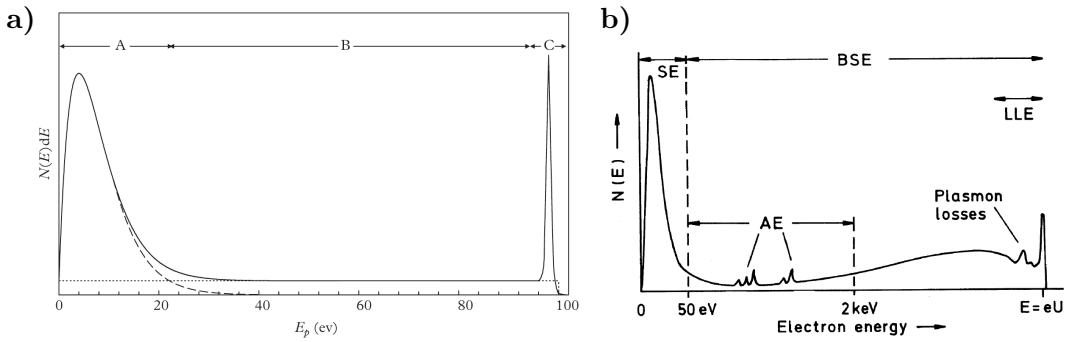


Figure 2.12: Secondary electron (SE) energy spectrum:
a) Diagram showing the essential features of the energy spectrum of emitted electrons following electron bombardment of a surface with primary energy of 100 eV, as found for metals, semiconductors and insulators; from [170; p. 211]. True secondary electrons (region A) have the lowest energies, inelastically backscattered primary electrons (region B) can be found at all energies with a more or less flat distribution, and elastically backscattered primaries (region C) form a peak slightly below the full primary energy.
b) Similar schematic for higher primary energies, showing deviations of the inelastic region from a flat plateau, in particular by loss effects through generation of Auger electrons (AE) and plasmons; from [268; p. 5]. The separation of secondary electrons and backscattered primaries at 50 eV is a convention in SEM. LLE denotes low-loss electrons in a window of about 10–100 eV below the primary energy, which includes elastic backscattering and backscattering after plasmon losses. In general, backscattered electrons (BSEs) with lower energies are from deeper inside the material [268; p. 5], which reduces escape probability. This leads to a rise towards large energies, which peaks below the elastic peak, however, since the interaction probability of the primary rises with traversed depth.

Actually the inelastic distribution is slightly convex and not flat, confer [232] and fig. 2.12b. However, for a Cs_3Sb dynode in a PMT, primary electrons with low energy enter a material with high effective atomic number Z_{eff}, which results in a flat interaction volume close to the surface (see fig. 2.13b). Thus the inelastic region in the SE energy spectrum shows no pronounced peak and is more or less flat, which justifies fig. 2.12a as model for PMT dynodes.

For all three types of emitted electrons the escape angle distribution follows a cosine law.

Backscattered electrons [41]

The term backscattered electrons (BSEs) denotes primary electrons which are emitted from the entry surface again through scattering inside the material. In case of elastic scattering the primary is emitted (nearly) without losing kinetic energy, for inelastic scattering after undergoing loss processes.
The fraction k of BSEs (backscatter coefficient) increases with the Z_{eff} of the material [42], the kinetic energy [43] and the incident angle [170, 268]. For common dynode materials Z_{eff} amounts to: BeO 6.7, GaP 26, Na_2KSb 34, K_2CsSb 44, and Cs_3Sb 54.

[41] Content based on [170; pp. 217–218], [268; pp. 135–152].
[42] For low energies as in PMTs, there is a slight drop of k with Z after a peak, for 1 keV after $Z \approx 30$ [268; p. 145].
[43] For small $Z < 30$ there is a slight *decrease* with energy [268; pp. 144–145]. This concerns many elements which can be hit by electrons in PMTs (Be, O, Na, Mg, Al, P, K, Ni, Cu).

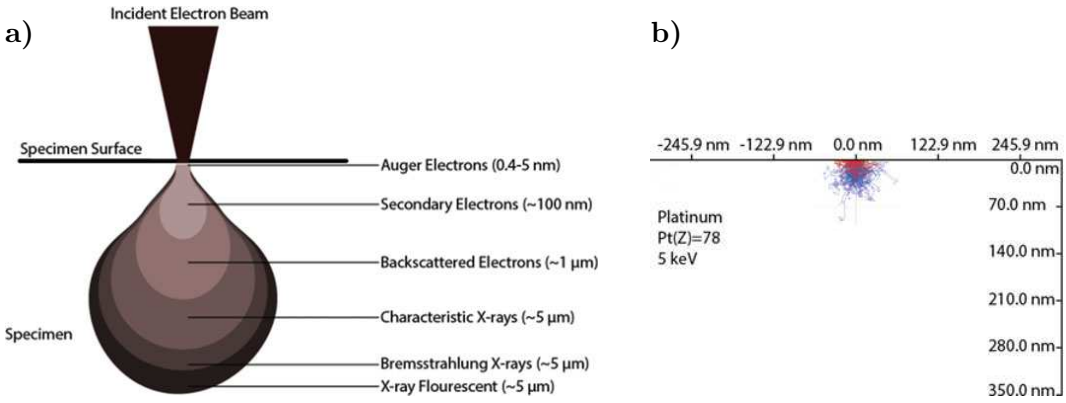


Figure 2.13: Interaction volumes of incident primary electrons in scanning electron microscopy (SEM):
a) Particle origins in the interaction volume for incident electrons with 20 keV and a material with low effective atomic number Z_{eff} (teardrop shape). From [269; p. 90].
b) Simulated interaction volume for a lower energy of 5 keV and a high-Z_{eff} material (hemispherical shape). From [269; p. 92].

The energy dependence of the backscatter coefficient $k(E)$ shows a steep rise at low energies and levels out at high energies (fig. 2.14). For Cs_3Sb k amounts to $\approx 25\%$ at 300 eV, lies above 30% for energies over 500 eV and reaches 36% at 1000 eV. The values for K_2CsSb are slightly lower. Wright concludes that backscattering of 30% is a reasonable number for the first dynode [170].
With increasing inclination θ from the surface normal the backscattering coefficient k increases $\propto (1 + \cos\theta)^{-9/Z_{\text{eff}}}$ [269].

Particles produced by electron bombardment

Much about the effects occurring in PMTs can be learned from the findings of scanning electron microscopy (SEM), transmission electron microscopy, and Auger electron spectroscopy [255–257, 268, 270]. The interested reader can delve far deeper into the subject of SE and BSE properties, as well as bremsstrahlung and cathodoluminescence in any good book on SEM, e.g. [268].

Based on the above sources, the (quasi)particles produced by electron bombardment encompass:

- **Secondary electrons (SE):** a) Excitation of electrons in the VB or CB above the vacuum level, the SE energies are typically below 50 eV. b) Ionization by ejection of inner shell electrons, which also causes the emission of characteristic X-rays or Auger electrons. The inner shell SEs can have larger energies, depending on primary electron energy. Due to scattering only SEs produced close to the surface can be emitted (also [268; p. 5]), see fig. 2.13a.

- **Auger electrons:** A higher shell electron fills the vacancy resulting from the ejection of an inner shell electron, transferring the energy difference to another electron, which is emitted with energies between 10 eV and up to several keV. For atomic Cs a broad range of Auger electron lines in the range of 15–75 eV was observed by [271] due to transitions between the outer shells. As for low

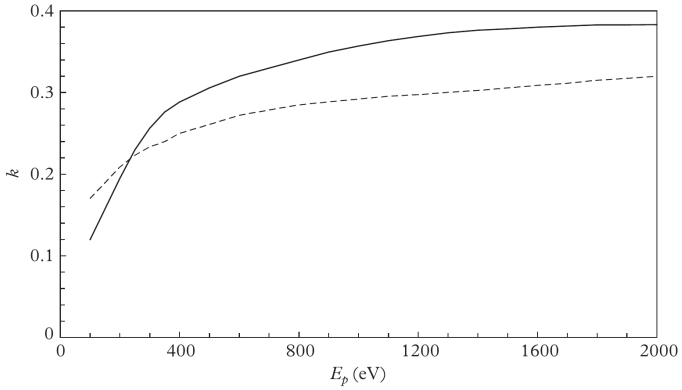


Figure 2.14: Backscatter coefficient k vs. primary electron energy for Cs_3Sb (solid curve) and Na_2KSb:Cs (dashed curve), from [170; p. 218]. K_2CsSb is expected to lie between both curves.

energy SEs, only Auger electrons generated close to the surface can escape (also [268; p. 5]), compare fig. 2.13a. In general, the Auger peaks in the SE energy spectrum are small [272] (see fig. 2.12b).

- **Characteristic X-rays:** Energies of Kα_1 range from 54 eV (109 eV for Be) to many keV [273]. For transitions between higher shells, the energy differences can become small, and in principle even cathodoluminescence (CB electron excited to the VB emits photon on recombination) could be considered a type of characteristic X-ray occurring at the outermost shells, where the excited electron later fills its own vacancy.

- **Backscattered electrons:** The BSEs in the elastic peak are from *elastic* Rutherford scattering of the primary electron in the Coulomb field of an atomic nucleus. The differential cross section is $\propto Z_{\text{eff}}/\left(E \sin^2 \frac{\Theta}{2}\right)$, with Θ the deflection angle. Due to the large mass difference the electron energy stays essentially constant. Elastically scattered electrons are only observed from a thin surface layer of a few nanometers depth, comparable in thickness to the origin regions of SE and Auger electrons [268; pp. 5–6] (see fig. 2.13a).
  BSEs with smaller energies stem from primaries which are subject to inelastic scattering (see next item) or elastic scattering after and/or before (inward/outward) incurring losses through production of other particles. These inelastic BSEs in general reach greater depths in the material [268; p. 6] as shown in fig. 2.13a.

- **Bremsstrahlung X-rays:** The *inelastic* deceleration of a primary electron near a nucleus through Coulomb force produces X-ray radiation with a continuous spectrum and energies ranging from 0 to the total primary energy [170; pp. 224–225] [195; p. 40]. The cross section is $\propto \frac{Z^2}{\beta^2}$ with $\beta = v/c$ [268; p. 383] and the intensity spectrum of produced photons drops linearly with rising energy [268; pp. 381–385]. At low energies the bremsstrahlung is absorbed unless it occurs close to the surface. The energy loss due to bremsstrahlung for primaries with < 1 keV is $<0.1\%$ of that from ionization and excitation [170; p. 224]. The energy distribution vs. primary energy can approximately be assumed to be constant over energy according to Kramer's relation [268; pp. 383–384]. The

differential number $N_b(E_b)\,\mathrm{d}E_b$ of photons at energy E_b emitted from a solid target is then [268; pp. 384–385]

$$N_b(E_b)\,\mathrm{d}E_b = K Z \frac{E_e - E_b}{E_b}\,\mathrm{d}E_b \tag{2.8}$$

where K is a constant, Z is the atomic number and E_e is the energy of the incident electron. Note that this is $\propto Z$ and not Z^2. Multiplying N_b by E_b yields the intensity, whose integral [44] over E_b is $\propto E_b^2$. The number of photons created at a fixed energy is $\propto E_e$.
The majority of bremsstrahlung photons are emitted parallel to or slightly away from the surface, depending on the fraction of primary energy they carry, but still a large amount can leave the material [268; pp. 382–383].

- **Electron–hole pair (e–h pair):** This is the basis for the internal photoeffect employed in semiconductors, where both charge carriers are separated, possibly amplified, and collected.

- **Cathodoluminescence:** In the recombination of an e–h pair produced through excitation in a semiconductor, a photon can be emitted with energy equivalent to the band gap [45], which typically lies in the visible range. This is the inverse of the photoelectric effect.

- **X-ray fluorescence:** X-rays ionizing atoms, which leads to emission of another X-ray, generally with lower energy.

- **Phonons:** Scattering on the lattice, heating up the material. The energy loss of a single phonon-scattering is small (in the meV range, $<1\,\mathrm{eV}$ acc. to [270]), nevertheless due to the abundance of phonons at room temperature this constitutes the main loss process and the largest fraction of primary energy is converted into phonons [268; p. 6]. The scattering angle is typically large [270], which should contribute to backscattering.

- **Plasmons:** Collective electron gas excitations occurring in materials with free or weakly bound electrons (metals and semiconductors). This causes energy losses in the range of 5–30 eV [257, 270]. Volume and surface plasmons are excited [172; p. 1-33]. Inter- and intraband transitions have a similar energy range as plasmons [257].

- **Change of chemical bound state** and **creation of crystallographic defects**

Auger electrons, followed by SEs, are from the regions nearest to the surface, BSEs occur deeper in the interaction volume, and characteristic X-rays, bremsstrahlung and X-ray fluorescence are created deepest in the material (fig. 2.13a).

In the above list, every effect except for elastic BSEs is inelastic in nature and causes energy losses for the primary electron. Many effects only become relevant

[44] In in (2.8) $N_b \overset{E_b \to 0}{\to} \infty$ and thus the approximation formula cannot be integrated over energy.
[45] In the unlikely case of recombination before thermalization the energy can be higher

at higher primary energies than present in PMTs, such as characteristic X-rays or X-ray fluorescence. The remainder can produce additional structures in the energy spectrum, such as plasmon peaks below the elastic peak and Auger peaks [257,268,272] (compare fig. 2.12b). SE energy spectrum measurements for low primary energies are scarce, and Wright knows of no measurement carried out for actual dynode materials [170], so fig. 2.12a might be a simplification.

Backscattering losses, early pulses and late pulses

The presence of backscattering, both elastic and inelastic, has serious implications for PDE, gain and time response.

BSEs have higher energies than SE, so they do not follow the SE paths the dynode geometries and fields are designed for and frequently hit the surrounding ceramic dynode mount, the glass bulb, the deck, the grid or other structures of the input system, where they are absorbed without secondary electron production [274].
For elastic scattering on d1 (ε-type, see 2.3.5.1) no secondary electrons are produced in the dynode, so the BSE does not trigger a cascade and no pulse is generated at this point in time.
For inelastic scattering, energy is dissipated either with production of SEs (B-type) or without (Λ-type) before the primary is ejected from the dynode with reduced energy. For B-scattering less SEs than normally are created, resulting in an undersized pulse, which can elude detection if it is smaller than the threshold.
All losses from undetected pulses are subsumed under the term backscattering losses (BSL) and contribute to the multiplication efficiency being lower than 1. Lorenz gives a number of as high as 10–30% for BSL in PMTs [274].

However, since the emission occurs with a cosine law, some backscattered electrons are emitted in direction of d2 where they *can* produce secondaries, start a cascade and generate a pulse [275].
For elastic scattering this yields an undersized pulse since amplification at d1 is skipped. Due to the higher initial kinetic energy than that of SE generated on d1 in a regular pulse, the distance from d1 to d2 is traversed faster and the pulse occurs some nanoseconds earlier; an early pulse (EP) of type [46] E1 results.
The same effect can occur for inelastic scattering (ΛE1 and BE1), in which case the pulse is not quite as early as for E1 due to the lower velocity between d1 and d2 and the resulting charge is even smaller, since the impact energy on d2 and thus the SEY is lower. However, for BE1 the secondary electrons produced at d1 create a *second* pulse (afterpulse, AP) following the inelastic EP at the normal transit time and with reduced charge, since only a fraction of the pe energy was used for SE creation. The delay between both pulses can be too small to resolve them. [232] notes that inelastically scattered primaries with less than about half the impact energy are successfully focused onto the next dynode in most linear-focused and box-and-grid dynode stacks, which should increase the probability of ΛE1 and BE1 observation, resp. reduce BSL.

Emission of BSEs can also occur towards the photocathode. For appropriate emission angles the electron will run against the potential of k–d1 towards the

[46] The nomenclature of this and the following effects is detailed in 2.3.5.1.

cathode, stop short of it (due to lateral momentum or very small losses in d1) and then get accelerated back onto d1 where it will produce a regular pulse with normal charge but delayed by the extra drift time spent in the loop [264,275]. This therefore produces a late pulse (LP).
For elastic scattering a L1-type LP occurs. For inelastic scattering (ΛL1, BL1) the pulse arrives earlier, since the traversed loop is smaller, and has a smaller charge. BL1 also produces a regular-timed small pulse running ahead of it.

Backscattered photoelectrons can also *directly* hit d1, since this dynode is especially large in order to increase collection efficiency, which will be termed E11 (pe hits d1 closer to d2 than emission position) and L11 (hit closer to cathode) for elastic scattering.

Inelastic scattering can also produce photons through e.g. bremsstrahlung, which can start a cascade on photosensitive structures (EP or LP depending on the structure) and produce a pulse even if the scattered primary does not.

For the mentioned types of EPs and LPs, no BSL are incurred if the pulses are large enough to pass the instrumental threshold. The backscattered photoelectrons are recycled, so to speak.
However, these effects only comprise a part of the BSEs and the largest fraction is lost, which gives rise to the BSL.
Merely interactions of a photoelectron from k with d1 were addressed so far; however, other structures can be hit as well, which produces different types of EPs, LPs and APs and also affects BSL. The various possible interactions and their probabilities are discussed in detail in 2.3.5, 2.3.6 and 2.5.2.2.

Reduced secondary electron emission

Another effect contributing to losses affecting the ME is if no secondary electrons are emitted from d1 despite a regular hit without backscattering (zero SE emission) [274]. Secondary electron production is a statistical process, which can be described by Poisson statistics with a few amendments [170], and the probability that no SE are produced close to the surface where they can escape is finite.
In addition, if only very few SEs escape d1, the pulse may slip under the hardware threshold and cannot be detected either (threshold losses).

Dynode types

For the electron multiplier a multitude of different layouts has been developed (figs. 2.15 and 4.3) [169, 170, 172]: linear focused, circular focused, box-and-grid, box-and-line, circular-and-linear focused, venetian blind, coarse mesh, fine mesh, foil, and metal-channel. The layout and properties of all dynode types will be discussed in detail in ch. 4 on p. 577. Some sources [169] even regard micro-channel plates (MCPs) and electron bombardment (hybrid photosensors such as the Hamamatsu HPDs, see p. 683) as electron multiplier subtypes[47].

[47] Since these multiplication structures are not based on discrete dynodes but on continuous dynodes (MCP) or create electron–hole pairs instead of emitting SEs (HPD), they are not considered PMTs but separate sensor types here. See 4.1.1.4 for a detailed discussion of the amplification processes employed in photosensors.

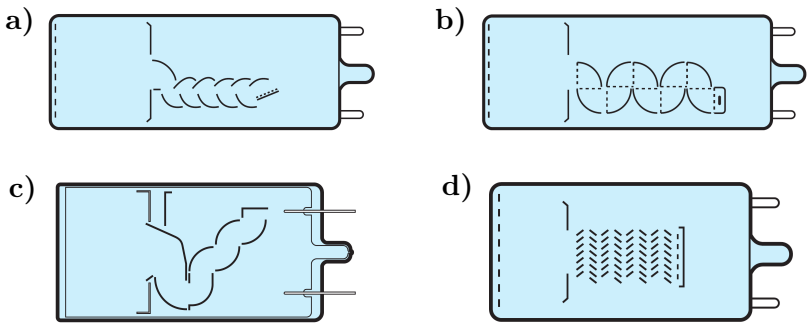


Figure 2.15: Dynode types: a) Linear focused (very fast, medium CE). b) Box-and-grid (slow, very good CE). c) Box-and-line, the combination of both (fast, good CE). d) Venetian blind (large TTS, low CE). All figures from [169; p. 45].

Most large modern PMTs use box-and-line dynodes with large first dynodes, which combines the benefits of a high CE from the box-and-grid type with shorter transit time, lower TTS and better linearity from the linear focused design.
In general, PMTs with small TTS tend to have low CE, since the electron optics is optimized for uniform transit times and not collection (see pp. 80 and 179).

Collection efficiency [48]

Dependencies

The collection efficiency – the probability that an emitted photoelectron hits the first dynode – depends on the emission position, initial energy (wavelength-dependent) and emission angle of the photoelectron, the geometry of the electron optics (cathode, bulb, focusing electrodes, grids), the voltage applied to all components and the type and shape of the dynodes. It is affected positively by using a spherical cathode, a large area first dynode or a focusing electrode with tunable voltage which can compensate assembly tolerances [172].

Loss paths

The pe can miss d1 and instead hit the cathode, the glass bulb, the grid in front of d1, the deck, d3, or the light shield behind d1 through the slit between d1 and the deck (fig. 2.16) [170, 276, 277]. If a pe terminates on the cathode, this in fact also affects the measured value of the QE (see below), since no net current is generated [277].
To give an example, according to simulations done by Wright for a 2" ETE 9214B PMT (see fig. 2.16) [277] with $V_{k-d1} = 300\,V$ a fraction of 72% of pe from k land on d1, 23% on the backside of d3, 3% on the dynode support structure, 1% on the metal light shield around d1, and the flux of photoelectrons hitting the cathode is nearly negligible (however, 1% at $V_{k-d1} = 50\,V$).

Non-uniformities

The electric fields must focus photoelectrons emitted from the complete photocathode onto d1, which is not possible for all trajectories. This can lead to large non-uniformities over the cathode surface, especially for fast, low-TTS PMTs, where the

[48] Content based on [170, 172, 274].

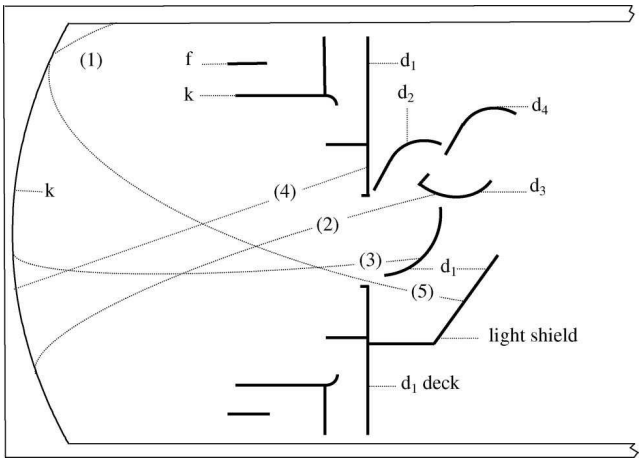


Figure 2.16: Simulated loss paths for photoelectrons in an ETE 9214B with potentials k = 0 V, d1 = f = 300 V, d2 = 400 V, d3 = 500 V. From [277].

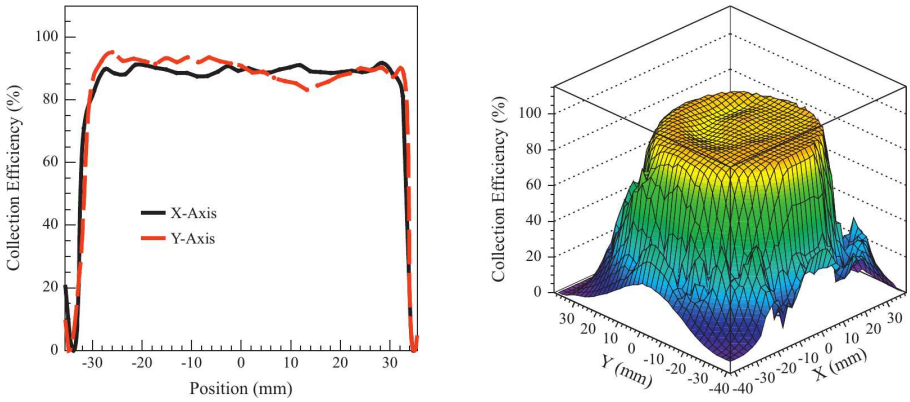


Figure 2.17: Pulse efficiency (PE) uniformity of a 3" Hamamatsu R11410-10. From [278].

orientation of d1 can often be clearly seen in the CE uniformity map [172]. These deviations are larger for photoelectrons originating in direction of the d1 symmetry axis than perpendicular to it [169]. The reason for this is that electrons from the lower part of the cathode of fig. 2.16 see a smaller d1 cross-section. The pe trajectory is more parallel to the d1 surface and can miss it more easily.
Photoelectrons from the region near the cathode brink are particularly hard to collect (fig. 2.17) since the extraction fields are smaller and the lateral momentum is larger; the 20" Hamamatsu R12860 for example has a CE of 95% for a reduced diameter of 46 cm but only 85% if the full diameter of 50 cm is illuminated [279].

CE uniformity is challenging to measure (see p. 86). Most published measurements are for PDE uniformity, which is largely affected by CE variations, and only a few publications show PE uniformity. The PDE uniformity over the photocathode typically amounts to 20–40% and can be even higher for side-on PMTs with reflective cathodes [169; p. 63].

Transit time spread versus collection efficiency

Often a small TTS will lead to low CE, since the electrostatic focusing is used to optimize both CE and TTS [177]. The CE can be sacrificed for a smaller TTS by

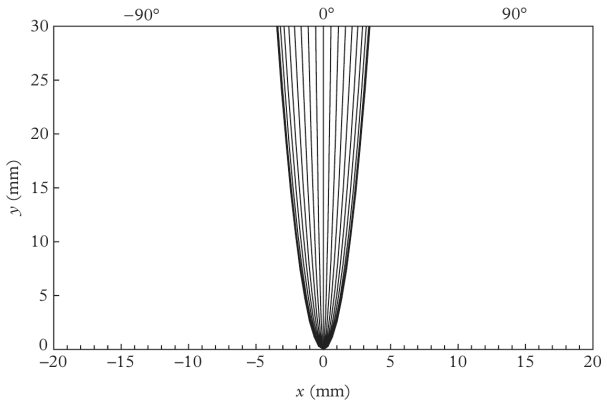


Figure 2.18: Electron trajectories in a linear field along the y-axis for launch angles from the surface from −90 to 90° for typical values of a photoelectron moving from the cathode to d1: 1 eV emission energy, 300 V interstage voltage and 30 mm distance k–d1; from [170; p. 332]. x is the transverse distance from the origin, y the interstage distance; the electron starts at (0,0) and shows parabolic trajectories with a rather high spread in x when arriving at d1 as a result of the varying emission angle.

incorporating shorter dynode distances or higher fields [280].
A good example is the difference between the NNVT GDB-6201 and GDB-6203 large area MCP-PMTs [251, 279]. The former shows a TTS FWHM of ≈14 ns and a CE of 98%, the latter was developed based on the GDB-6201 with the target of improving the TTS by adding a flower-shaped focusing electrode and shows a lower TTS FWHM of 4.3 ns but a reduced CE of only 85%.

Electron optics simulations

Recently developed PMTs are based on simulations of the input system taking account of the above-mentioned effects, which helps to reduce losses and optimize CE.
The newly developed 1.5" Cherenkov Telescope Array (CTA) PMT Hamamatsu R12992-100 boasts a CE of 94.6% at 400 nm and up to 98% for longer wavelengths at $V_{k-d1} = 350$ V with a good TTS of 2.4 ns [177]. The 20" Hamamatsu R12860 has a CE of still 85% (compare typical values on p. 82) while showing a remarkably low TTS of under 2.7 ns.

Transverse momentum

The transverse momentum of the photoelectron, which results from its emission energy and angle (fig. 2.18), is not focused by the electrical fields and therefore is one cause for loss paths [172, 274].
The emission energy is linked to the photon energy (fig. 2.4) and thus is wavelength-dependent [190, 247]. Above ≈ 500 nm the CE is nearly 100% due to the low exit energies, and decreases with smaller wavelengths due to higher initial velocities. Below ≈ 350 nm, the CE rises again, since the higher photon energy can sometimes produce two photoelectrons [274]. [172] observed the same effect below 200 nm but had no satisfactory explanation for it at the time.

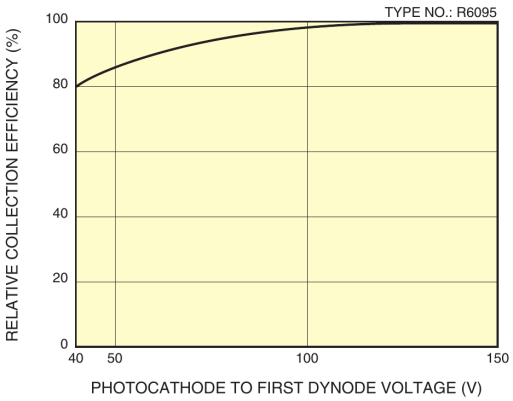


Figure 2.19: Relative CE vs V_{k-d1}, from [169; p. 48].

Dependence on high voltage

Raising the k–d1 HV increases the CE, since the drift time from k to d1 is smaller which reduces the lateral drift.
When increasing the HV, at the beginning the CE shows a steep rise and then reaches a plateau between 100 and 200 V (fig. 2.19), depending on the PMT [169, 170].

High voltages in combination with more uniform geometries are the reason for the extremely high CE of photosensors such as the X-HPD (20–30 kV, spherical cathode, CE ≈ 100%, but 50% BSE [242, 281]), Abalone (25 kV, hemispherical cathode, CE 99.99%, 30–40% BSE [282, 283]), Hamamatsu HPD R12850 (8 kV, mushroom shape, CE 97% for a reduced diameter of 46 cm, 80% for 50 cm [279]), and NNVT GDB-6201 (1.8 kV, ellipsoidal, CE 98% [251, 279]). As can be seen, the high CE due to strong fields comes at the cost of increased backscattering and corresponding BSL. For the R12850 and the GDB-6201 the BSE fraction is not known to date, but can be expected to be high as well.

It is often reasonable to keep the voltage between k and d1 fixed at the optimum performance point, since the electron optics are designed for a certain potential difference, the gain δ can decrease for too high HV, and the backscatter coefficient k will increase with pe energy. This can be done by using Zener diodes in the voltage divider [169; p. 100] or active dividers [171; p. 24] (see p. 144).

Values

For PMTs, the CE typically lies between 60 and 90% [169, 244], depending on whether the PMT is optimized for timing or CE. Newer PMTs developed using sophisticated electron optics simulations can range higher (Hamamatsu R12992-100: 94.6% at 400 nm, see p. 81). Photosensor types using large accelerating voltages can reach even higher values, often close to 100% (see last section). This, however, results in large BSE fractions and correspondingly low ME, which poses the question which HV might yield the highest PE.
However, since PE and CE are difficult to measure (see p. 86), in most publications numbers obtained from Monte Carlo simulations are quoted, which have to be treated with caution, since no simulation will incorporate all of the multitude of relevant factors.

Multiplication efficiency [49]

The multiplication efficiency (ME) denotes the probability that a photoelectron hitting d1 starts a cascade, which results in a detectable anode pulse.
The likelihood of successful amplification at d1 depends on hit position, incident angle, pe energy, as well as dynode surface material, layout and geometry. For pulse generation, however, also successful multiplication at the later dynodes is essential, for which the dynode type, geometries and field distributions are important.

Effects

An effect reduces the ME if due to it either no pulse or an underamplified pulse, which lies below the detection threshold, is produced. The latter depends on the electronics employed in the measurement, so strictly speaking one could also define a threshold efficiency (TE) in addition to QE, CE and ME. In [284] the TE is called front-end efficiency.
A reduction of the ME below 100% can have the following causes:

- **No pulse**
  - **BSL from scattering on d1** (elastic or Λ-type): The backscatter coefficient k rises with Z_{eff}, impact energy (saturates at high energies), and incident angle. However, not all backscattered photoelectrons are lost, since some are refocused onto active stages where they can start a pulse (e.g. E1, E11, L11, L1). Thus the BSL are smaller than k.
  - **Zero SE emission from d1** due to statistical processes: The mean SEY depends on material (composition, activation, vacuum, age), and increases with energy – i.e. HV (if too high the SEY drops again) – and incident angle. On the particle level, the SEY is determined by the minimum energy required to produce an SE, inelastic processes withdrawing energy from the primary electron, SE scattering, and potential barriers. Wright describes a statistical model for electron multiplier amplification [170; p. 161]. The zero SE probability on a dynode amounts to $e^{-\delta}$, resulting in losses of 0.67% for $\delta 1 = 5$, and a negligible $4.5 \cdot 10^{-3}$ % for $\delta 1 = 10$.
  - **Zero SE emission from later dynodes** [170; p. 161]: There is a small probability that none of the SEs in the cascade themselves produce SEs at d2 or even d3. V_{d1-d2} and V_{d2-d3} are smaller than V_{k-d1}, typically in the order of 100 V, so their δ is smaller. However, when assuming the same gain δ for all dynodes, for $\delta \geq 5$ the zero SE emission is already small on d2 (0.02%) and negligible beyond.
  - **Total SE loss:**
    - ▷ **Between d1 and d2** (CE_{12}): This can be seen as the dynode collection efficiency between d1 and d2. If all SEs are lost on their way to d2, no pulse is generated. The probability of complete SE loss is $(1-CE_{ij})^n$ with CE_{ij} the collection efficiency of an electron emitted from stage i onto stage j and n the number of electrons. Between

[49] Content based on [170, 274]

cathode and d1 this becomes $1 - \mathrm{CE}_{k1}$ (1 pe), between d1 and d2 $(1 - \mathrm{CE}_{12})^{\delta 1}$. In 2.3.5.3 the CE_{ij} are estimated for a Hamamatsu R5912.
SE drift losses depend on the hit position on d1, geometry and fields. Only photoelectrons entering the effective area of d1 are efficiently multiplied on later dynode without deviating from favorable trajectories (edge effect [50]) [169], otherwise a fraction of d1 secondaries lands on the ceramic support structure or hit the back of d3 [170; p. 225], which reduces the resulting charge. A large enough HV ensures that the pe paths end in this area. In fig. 2.20, the trajectories 1 and 5 come dangerously close to the edges of d2 resp. d3 and some SEs with high initial velocities might not hit the next dynode.

▷ **Between later dynodes** (CE_{ij} with $i \geq 2$, $j = i + 1$): This only results in complete pulse losses if *all* SEs are lost, which becomes increasingly unlikely for later stages and already is negligible at d2.

- **Threshold losses**

  – **BE1 and BL1:** If both the pulse from the scattered primary and the SE-pulse are smaller than the threshold, no pulse is detected. This is also the case if the backscattered photoelectron is lost (hits an inactive stage) and produces too few SEs on d1 before scattering to result in a pulse exceeding the threshold.

  – **E1, E11 and L11:** No detection occurs for pulses smaller than the threshold. E11 and L11 can have reduced charges through the edge effect.

  – **Reduced SE emission from d1:** When fewer than normal SEs are produced on d1 due to statistics the resulting pulse can be too small. This has the same reasons as zero SE emission.

  – **SE losses between d1 and d2**: This is the same effect as without pulse production, but far more likely, since it is sufficient that *most* SE are lost. CE_{12}, however, likely is already part of the effective $\delta 1$, which otherwise would be higher.

  – **SE losses between later dynodes**: These losses becomes increasingly *likely* for later stages, since more SEs are in the cascade and thus the number of trials increases. The losses, however, quickly become a statistical quantity, which is already part of the δ of each stage and the resulting pulse charge and has no influence on threshold losses anymore.

Values

To the author's knowledge, no measurements or simulations of the ME have been carried out so far; for most applications the knowledge of the PE, combining CE and ME, or the PDE is sufficient. Thus no typical values can be presented. However, the order of magnitude can be estimated from the effects listed above.

[50] In principle, the edge effect could also be considered a contribution to the CE if the definition is adapted from "fraction of photoelectrons which hit d1" to "fraction of photoelectrons which hit the effective area on d1".

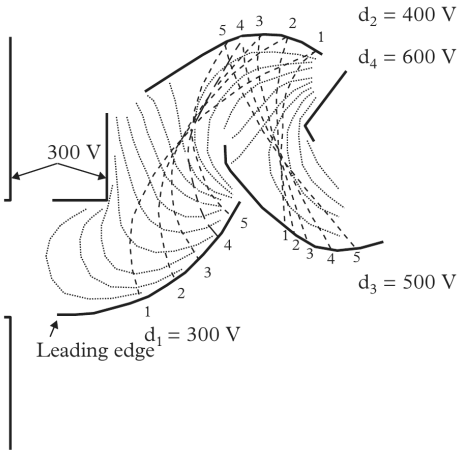


Figure 2.20: Electron trajectories and potential distribution for the first dynodes of a linear focused PMT (Philips XP1020), from [170; p. 338]. Dotted lines refer to equipotential lines. Note the crossover of all trajectories in between dynodes, which balances transit times, since long paths lengths (path 1 from d1 to d2) lead to short trajectories from d2 to d3 and vice versa. [170; pp. 338–339] This crossover is present for all dynodes. The extraction field at d1 is determined mostly by d3 due to its proximity, and not d2 as might be expected.

The fraction of BSL, the main contribution to ME < 1, can be deduced from the backscattering probability k and the likelihood of EP and LP appearance after backscattering, which reduce losses.
E1 occur with 0.05–0.75%, typically around 0.5%, L1 with 2–8%, typically 4% (see 3.1). BE1/ΛE1 and BL1/ΛL1 usually are included in the E1 and L1 numbers or are too small to be detectable and therefore constitute effective losses. E11 and L11 have not been measured prior to this work, but the effects are less likely to occur than E1 and L1.
Wright states a backscattering coefficient $k \approx 30\%$ for Cs_3Sb and slightly less for K_2CsSb [170].
Combining this with the recovery probability through EP/LP, the BSL assumed by [274] to lie in the order of 10–30% seems very plausible for these dynode materials.
For BeO dynodes, however, Z_{eff} is far lower and since backscattering scales with Z_{eff}^2, far less BSL, EPs and LPs are expected for this material.
For high voltages between cathode and multiplication stage entry, the BSE fraction can reach staggering values of as high as 50% (see p. 82). Correspondingly high BSL are expected.
Zero SE emission from d1 or later should only constitute a minute effect for realistic values of $\delta 1$. Reduced SE emission and SE losses on the way to d2 will produce additional reductions and could be estimated statistically resp. by simulations.

Putting everything together, while neglecting some effects with unknown dimensions, a ME of photoelectrons from the cathode on d1 of around 75% seems realistic for typical PMTs (compare the value of 65.6% calculated with time shift theory (TST) for a Hamamatsu R5912 on p. 233).
This value is as small as the CE, if not smaller, which shows that the ME has a large influence on PMT behavior and cannot be neglected as is often done.

Measurement [51]

Quantum efficiency

Quantum efficiency can be measured by illuminating the PMT with a known photon flux while operating it as a photodiode – i.e. by connecting all electrodes other than the cathode to the anode and applying the same voltage. This makes certain that all emitted photoelectrons and secondary electrons released by bombardment on the dynode structure are collected, measured as the output current. To determine the incident light flux, a calibrated photodiode is needed. The QE then is the ratio of incident photons n_p to emitted photoelectrons n_k.
PMT manufacturers typically measure the QE with parallel light incident on a defined central small area around the PMT rotational axis, but QE varies over the cathode surface, so QE uniformity scans are carried out by many experiments.

Photodetection efficiency

The PDE combines QE, CE, ME and TE, and thus depends on all effects affecting them. It is the probability that a detectable anode pulse is produced, and can be obtained from the simple relation n_a/n_p: the number of anode output pulses n_a after DCR subtraction divided by the number of photons incident on the window n_p. Thus, in order to measure the PDE an absolute photon source with known flux is required as well.
Since n_k depends on the threshold used for counting and the gain of the PMT, the PDE always refers to a specific combination of the PMT operating conditions (such as HV and temperature) and the measurement electronics, including the discriminator/flash analog-to-digital converter (FADC) with their settings and even the voltage divider which influences charge and pulse shape. The value is valid only if these exact conditions are reproduced, including the illumination (position, angle, polarization), which is why PDE surface scans are employed to obtain mean values or distributions as input for simulations.

Pulse efficiency

The pulse efficiency, the product of CE and ME, is notoriously difficult to measure without large systematic errors.
The PE can be obtained by measuring the constituents of any of the following formulas:

$$\mathrm{PE} = \frac{\mathrm{PDE}}{\mathrm{QE}} \tag{2.9a}$$
$$= \frac{n_a}{n_p \cdot \mathrm{QE}} \tag{2.9b}$$
$$= \frac{n_a \cdot e}{I_k} \tag{2.9c}$$
$$= \frac{G}{g} \tag{2.9d}$$
$$= \frac{\Delta 1}{\delta 1} \tag{2.9e}$$
$$\text{where} \quad I_k = n_p \cdot \mathrm{QE} \cdot e \tag{2.9f}$$

[51] Content based on [170, 277]

$$I_a = n_a \cdot g \cdot e \tag{2.9g}$$

$$G = \frac{I_a}{I_k} \tag{2.9h}$$

Here I_k is the cathode photocurrent, I_a the anode current, G the PMT gain (defined by currents), g the multiplier gain (defined by amplification, this is the "gain" normally used by physicists and in this work), and Δ_1 the current-gain of d1; see also p. 134.

Many measurement methods exist based on combination of these variables; however, the intrinsic problems of determining the respective properties limits the obtainable accuracy (see also p. 136).
In the classic method, first the cathode current I_k is measured at high photon flux, then a neutral-density filter (NDF) with optical density (OD) 5 is placed in front of the PMT and the spe counting rate is measured [285]. Eq. (2.9c) then becomes $\mathrm{PE} = n_a \cdot e \cdot f / I_k$, where f is the attenuation factor. The measurement of these variables illustrates many of the problems: Determining the photocurrent I_k poses severe difficulties, since at spe light fluxes the current would be smaller than picoampere. Thus, I_k must be measured at a high light level, which is then attenuated by NDFs or other filters to spe level to measure the anode count rate n_a, which is not possible at the high flux, since the output is not linear anymore. The attenuation factor f is measured in situ using the same PMT. Furthermore, n_a always depends on the chosen discriminator threshold, so to obtain an instrument-independent value, n_a has to be extrapolated to zero threshold. In addition, it is important to confirm that I_k is not so high as to be saturated due to the photocathode conductivity.
Other methods are based on the gain g, which can be obtained from the single electron response (SER) charge distribution or by measuring the anode current and count rate. Both methods introduce errors due to the non-zero threshold, which again would require extrapolation.
The accuracy can be improved by use of a calibrated light source or by a precise measurement of the attenuation factor with an integrating sphere and a monitor diode.
Wright has put forward a new method (eq. (2.9e)) which is based on determining the apparent gain $\Delta 1$ (defined by currents) at d1 and the true gain $\delta 1$ (defined by amplification) [277] and that uses no attenuation, the largest error source. This method has reduced systematic errors compared to previous methods, amounting to only 3%.
The most straightforward approach to measure PE might be to determine QE and PDE and use (2.9a), but this again depends on the employed electronics. Nevertheless, this is fine if measurement conditions are equivalent to operation conditions.

Which method to use is greatly disputed and the reader is referred to [170; ch. 10] for a thorough discussion of the benefits and drawbacks of seven different methods, including the one proposed by Wright himself.

Collection efficiency and multiplication efficiency

No simple method to measure CE or ME individually – and not their product PE – is known to the author.

The CE can be obtained from simulations (with the mentioned reservations), and in combination with a measured PE, the ME can be calculated. Measurements of the material properties and PMT effects can give an indication of the expected values as outlined above.

To measure the ME independently from the CE in order to better understand the influence of elastic and inelastic backscattering, one or more measurements determining all ME loss effects would be needed:

- **Losses from elastic scattering on d1:** The additional currents the backscattered pe produce on all other structures near d1 could be measured. However, the CE losses also hit these components. A gate for impact time might allow to separate both contributions if cathode time spread (CTS) and TTS from pe emission were not an issue and the drift time to d1 would be exactly known. In addition, leakage currents might further complicate this.

- **Zero SE emission from d1:** In this case the photoelectrons still enter d1 which results in a current out of d1. But the normal multiplication also produces a d1 current which has the opposite sign since many SEs leave d1. [277] provides a current model for various PMT components, which could be extended for that purpose in order to determine the zero SE emission through current measurements.

- **Zero SE emission from d2+:** Again, one could measure the current produced at dn; the same problem as at d1 applies.

- **Threshold-dependent losses** These could be measured by varying the threshold and extrapolating it to zero. This is complicated by the presence of additional underamplified EP, LP and AP effects (see 2.3.5, 2.3.6, 2.5.2.2).

Measuring the CE, on the other hand would require to count all photoelectrons which hit d1, including the ones which then get elastically backscattered and are subsequently lost, which produces no visible effect on d1. One could try to collect all backscattered photoelectrons, but this either becomes a QE measurement, changes the geometry and/or fields and thus the measured system, or requires gating either large fast voltage changes or acquisition windows with higher precision than CTS and TTS allow for large PMTs.

Overall, a clean ME resp. CE measurement – even only for photoelectrons from the cathode hitting d1 – seems challenging, if feasible at all.

2.2 Photon counting

2.2.1 Photon counting statistics [52]

For photomultipliers (PMTs) it is not possible to unambiguously determine the number n of concurrent photon hits (n pe) from the charge of the pulse alone. This is due to the strong fluctuations in gain, which result in a large charge resolution (defined on p. 119) and cause the distributions of neighboring n pe to overlap (confer fig. 2.21a).
However, for proper characterization of the sensor a clean sample of single-photon pulses (spe) with no admixture of 2- or 3-photon pulses (2 pe, 3 pe) is required. The most accurate way to achieve this is by using a very low light intensity, which, as will be shown, minimizes the probability of detecting two photons within the measurement window [288].

The case is easier for photosensors with a small charge resolution, e.g. silicon photomultipliers (SiPMs), since here spe-pulses can be selected reliably from the charge distribution (see fig. 2.25). Nevertheless, it is still recommendable to use a low light intensity, since otherwise the fraction of spe pulses is reduced, which increases measuring time.

Occupancy

A central variable to all low light level (LLL) photon counting applications is the occupancy o, the fraction of light pulses causing a PMT response:

$$o = \frac{n_p}{n_l} \tag{2.10}$$

where n_l denotes the number of light pulses from a light source (e.g. a laser) incident on the PMT and n_p is the number of times the PMT registers a light pulse by producing one or more anode pulses. n_l can also be called n_{tot}, as it denotes the total number of events; triggers which do not produce a pulse still fill the pedestal, if this is recorded.
The occupancy is either quoted in percent or as a number ranging from 0 to 1. Naturally, o depends on the chosen instrumental threshold. There are various other names for the occupancy, often confusingly closely resembling other terms: trigger probability, signal-to-trigger ratio, light intensity, light level, (detected) pulses per trigger, or detection probability.

Probability of observing a given number of pulses

A light pulse typically consists of many photons and even for LLLs it can statistically occur that more than one photon hits the PMT and produces a photoelectron (pe)

[52] Content based on [276, 286, 287]

respectively pulse. The resulting pulses are distributed in time according to the jitter of light source and sensor, but for a suitably chosen acquisition window all are contained.
The probability $p(n)$ of observing n PMT pulses per detected light pulse follows from Poisson statistics:

$$p(n) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{2.11}$$

$$p(0) = e^{-\lambda} \tag{2.12}$$

$$p(1) = \lambda e^{-\lambda} \tag{2.13}$$

$$p(2) = \frac{\lambda^2 e^{-\lambda}}{2} \tag{2.14}$$

λ is both the mean and the variance of the Poisson distribution.
Thus λ is the mean number of pulses produced by the PMT. It is usually quoted with the unit pe, even though in most cases output pulses are meant, and not photoelectrons which can get lost due to collection efficiency (CE) and multiplication efficiency (ME).

λ can be calculated from the occupancy, since o is the probability of observing one or more pulses, i.e. using (2.12)

$$o = p\,(n > 0) = 1 - p(0) = 1 - e^{-\lambda} \tag{2.15}$$

$$\overset{o\approx 0}{=} \lambda - \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots\,. \tag{2.16}$$

From this results

$$\lambda = -\ln(1 - o)\,. \tag{2.17}$$

For large incident fluxes the occupancy approaches 1; the probability of detecting pulses can not exceed 1. λ is always $> o$, but for values close to zero $\lambda \approx o$ and below $o = 2\%$ the relative deviation is $< 1\%$.

Multi-pulse contamination

For the measurements relevant here, typically a clean spe PMT response is desired, and contaminations by multi-pe pulses – mainly 2 pe pulses – should be minimized by setting a small value for o.
The absolute probability of observing 2 pe events is $p(2)$, which according to (2.14) for $o = 5\%$ ($\rightarrow \lambda = 0.0513$) is only 0.12%. However, since $p(1)$ is not large either (4.87% from (2.13)), the relative contribution of 2 pe pulses

$$\frac{p(2)}{p(1)} = \frac{\lambda}{2} \tag{2.18}$$

amounts to 2.56%, and the total contribution of 2 pe events to all detected events

$$\frac{p(2)}{o} = \frac{\lambda^2}{2\,(e^{\lambda} - 1)} \tag{2.19}$$

is 2.50%. This is a non-negligible amount despite the low occupancy of only 5%. The relative and total fraction of multi-pe events are

$$\frac{o - p(1)}{p(1)} = \frac{e^\lambda - 1}{\lambda} - 1 \tag{2.20a}$$

$$\text{and} \quad \frac{o - p(1)}{o} = \frac{e^\lambda - 1 - \lambda}{e^\lambda - 1} \tag{2.20b}$$

which in this example amount to 2.61% resp. 2.54%, slightly higher than the contribution from 2 pe pulses alone.

To keep the multi-pe contamination low, a low λ and accordingly a low o is required. To obtain the λ corresponding to a target contamination, (2.20a) can be equaled with a target value, but this can only be solved numerically. However, (2.18) can serve as a good substitute. For a target contamination $c_{2\text{pe}}$ with 2 pe pulses follows the maximum allowed mean number of photons $\lambda_{2\text{pe}}$:

$$\lambda_{2\text{pe}} \leq 2\, c_{2\text{pe}} \tag{2.21}$$

For example, if the relative 2 pe fraction should be $< 1\%$, $\lambda_{2\text{pe}} \leq 0.02$ pe and $o_{2\text{pe}} \leq 1.98\%$ are required.

Dark noise contamination and measurement time

Ideally one could improve the purity of a spe photosensor response arbitrarily by simply lowering o. Two aspects make this unfeasible.

For one thing, this increases measuring time, since $p(0)$ approaches 1 and most light pulses will produce no PMT pulse at all. This would still be manageable by increasing the pulse frequency (falling below a period of 30 µs, however, should be avoided to exclude spurious events from ionic afterpulses (iAPs) of the previous pulse in the detection window) and sufficient patience.

For another thing, random coincidences due to the dark count rate (DCR, f_d) of the photosensor pose a sensor-intrinsic and therefore unavoidable statistical background. The light source trigger opens a coincidence gate with length Δt for time-correlated photosensor pulse detection, with frequency equal to the light pulse rate f_l. The trigger rate caused by dark counts f_D is the probability per time to observe one or more dark noise (DN) pulses in the trigger window. This observes Poisson statistics and can be calculated similar to (2.15). For small f_D, however, (2.16) can be applied and the dark count trigger rate then is given by

$$f_D = f_d \cdot f_l \cdot \Delta t \tag{2.22}$$

whereas the event trigger rate f_e is

$$f_e = o \cdot f_l \tag{2.23}$$

From (2.17), (2.22) and (2.23) follows the recommended λ for a target DCR contamination $c_d = f_D / f_e$:

$$\lambda_d = -\ln\left(1 - \frac{f_d \cdot \Delta t}{c_d}\right) \overset{\lambda \approx 0}{\approx} \frac{f_d \cdot \Delta t}{c_d} \tag{2.24}$$

This means that the lower the desired dark count background fraction c_d, the *higher* λ and therefore o need to be. For e.g. $c_d = 1\%$, a DCR of 2 kHz and a window of 100 ns, a $\lambda_d \geq 0.02$ pe is required.

Non-ideal measurement conditions

These values match the maximum $\lambda_{2\text{pe}}$ from (2.21) for $c_{2\text{pe}} = 1\%$. However, if the DCR is higher (for example for large high quantum efficiency (HQE) PMTs) or the gate needs to be longer to register time shift effects (see 2.3) such as $\gamma 1$, L1 (possibly also L1L1), and early afterpulse (EAPs, see 2.5.2.2), for a PMT with a large transit time the value can quickly become unreasonably high for maintaining an adequate 2 pe-fraction.
In these cases the gate length should be kept at a minimum, and the DCR lowered as much as possible by letting the PMT run for some time with high voltage (HV) turned on or cooling it (if not conflicting with the planned environmental measurement parameters).

If this is not possible or sufficient, either a higher 2 pe contamination or a higher DN background have to be accepted.
The former produces an additional 2 pe distribution in the charge spectrum – typically greatly overlapping with the spe distribution due to the large charge resolution (see 2.2.2 and fig. 2.21a). A larger 2 pe fraction also artificially improves the transit time spread (TTS) a bit, since pulses can overlap. This causes a more average point in time when exceeding a set threshold, and the variation of this crossing time is used to determine the TTS (see 2.3.3 and eq. (2.131)). In addition, late pulses (LPs) will not be detected correctly if a regular pulse (RP) or early pulse (EP) occurred before. In case of the latter, the DN background covers all fine (i.e. rarely occurring) features in the transit time distribution (TTD) (see 2.3.2). Furthermore the contribution of dark counts to the charge spectrum lowers the mean charge due to the underamplified dark count pulses originating from the first dynodes (see p. 106).
The experimentalist has to weigh which issues are more grave and select the occupancy accordingly.

Recommended occupancy

Typically, the occupancy should lie in the range between 1 and 10%, as lower values raise the DN plateau in the TTD too far and higher values lead to a unacceptably large 2 pe contribution ($> 5.3\%$ for $o > 10\%$). If the photosensor shows a low DCR, o can be further lowered to improve spe purity, or be kept at a medium value to resolve more features in the TTD. [276] recommends a value of $o = 5\%$, [286] used settings around 2% for the ET [53] 9351, which had a moderate DCR, [170; p. 367] advises to use values between 1 and 10%.

Early afterpulse measurements

To study early afterpulses (EAPs, see 2.5.2.2), a low o is beneficial for several reasons. For one thing, the first pulse (primary pulse, PP) should be from single photons only,

[53] Electron Tubes Ltd.

as a large 2 pe fraction makes it harder to associate afterpulses (APs) to the primary pulses which created them and also to determine whether an AP was caused by a PP or an AP (secondary afterpulses).
In addition, when a second photon creates an LP (e.g. L1 or ΛL1; see 2.3.6.4 for all possible effects) following a regular pulse from the first photon, the LP is indistinguishable from an EAP. The likelihood $p_{2\text{pe LP}}$ that an LP-effect occurs for one of two detected photons can be calculated using (2.19), the probability p_l that the LP effect occurs in the TTD (see 2.3), the integrated probability p_b that a pulse lies before that effect in the TTD, and that the order of the two pulses is irrelevant:

$$\begin{aligned} p_{2\text{pe LP}} &= p_b \cdot p_l \binom{2}{1} \frac{p(2)}{o} = p_b \cdot p_l \cdot 2 \cdot \frac{\lambda^2}{2\left(e^\lambda - 1\right)} \\ &= p_b \cdot p_l \cdot \frac{\lambda^2}{e^\lambda - 1} \overset{\lambda \ll 1}{\approx} p_b \cdot p_l \cdot \lambda \end{aligned} \tag{2.25}$$

For small λ this contamination of the EAPs is proportional to λ. Taking typical values of the L1 probability (1.5%) a value of 0.16% results for $o = 10\%$. This already constitutes a noticeable contamination and other LP effects which will be introduced in 2.3.5 can have far higher probabilities of up to 10%, so a lower occupancy should be chosen.
The EAP rate (see 2.5.2.1) also depends on the occupancy, since additional detected photons can also create afterpulses, but to a far smaller extent. As will be derived in (2.203), the EAP rate grows with occupancy with the factor $\frac{\lambda}{1-e^{-\lambda}}$, which for $\lambda \ll 1$ is approximately 1.
Thus it is recommendable to study EAP effects at low occupancies, since it improves the signal-to-noise ratio (here EAPs as signal, 2 pe LPs as noise) and allows unambiguous association of APs to PPs.
If two measurements at different levels of o are performed it is even possible to identify the LP populations through their different occupancy scaling factor.

On the other hand, some EAP effects have small occurrence probabilities and are affected by DN background, so o should not be too low.

Ionic afterpulse measurements

For timing and EAP measurements low occupancy values between about 1 and 10% are recommended due to the compromise between 2 pe contamination (eq. (2.21)) and DCR (eq. (2.24)).
The measurement of iAPs (see 2.5.2.3) is a different matter. Here, some groups use very large λ values of 20–100 pe [227]. Response linearity and independence between primary pulses can still be assumed at this number, so Poisson statistics are therefore still valid. [289] measured a linear relation of AP number vs. primary pulse charge up to $\lambda = 400$ pe for a Hamamatsu[54] R5912, IceCube performed linearity measurements up to $\lambda = 4.4 \cdot 10^6$ pe in 40 ns pulses for a Hamamatsu R7081-02, which showed near linear behavior up to the maximum intensity [290].
The advantage of measuring with a high o is that measurement time can be greatly reduced, since iAP rates lie in the order of only a few percent per primary pulse. Otherwise for e.g. flash analog-to-digital converter (FADC) waveform acquisition a great number of waveforms have to be recorded, which do not contain iAPs and waste

[54] Hamamatsu Photonics K. K.

storage space – FADC readout often is the limiting factor in obtaining sufficient statistics. An alternative is to use a coincidence and record only waveforms containing APs, but then the AP probability has to be measured electronically and cannot be obtained in offline analysis for varying thresholds.
Another benefit of a large occupancy is the tremendous reduction of the DN contamination, which allows to observe the trailing flank in the delay distribution (DD) up to far longer delays before it drowns in dark noise.

Many groups, however, use spe light levels (o around 10%) for iAP studies [276, 291, 292]. This allows to better identify higher-order APs, i.e. APs caused by previous APs. Furthermore, linearity consideration are not an issue.

If the experimentalist has sufficient time, therefore, the recommendation is to perform measurements both at low and high o and compare the results.

Generally, for iAP measurements, the threshold should be as low as possible, since many iAP populations originate from interdynode ionic afterpulses (iiAPs) and have mean charges of only a fraction of the mean spe charge.

Measuring occupancy and mean number of pulses [55]

Since the occupancy o and the mean number of pulses λ per light pulse are such central quantities for each LLL measurement, this raises the question how they can be measured.
For the methods presented in the following, in principle both the distribution of the pulse charge (see 2.2.2) and of the pulse height (see 2.2.3) can be used. However, the height is more sensitive to electronic noise on the baseline. This will exaggerate the maximum pulse height obtained with any analog-to-digital converter (ADC) for empty waveforms, which can require to raise the threshold and exclude more underamplified events. For PMT pulses the region around the pulse peak is far shorter than the complete waveform so noise fluctuations increase the obtained maximum value on average less strongly.
In the charge integration the contribution from noise is mostly averaged out, making the charge spectrum slightly more reliable.

Counting events without sensor response

The easiest method to measure the light level is to use the fraction of events without sensor response, which equals the events contained in the pedestal of the charge or height distribution (cf. fig. 2.21a):

$$p(0) = \frac{n_{\mathrm{ped}}}{n_l} \tag{2.26}$$

where n_l was the number of light pulses and n_{ped} is the number of events contained in the noise pedestal, the Gaussian distribution around zero charge (resp. height) after baseline subtraction.

[55] Content based on [286]

From eqs. (2.12) and (2.26) follows

$$\lambda = -\ln\left(p(0)\right) = -\ln\left(\frac{n_{\text{ped}}}{n_l}\right). \tag{2.27}$$

In order to count n_{ped}, a cut separating pedestal from underamplified pulses must be defined. This can be done by fitting a Gauss function to the pedestal and comparing the fit function with the charge distribution (see fig. 2.21).
A possible improvement is to shift the underamplified events covered by the pedestal from n_{ped} to n_l. The fraction of hidden undersized pulses can be obtained by extrapolation of the underamplified pulse distribution (often approximated by an exponential decay) to zero charge or height.
λ can be corrected for DN contributions if the DCR and the DN charge distribution are known through [293]

$$\lambda = -\ln\left(\frac{n_{\text{ped}}}{n_l}\right) + \ln\left(\frac{n_{d,\text{ped}}}{n_{d,\text{l}}}\right). \tag{2.28}$$

where $n_{d,\text{ped}}$ is the expected number of dark count pulses below the chosen threshold and $n_{d,\text{l}}$ is the total number of dark count pulses expected in the acquisition window for n_l light source triggers. The DN contributions to the recorded events can be obtained from (2.22).

Counting events with sensor response

Alternatively, another simple method is to directly measure the occupancy $o = n_p/n_l$ by counting the events which produced one or more pulses n_p with a discriminator and a counter and measure the number of light pulses n_l produced during the same time window. From (2.17) and (2.27) follows that this is equivalent to the last method, since $n_l = n_{\text{ped}} + n_p$.
This method also assumes knowledge of the charge/height distribution because a cut has to be set to discriminate events from noise, and the same improvement through redistribution of the extrapolated underamplified pulses can be applied. It is thus recommendable to first have a look at the charge distribution with a digital oscilloscope or FADC to define a suitable threshold. This instrument then can also be used in place of both discriminator and scaler.

Mean single photon charge

Since the mean charge of the whole charge spectrum, including the pedestal, is $\overline{q} = q_1 \cdot \lambda$ with q_1 the mean spe charge, the mean number of pulses follows to

$$\lambda = \frac{\overline{q}}{q_1} \tag{2.29}$$

This requires exact knowledge of q_1. This can be obtained by low (pedestal) and high (2 pe) cuts in the charge spectrum, more accurately from fits of the spe, 2 pe and underamplified populations, or from a clean spe sample (low o). As a crude first estimate, the spe modal value can be used in lieu of q_1, but this neglects the underamplified pulses and will be 10–20% larger than the true value.

Charge variance

For an ideal photosensor one could derive λ from the relative uncertainty (resolution) σ/μ of a Poisson distribution [287; p. 38]:

$$\frac{\sigma}{\mu} = \frac{\sqrt{\mu}}{\mu} = \frac{1}{\sqrt{\mu}} \tag{2.30}$$

where σ is the standard deviation and μ the mean. Here $\mu = \lambda$.
The mean charge in the anode output pulse is

$$\overline{q} = \lambda \cdot g \cdot e \tag{2.31}$$

where g is the mean gain. Assuming ideal amplification by adding a factor $g \cdot e$ to both numerator and denominator on the left side of (2.30) one obtains

$$\frac{\sigma_q}{\overline{q}} = \frac{1}{\sqrt{\lambda}} \tag{2.32a}$$

$$\text{and} \quad \lambda = \frac{\overline{q}^2}{\sigma_q^2} \tag{2.32b}$$

where σ_q is the standard deviation of the charge.
To measure λ, then all which would remain to be done, is to measure σ_q and the mean charge $\overline{q}$ over the whole charge spectrum including the pedestal.

In reality, PMTs are *not* ideal amplifiers and (2.32b) cannot be used without incurring large errors. The excess noise factor[56] (ENF) F describes the deviation of an amplifier from the ideal case and reads

$$F = \frac{\sigma_{\text{out}}^2}{\sigma_{\text{in}}^2} \overset{\text{PMT}}{=} 1 + \frac{\sigma_g^2}{g^2} \tag{2.33}$$

where σ_g is the standard deviation of the gain, i.e. taking spe events only. Eq. (2.33) will be derived in 2.2.2. For good PMTs with high $\delta 1$ the ENF lies close to 1, for example for $\delta 1 = 10$ it can be derived to be about 1.2 [267].
Using the relative variance of the charge $v_q = \sigma_q^2/\overline{q}^2$, (2.32a) becomes

$$\frac{\sigma_q^2}{\overline{q}^2} = \frac{1}{\lambda}\left(1 + \frac{\sigma_g^2}{g^2}\right)$$

$$\text{and} \quad \lambda = \frac{\overline{q}^2}{\sigma_q^2}\left(1 + \frac{\sigma_g^2}{g^2}\right). \tag{2.34}$$

This equation can be used to measure λ by determining the standard deviation and mean from the complete charge spectrum (σ_q and $\overline{q}$) and from the spe response only (σ_g and g). Yet another time, this depends on the used cuts or fits.

[56] Not to be confused with the statistically similar Fano factor F_F – the underlying physical processes are very different: The excess noise factor (ENF) describes *gain noise* due to the multiplication, whereas the Fano factor quantifies the *conversion noise* of e.g. a primary particle losing energy by production of secondaries in a Germanium detector.

Fitting the charge spectrum

Fitting the complete charge spectrum with all populations (pedestal, spe peak, spe underamplified pulses, and all n pe distributions with curves obtained from the spe contributions) was found to be the most accurate method by [286]. λ is then just a fit parameter and is obtained from the ratio of all n pe populations to each other.
In [286] the spe peak and pedestal were approximated by a Gaussian and the underamplified pulses by an exponential decay; other authors use Pólya (fig. 2.21b) [276] or Poisson distributions [170] for the spe peak, and a Poisson distribution with reduced mean for low-charge pulses [170]. More details about the fit functions can be found in the respective papers.
Alternatively, somewhat less accurately, λ can also be reconstructed from fitting only the spe and 2 pe peaks and using formula (2.18). This of course assumes a negligible influence of underamplified pulses and the 3 pe population in the fit window around the spe and 2 pe peaks. One could even use the ratio of the modal values of the spe, 2 pe, and maybe larger n pe distributions (if present) as a first estimate.

Gain correction factor

The photosensor gain can be determined from the mean value of the charge distribution. The pedestal can be excluded by a low threshold, which introduces errors from neglected low-charge pulses, which can be amended by extrapolation. On the high charge end, due to the large charge variance a high cut would either cut away a large fraction of spe pulses while still not completely excluding the 2 pe population, or encompass all spe pulses at the cost of a large 2 pe contamination.
Luckily, if o or λ is obtained by any of the methods previously mentioned, one can determine the gain by integrating over the charge distribution without pedestal and simply multiply the obtained value with a gain correction factor η_g to automatically eliminate the influence of multi-pe pulses. Since λ is the mean number of pulses per trigger, o is the probability to observe any number of pulses per trigger, and the charge is proportional to λ (see (2.31)), this reads

$$\eta_g = \frac{o}{\lambda} \tag{2.35}$$

λ resp. o can be calculated from each other using (2.17) or (2.15).

2.2.2 Charge

When a photon produces a PMT pulse, a wave of secondary electrons produced in the cascade and collected by the anode leaves the output. This can be measured either in form of a current pulse or, since the current will be connected to ground via a load resistor of typically 50 Ω, as a voltage pulse. The charge contained in this pulse can be obtained by integrating over a suitable signal window. If this is repeated many times for a light flux with an occupancy in the spe range with constant operating conditions, a charge spectrum such as shown in fig. 2.21a will be obtained.

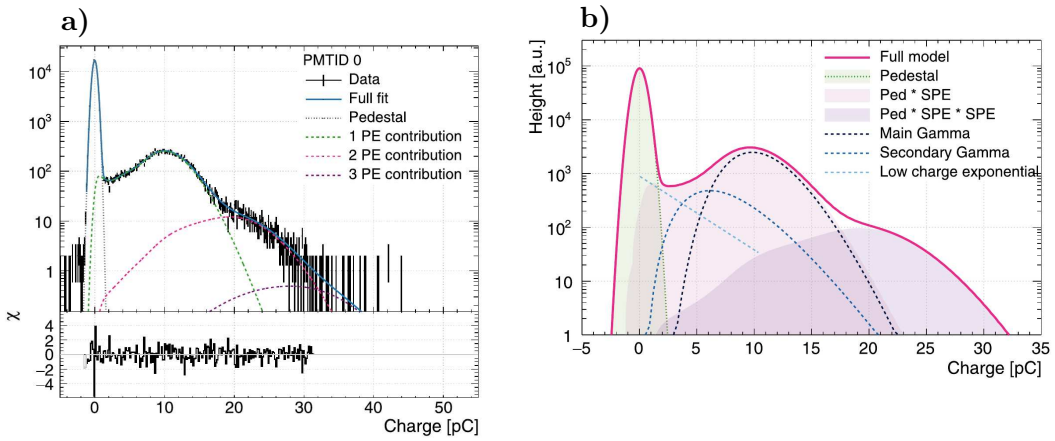


Figure 2.21: Charge distribution (SER):
a) SER of a Hamamatsu R5912 high quantum efficiency (HQE) PMT recorded at $o = 15\%$ with $g = 6 \cdot 10^7$, fitted with the function shown in b). From [276].
b) Advanced fit model for the charge spectrum, containing: pedestal (Gaussian); spe contribution: main spe peak (gamma function, approximating a continuous Pólya distribution), underamplified spe pulses with reduced gain (gamma function), underamplified spe pulses from inelastic scattering (exponential decay). Higher n pe contributions are obtained by convolving the spe distribution with itself n − 1 times. As a final step all resulting n pe distributions are convolved with the electronics noise, i.e. the normalized pedestal. From [276].

The charge response of the PMT to a single photon source is called the single electron response (SER) [170, 267] – named after the single pe which is created – and can be used to determine the gain, photon counting capability and the incident light flux (see 2.2.1).
The charge is usually quoted in pC, but can also be normalized to the mean spe charge, using the unit pe.

Components of the charge distribution [57]

The SER contains:

- **spe pulses:** The spe response consists of two parts:
  - **Regular pulses (RP):** The normally amplified pulses
  - **Underamplified pulses:** Created by a multitude of effects which will be detailed later.
- **Multi-pe pulses:** For realistic conditions, the occupancy will be > 0 and small multi-pe populations (2 pe, 3 pe, ...) will be visible in the SER, as in fig. 2.21a.
- **Pedestal:** Furthermore, the SER will contain a large Gaussian pedestal around or near zero charge caused by the noise of the electronics used to record the spectrum. The pedestal is *not* an intrinsic property of the photosensor, but of the setup.

[57] Content based on [169, 171]

- **Dark counts:** Also, dark count pulses are present in the spectrum with on average smaller charges than spe pulses (see p. 108).

- **Afterpulses:** Afterpulses from previous events or caused by dark noise can be found at large charges up to above 15 pe, but also at spe values or below due to interdynode ionic afterpulses and early afterpulses.

- **Ionizing radiation:** Finally, there are contributions from radioactive decays inside or near the sensor with charges between spe and about 15 pe, and occasional cosmic ray hits (mainly muons and secondary electrons), depending on photosensor cross-section and overburden, which produce very high charges. These effects create PMT pulses by Cherenkov radiation in the glass and pe production in the cathode.

Thresholds

Due to the unwanted contributions to the charge spectrum usually two cuts are employed, one at a low threshold to exclude the electronics noise (see p. 47), and one at a high value which cuts off large APs and particle hits and whose exact position depends on the occupancy and the aim of the measurement.

Regular pulses

Normal distribution

The easiest way to describe the regular pulse distribution is with a normal distribution. The underlying amplification processes are governed mainly by Poisson statistics, but for a large mean value λ of the Poisson function, a Gaussian distribution with mean λ and standard deviation $\sqrt{\lambda}$ is a very good approximation [287]. In many cases the spe peak can be well described by a Gaussian fit.
[286] gives a fit model which incorporates the error function to account for the part of the normal distribution which is cut off at 0, which can well account for a non-negligible fraction due to the typically large standard deviation. The suggested fit function is[58]:

$$p(q) = \begin{cases} q > q_p : \frac{1}{\sqrt{2\pi}\,\sigma}\frac{1}{G_N}\exp\left(-\frac{1}{2}\left(\frac{q-\overline{q}-q_p}{\sigma}\right)^2\right) \\ q \leq q_p : 0 \end{cases} \tag{2.36}$$

$$\text{with} \quad G_N = \frac{1}{2}\left(1+\operatorname{erf}\left(\frac{\overline{q}}{\sqrt{2}\,\sigma}\right)\right)$$

where q is charge, q_p the pedestal position, $\overline{q}$ the mean charge, σ the standard deviation, and G_N the normalization of the Gaussian.

[58] Comment of the present author: For a Gaussian function with mean μ and standard deviation σ, the part lying below zero is $f_0 = \frac{1}{2}\left(1+\operatorname{erf}\left(\frac{-\mu}{\sqrt{2}\sigma}\right)\right)$, which leads to a correction factor $\frac{1}{1-f_0}$. Using $\operatorname{erf}(-x) = -\operatorname{erf}(x)$, $1-f_0 = \frac{1}{2}\left(1+\operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right)$.

Poisson distribution

Using a Poisson distribution is often more accurate. By using the gamma function in place of the factorial, a continuous distribution can be obtained:

$$p(n,\lambda) = \frac{\lambda^n}{n!}\, e^{-\lambda} = \frac{\lambda^n}{\Gamma(n+1)}\, e^{-\lambda} \tag{2.37}$$

$$= e^{\ln\left(\frac{\lambda^n}{\Gamma(n+1)}\, e^{-\lambda}\right)} = \exp\left[n \ln\lambda - \ln\left(\Gamma(n+1)\right) - \lambda\right] \tag{2.38}$$

Eq. (2.38) has the added advantage of being numerically stable to compute; in the original expression (2.37) both the factorial and the term λ^n are problematic.

For a gain in the order of 10^7 a Poisson with such a high mean, i.e. λ, looks like a narrow Gaussian, which is very different from the observed distribution (fig. 2.21). As will be elaborated in the next paragraphs, this is because instead of one amplification step multiple gain processes are chained where the overall shape is mostly determined by the gain on d1 (dn denotes the n-th dynode). Thus the SER most likely can not adequately be fitted with a Poisson function using (2.38), since the unit of the charge axis (pC) does not correspond to the number of SEs produced on d1, which governs the shape of the distribution. The fit will randomly produce good or bad results depending on the total gain.

To solve this discrepancy, the x-axis can be scaled by multiplying n and λ with a factor s, which is a free fit parameter. This is equivalent to assuming one Poissonian gain step on d1 with an effective λ, which takes into account $\delta 1$ and a small smearing from later dynodes which reduces λ (see below), and a fixed gain from all dynodes after d1. Inclusion of the scaling parameter for the x-axis provided satisfactory fit results for all measurements made during this work.

Compound Poisson distribution

The Poisson distribution actually is *discrete*, and not one, but multiple serial Poisson processes take place due to the subsequent amplification on each dynode. [294] notes that this results in a compound Poisson distribution.

Wright describes a theory allowing to calculate the resulting charge distribution after the k-th dynode based on a few simple iterative formulas [170; ch. 4]:

$$n \neq 0: \quad p_k(n) = \frac{\delta}{n} \sum_{i=0}^{n-1} (n-i)\; p_k(i)\, p_{k-1}(n-i) \tag{2.39a}$$

$$p_1(n) = \frac{\delta^n}{n!}\, e^{-\delta} \tag{2.39b}$$

$$n = 0: \quad p_k(0) = e^{-\delta} \exp\left(\delta\; p_{k-1}(0)\right) \tag{2.39c}$$

$$p_1(0) = e^{-\delta} \tag{2.39d}$$

where n is the number of produced electrons and δ is the dynode gain.
Unfortunately, this is only valid if the gain δ is the same for all dynodes and cannot be used to calculate the distribution for real voltage dividers. It still is instructive though, as $\delta 1$ mainly governs the shape of the spectrum and later dynodes only add small changes, so the resulting function is not too far off.

In order to assess how well a measured charge spectrum can be described with a compound Poisson distribution, the present author developed an iterative model, which is more flexible. In this, the probability $p_j(n)$ to obtain n secondary electrons (SEs) after dynode j is the sum over the probabilities $p_{j-1}(i)$ to obtain i electrons at the previous dynode multiplied with the probability $f_i(n)$ that i electrons hitting dj produce n SEs. $f_i(n)$ was assumed to be a Poisson distribution with mean $\lambda = \delta\mathrm{j}$, which is folded with itself i times. This is summarized in the following equations:

$$p_1(n) = \frac{(\delta 1)^n}{n!}\, e^{-\delta 1} \tag{2.40a}$$

$$p_j(n) = \sum_{i=1}^{\infty} p_{j-1}(i) \cdot f_i(n) \tag{2.40b}$$

$$f_i(n) = \prod_{i}^{*} p_{\mathrm{SE},j}(n) \tag{2.40c}$$

$$p_{\mathrm{SE},j}(n) = \frac{(\delta \mathrm{j})^n}{n!}\, e^{-\delta \mathrm{j}} \tag{2.40d}$$

$\prod_{i}^{*}$ denotes a series of i convolutions of $p_{\mathrm{SE},j}$, which is the SE probability density function (PDF) at dynode j. $i = 0$ needs not be considered, since then no SEs hit the j-th dynode.

Based on this theory, the charge output was calculated for a Hamamatsu R5912 at +1425 V, corresponding to a gain $g = 1.3 \cdot 10^7$, which can be directly compared to measurements performed in this work. For the calculations, the known resistance ratios of the tapered voltage divider of Hamamatsu were used, as well as the curve shown in fig. 2.10 for Cs_3Sb, which was linearly extrapolated to zero, since according to the reduced yield curve the behavior should be quite linear in the lower voltage range (e.g. [267]). This procedure slightly underestimates δ for the last stages, if the true (concave) curve lies above the linear extrapolation.
Using these values, the total gain $g = \delta 1 \cdot \delta 2 \cdots \delta 10$ results in $1.42 \cdot 10^6$, in part due to the underestimation of δ for low impact energies, but also from the properties of material and surface differing between the data of Hamamatsu and the specific device. The difference factor to the measured value 9.10 was split up uniformly on all ten dynodes ($g' = g \cdot \eta^{10}$ with $\eta = \sqrt[10]{9.10}$), resulting in a δ correction factor η of 1.247 per dynode.

Using eqs. (2.40) the resulting secondary electron distributions (compound Poisson distributions) after d1–d4 were calculated. The repeated convolution is numerically costly, but as will be seen the shape of the distribution is determined mostly by the first dynodes, so this range is sufficient. As expected, the resulting distributions have a far larger charge resolution than the measured curve, which is due to d1 having a high-gain material in place of Cs_3Sb, as detailed on p. 69. Since the d1 material and secondary emission yield (SEY) curve is unknown, $\delta 1$ was varied while adapting the correcting factor for the other dynodes such that the total gain was correct, until the best match with data was found. This was for $\delta 1 = 15.00$, which corresponds to a correction factor for d1 of 1.688, reducing the correction factor for the other dynodes to 1.206.

The resulting dynode gains are:

$$\{\delta 1,\, \delta 2,\, \ldots\, \delta 10\} = \{15.00,\, 8.69,\, 9.46,\, 8.06,\, 4.89,\\ 3.06,\, 3.06,\, 3.06,\, 3.06,\, 3.06\} \tag{2.41}$$

The corresponding calculated dynode charge spectra are plotted for d1 to d4 in fig. 2.22a and b. Figure 2.22c shows the measured counterpart including the calculated distribution.

At d1 the discrete values of the Poisson distribution can still be seen in form of kinks in the function. The regions between the integral values are filled out by the subsequent amplification steps, and the distributions become increasingly dense; after three dynodes it is essentially continuous. The distribution is governed mainly by the d1 gain, subsequent dynodes only broaden and shift the charge spectrum slightly. This was already observed in [170] using eqs. (2.39). The shape converges rapidly, and the final shape is well approximated by the d2 distribution already, d3 is virtually indistinguishable. Note for d1 the curve has already almost Gaussian shape due to the high $\delta 1$. Generally a Poisson distribution can be decently approximated by a Gaussian for $\lambda > 10$; $\delta 1 = 16$ is above this value. This explains the frequent successful use of a Gaussian fit function.
The calculated spectrum is in very good agreement with data between 2.1 and 4.1 pC. Above, deviations occur due to the 2 pe peak which is centered around 4.8 pC; below, the underamplified pulses contribute to the spectrum. The data shows a tiny excess around 3 pC; this is likely *not* due to model mismatch, but caused by populations of overamplified pulses, which will be explained on p. 110.

In conclusion, the compound Poisson distribution gives a correct description of the SER if the dynode gains are Poissonian. No additional phenomenological shape parameters are required, but in principle could be added, such as an x-axis scaling. The individual δn can be either calculated ab initio, if the dynode materials and their SEY–energy functions are known, or can be used as fit parameters and retrieved from the fit.
The compound Poisson distribution is therefore clearly recommendable for fitting the SER. However, it is numerically costly. Since the d3 distribution already is a good match for the final SER, it can be used instead of the d4+ SER to save computation time without compromising the accuracy of the model notably.

In [295] an alternate formulation of a multi-stage Poisson distribution is given in form of a recursion formula, which was derived from a compound Pólya distribution using the special case $b = 0$ (see next section). Due to the recursive nature, this method is also numerically expensive. The authors of [295] stopped computation after d6 and also noted the rapid convergence of functions with dynode number.

Connection to scaled Poisson

The scaled spe distributions (mean and peak height set to 1) in fig. 2.22a are visibly broadened by inclusion of the stages after d1 (mostly d2). Thus, after d2 the compound Poisson distribution is not a Poissonian anymore: Since a Poisson distribution has a relative standard deviation (standard deviation divided by mean) of $1/\sqrt{\lambda}$, and the mean of the unscaled compound Poisson distributions is the product

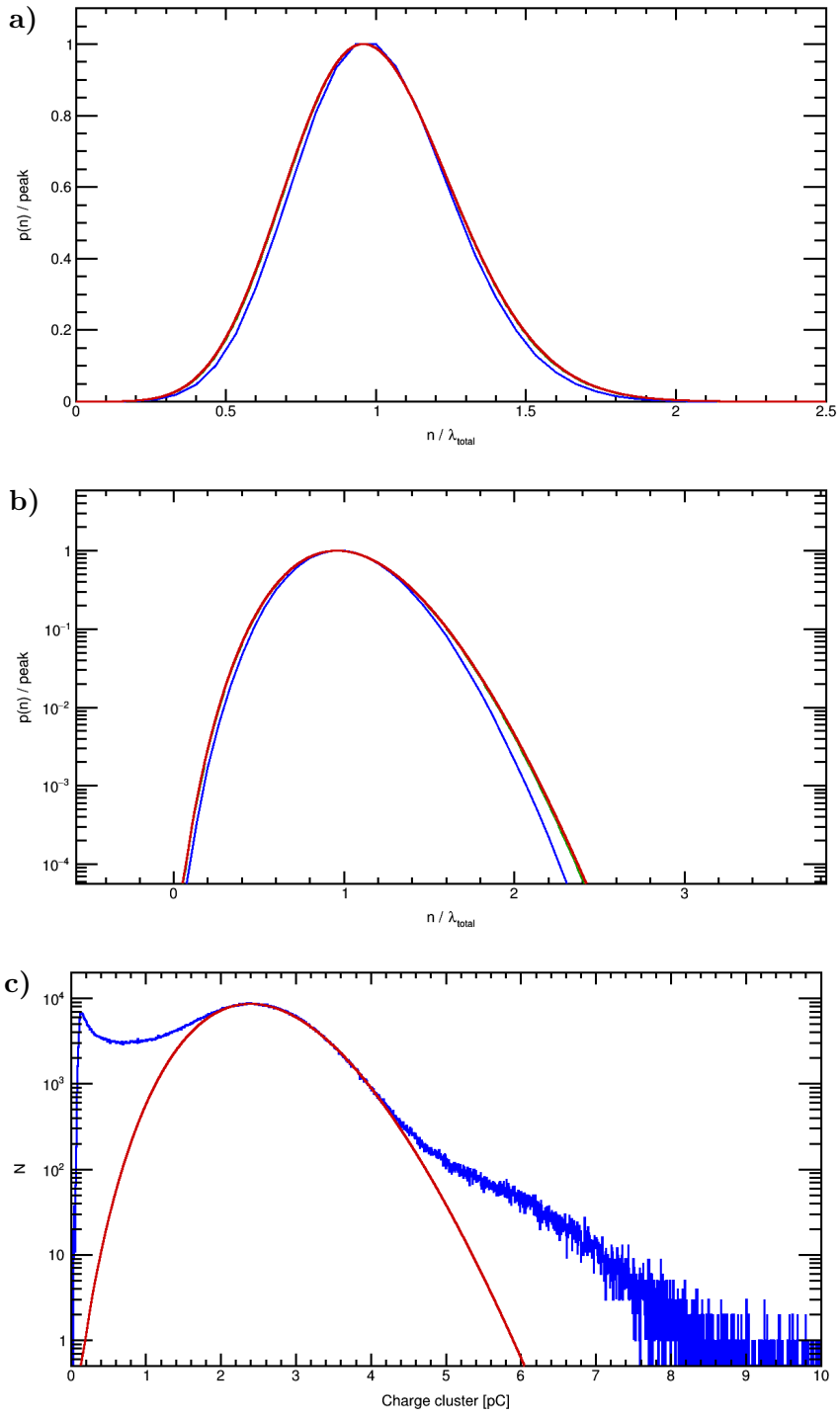


Figure 2.22: Compound Poisson distribution vs. measurements:
a) Calculated compound Poisson distributions after d1 to d4 (blue, green, orange, red) for a Hamamatsu R5912 at $g = 1.3 \cdot 10^7$, normalized to the resulting mean and peak height for comparison. The distributions quickly converge and cannot be distinguished after d3.
b) Semilogarithmic plot of the same distributions.
c) Measured charge spectrum (blue) for the same PMT at identical gain, $o = 2.9\%$, diffuse integral illumination, magnetically shielded, and in a Faraday cage. The calculated compound Poisson distribution after d4 (red) was scaled to the mean charge and number of events and is in perfect agreement with the spe events. Abscissa labeled "charge cluster" due to the presence of overlapping unrecognized early afterpulses (EAPs) in a classic measurement without pulse shape analysis (see 2.3.5).

of the mean stage gains, the scaled distributions would become more *narrow* after additional stages if they were Poissonian (after d10 a sharp Gaussian would result). While the shape of the compound Poisson distribution thus does not match that of a Poissonian with the same mean, it might be equivalent to a *scaled* Poisson, however. In this case one could use a slightly lower effective λ to approximate the broadening and then scale the charge axis to arrive at the mean gain after 10 dynodes. It would be interesting to mathematically derive or refute this conjecture, since if it is proven to be correct, a simple fit function instead of the complicated and numerically costly compound Poisson can be used to accurately describe the charge behavior. However, then the δn would constitute correlated fit parameters, which would make it hard to obtain the stage gains from the fit alone, and instead the method employed above should be used.
In practice it was found that a scaled Poisson is a reasonable approximation of the compound Poisson for all measurements presented in this work, so this far simpler fit was used to obtain the results listed in 3.2.2.

Pólya distribution and gain non-uniformities[59]

The Poisson and compound Poisson distributions assume that the gain processes at the dynodes are Poisson-distributed and uniform over each dynode. In the past this was a matter of much dispute as e.g. for venetian blind dynodes (see fig. 2.15d) the gain appears to not be uniform over the dynode surface (edge effect), and for many PMTs the SER could be better described by a Pólya distribution [170, 195, 296, 297]. In this, SE emission from a dynode is assumed with a mean SEY for the whole dynode of λ. For each point on the dynode the distribution of the number of secondary electrons is Poissonian, but the local mean SEY varies across the dynode as described by a Laplace distribution. This gain uniformity is described by a parameter b with $0 \leq b \leq 1$.
Also called negative binomial distribution, the Pólya distribution is a generalization of the Poisson distribution[60], which has the Poisson ($b = 0$) and geometric distributions ($b = 1$), i.e. a discrete negative exponential, as limiting cases. This explains why b acts as a shape parameter; the more uniform the gain is, the more Poissonian the distribution will be.

The Pólya distribution $p_b(n)$ of n emitted SEs for a single amplification step with mean gain δ is [170, 295]:

$$p_b(n) = \frac{\delta}{n!} (1 + b\,\delta)^{-n-1/b} \prod_{i=0}^{n-1} (1 + i\,b) \tag{2.42}$$

In [295] a recursion formula based on (2.42) can be found, which allows to compute the charge probability distributions after dn but is numerically very costly.
[276] gives a continuous function for the charge q, since for large numbers of multiplied electrons the Pólya distribution can be approximated by a gamma distribution:

$$p_b(q) = \frac{1}{b\,\delta\,\Gamma\left(\frac{1}{b}\right)} \left(\frac{q}{b\,\delta}\right)^{\frac{1}{b}-1} \exp\left(-\frac{q}{b\,\delta}\right) \tag{2.43}$$

[59] Content based on [170; pp. 222–224], [295]
[60] The Pólya distribution, in turn, is a special case of a compound Poisson distribution, which is a rather broad class of probability distributions [298].

For venetian blind PMTs, the effective gain varies over dynode surface, where effective means that actually the *collection efficiency* from the previous dynode varies due to shadowing, and skipping of dynodes can occur [170]. Statistically, this looks like a gain variation, but has different physical causes.
For Cs-activated dynodes, which are common in PMTs, the surface itself is rather uniform [170; p. 222], so no large deviations from $b = 0$ are expected.

Independent of the underlying phenomena, the Pólya distribution is a viable approach since it adds one more shape parameter and therefore is used in many current works with good results (e.g. [276]). Nevertheless, modern PMTs typically have rather Poissonian dynode amplifications, which is why the extra shape parameter b not always is needed and in many cases other distributions such as the compound Poisson, the single Poisson or even the Gaussian give better approximations of the SER – the Pólya distribution can only model the charge response for a limited number of PMTs [193, 299, 300], and e.g. in [261] a Pólya fit was far worse than a Monte Carlo generated compound Poisson function for the Hamamatsu R11780.

Wright states that the Pólya distribution should not be used in an attempt to fit the underamplified pulses which are caused mainly by backscattering and can not be modeled in this way [170; p. 224].
In the present work, the effects causing undersized pulses were studied in great detail (see 2.3.5, 2.3.6), which shows that the underamplified charge region is populated by a multitude of effects whose reduced charge stems from multiple reasons – e.g. skipped amplification stages, reduced impact energy, pulse branching, and (quasi-)particle generation, but also a reduced dynode gain due to varying hit positions and incident angles for the following reasons:

- For modern PMTs the focus region on d1 and the later dynodes for the majority of electrons is quite narrow, so potential gain variations over the surface of d1 (e.g. from varying material composition, surface structure or crystallinity) will not play a large role and will not affect the SER shape strongly. Thus the absence of positional gain fluctuations cannot be concluded with certainty from an SER fit.
- In addition, the SEY depends on incidence angle (see p. 71) and therefore the origin of the photoelectron, since d1 is not rotationally symmetric like the cathode (cf. the trajectories in fig. 2.16).
- Even if $\delta 1$ were completely uniform, the hit position could cause edge effect (see p. 84) due to a variation of CE_{12}.

The present author therefore presumes that gain variations on d1 *can* in fact occur for modern PMTs, although not necessarily with notable influence for the regular pulses. However, for the effects derived in 2.3.5 and 2.3.6 gain fluctuations should play a role and are included in the calculations.
This makes the Pólya distribution an important fit function for *individual* effects contributing to the undersized charge region, if they are isolated from other effects through cuts; in the SER a great number of distributions overlaps at small charges (confer figs. 2.45 and 2.47), which severely complicates fits of single effects.
Attempts of a phenomenological Pólya fit of the *complete* spe charge distribution including spe peak and underamplified events, however, likely will fail due to the large number of contributing effects with very different causes.

Comparison of single photon peak fit functions

In general, the charge distribution behavior varies from series to series, and which function fits best must be found through trial and error. It is recommended to use the compound Poisson and Pólya distributions first since they take into account more physical parameters than the other functions.
Whenever possible, the chosen charge distribution should be convolved with the normalized pedestal before fitting, since this constitutes the noise response of the electronics and as such is also present in the PMT output. The pedestal is best obtained with the PMT HV turned off but with otherwise unchanged electronics [276].

Another method not detailed here is by using a Markov diffusions approach (see [295] and the articles cited therein). [295] stresses the difficulties of this approach, that using a gamma distribution produces virtually identical results and is easier to compute, and that no stage gains or gain uniformities can be implemented. Therefore they employed a recursive Pólya distribution instead, and found that data was best matched for $b = 0$, i.e. purely Poissonian gain distributions.

Underamplified pulses

Causes

As can be seen from fig. 2.21 and 2.22c, there is a sizeable fraction of pulses having smaller charges than predicted from a model assuming Poissonian amplification (with gain non-uniformities) at each dynode, which is used to fit the spe peak with excellent agreement. [172; p. 2-9] notes that the fraction of undersized pulses $< 1/3$ pe typically amounts to 10–20%. [276] notes a reduction of the spe mean charge by 9% for the R5912 if the underamplified pulses are included.
Therefore, other gain paths must exist, many of which already were discussed in 2.1.3. Some possible effects are:

- **γ1:** A photon passes the cathode, and creates a pe on d1 which creates a pulse. The first amplification step is skipped, but apart from that the process is similar to a regular pulse and can be described by the same distribution with the mean charge lower by a factor of 10–20, potentially with adapted shape parameters.

- **E1:** Since an E1 is produced by a pe scattered elastically off d1 onto d2, the charge is expected to be reduced by the skipped gain from d1. The larger bombardment energy released at d2 can yield an increased gain; however, the SEY curve levels out for large energies and later even sinks, so this effect is not expected to be large and can even lead to a *decreased* gain. Altogether it seems reasonable to assume a Poisson gain distribution at d2 and all subsequent dynodes, so the same models as described for the main peak can be used with a smaller total gain. This will produce an spe-like peak with lower mean and somewhat different shape parameters.

- **Inelastic backscattering on d1:** After scattering with SE production (BE1, BL1), either the pe is lost or produces a pulse which is separated in the charge integration window from the charge-reduced PP, which is analyzed in the following. Therefore, the charge integral has a lower value and can range anywhere from 0 to 1 pe plus statistical spread from later dynodes.

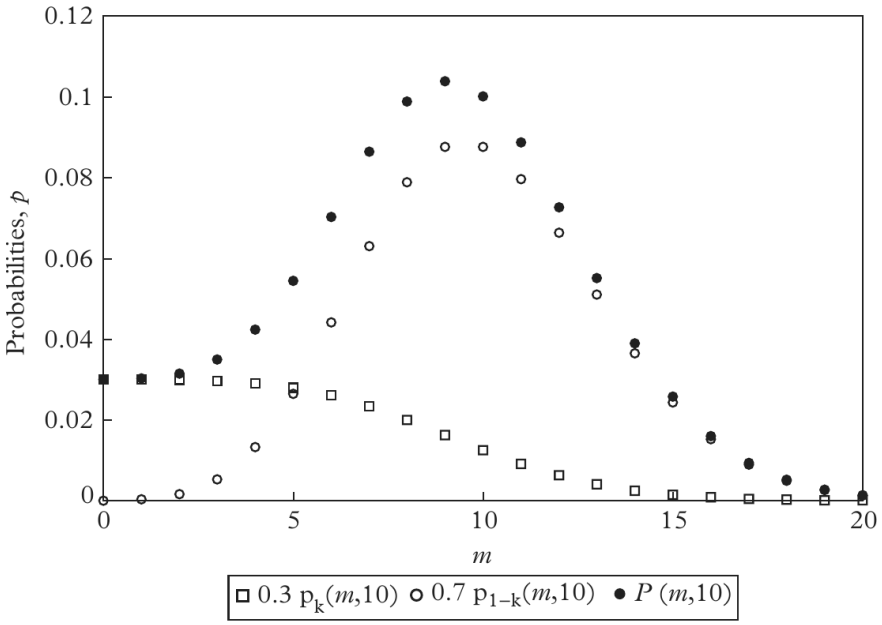


Figure 2.23: Calculated probability distributions of secondary electrons produced by inelastic scattering on d1 (eq. (2.44), squares), normal secondary electron (SE) production on d1 (Poisson distribution, circles) and the sum of both distributions (filled circles), from [170; p. 220]. $\lambda = 10$ and a fraction of underamplified pulses $k = 0.3$ were assumed.

In [170; p. 219] a formula is derived for the charge distribution of the primary pulse produced by inelastic scattering, assuming that the energy distribution is flat (see fig. 2.12a), and that the SE production follows a Poisson distribution with reduced mean $\lambda_r = \lambda(1 - x)$, where x is the fraction of primary energy carried away from SE production by the scattered primary. Integrating over the flat energy spectrum yields the SE distribution $p_i(n)$ for n emitted SEs:

$$p_i(n) = \int_0^1 e^{-\lambda(1-x)} \frac{\lambda^n (1-x)^n}{n!}\, dx \tag{2.44a}$$

$$n = 0: \quad p_i(0) = \frac{1 - e^{-\lambda}}{\lambda} \tag{2.44b}$$

$$n = 1: \quad p_i(1) = \frac{1 - e^{-\lambda}}{\lambda} - e^{-\lambda} \tag{2.44c}$$

$$n = 2: \quad p_i(2) = \frac{1 - e^{-\lambda}}{\lambda} - e^{-\lambda} - e^{-\lambda}\frac{\lambda}{2!} \tag{2.44d}$$

This is a monotonically decreasing function (fig. 2.23).
In order to obtain the charge distribution after the last dynode, the Poisson-based amplification steps of the following dynodes have to be applied to (2.44), as detailed before. This will broaden the spectrum slightly.
The spe peak distribution and the underamplified pulses are connected by the fraction k of photoelectrons which undergo inelastic scattering; if the total spe distribution is normalized, k is a prefactor for the underamplified pulses, given by (2.44), whereas $(1 - k)$ is a prefactor for the regular pulses.
Unfortunately, (2.44) does not take into account all effects. The SEY curve is not linear and shows reduced yields for higher energies for usually employed

dynode materials. Strictly speaking, the scattered pe energy distribution is not flat either, but this should constitute only a secondary effect. Furthermore, some scattered photoelectrons hit d2 or d1 (BE1, BL1) which produces a second pulse which will increase the total charge if the integration window covers both pulses. Regarding BE1, it must be taken into account that scattered photoelectrons with higher energies are less likely to be focused onto d2.
The higher "density of states" if the curve SEY(E) is projected onto the y-axis in fig. 2.10 should cause a maximum to appear toward the 1 pe peak instead of a monotonically decreasing function. This is further heightened by the charge contribution from BE1 which are more likely to hit d2 for lower energies, i.e. higher primary pulse charges. The energy dependence of BL1 collection probability is not known, but around even energy splitting between BL1-pulse and PP the total SEs produced on d1 should reach a maximum due to SEY(E), which would further contribute to a peak at larger charges.
Summarizing all findings, the charge distribution for inelastically scattered photoelectrons on d1 is expected to follow a curve similar to (2.44) shown in fig. 2.23, but with a peak in the upper half of the distribution and smeared by subsequent dynodes.

- **Loss processes in d1:** The pe can undergo loss processes not producing any SE in d1, such as phonon or plasmon scattering or Bremsstrahlung generation. Phonon scattering causes $\mathcal{O}$(meV) energy losses and is Poissonian, but if other effects are non-Poissonian or withdraw large energies from the SE production (such as characteristic X-rays, Auger electrons, or bremsstrahlung; the latter can radiate any energy from 0 to the total primary energy [170; p. 224]), the shape of the spe distribution may be changed or additional distributions at lower charges may arise. No evidence for a shape change was seen in the fits shown previously, so it might be concluded that the processes without large energy losses behave Poissonian.

- **Scattering on later dynodes:** These effects are introduced in 2.3.5.1 and comprise elastic and inelastic scattering as well as loss processes, occurring for one or more *secondary electrons* on d2 or later. This can cause EP, LP which can reduce charge, albeit not as much as on d1, since only an SE and not the pe is affected. Additional[61] charge reductions of up to $1/\delta 1$ ($\approx$ 5–10%) are expected; if such effects occur for several SE, the charge can be further reduced.

- **Additional time shift effects (TSEs):** Beyond the effects modifying the transit time mentioned so far, a plethora of other interactions can occur, which are detailed in 2.3.5, 2.3.6 and 2.5.2.2, and would be too numerous to repeat here. On p. 301 the contributions to underamplified pulses from TSEs are reviewed.

- **Dark counts:** Not pertaining to the spe distribution, but inevitably contained in every charge spectrum are contributions from dark count pulses (fig. 2.24). These consist mostly of thermionic electron emission from the cathode and the dynodes. The former produces a charge distribution virtually identical to the

[61] The effects can occur for underamplified pulse types, too.

spe spectrum[62], so it just contributes to the 1 pe distribution. Emission from the dynodes creates far smaller charges due to the missing amplification step(s) (factors $\delta 1$, $\delta 1 \cdot \delta 2$, ...) and constitutes a large fraction of all dark count events. Therefore the mean dark count charge is *smaller* than 1 pe. Contributions are expected from every dynode, but after d1 or d2 the distributions are completely covered by the pedestal – they are visible in the dark current, however (see p. 349). All dark count distributions should follow a Poisson-based distribution[63], with reduced mean for dynode emission.

- **Interdynode ionic afterpulses:** Even if large delays between incident light pulses are used in order to suppress APs leaking into the next acquisition window, there will always be a contribution of APs caused by dark count pulses present in the spe spectrum. Afterpulses can also originate from ionizations inside the dynode chain (see 2.5.2.3), where the ions are accelerated onto the previous dynode (interdynode ionic afterpulse, iiAP). The produced number of SEs and thus the pulse size depends on molecular mass, charge and drift distance just as for APs hitting the photocathode. Similar to dark count pulses originating on the dynodes one or more amplification steps are skipped, in addition the traversed potential difference is smaller than between d1 and k, all of which reduce the mean charge accordingly. Cathode iAPs can have charges exceeding 15 pe, and typically several ionic species are observed in the AP delay distribution, therefore multiple interdynode AP populations with charges below 1 pe and Poisson-based distributions are expected.

Fit function

An ideal fit function of the underamplified pulses would consist of multiple functions describing the various Poisson-based and non-Poissonian contributions. $\gamma 1$, E1 and dynode dark counts all have very small charges (mean $\lesssim 0.15$ pe). L1 and scattering on d2+ have charges not much lower than 1 pe ($\gtrsim 0.9$ pe). This also explains why fitting the spe peak alone can only be performed reasonably well if the lower border is closely below the peak. Finally, inelastic backscattering, time shift effects, loss processes and iiAPs can occur in the complete range from 0 to 1 pe.

Often, this conglomerate of effects can be described phenomenologically by an exponential decay [226, 276, 289]. The complete spe distribution is then the sum of the exponential term and the spe peak function and should be convolved with the noise distribution of the setup (the normalized pedestal).
[276] uses two fit functions, one continuous Pólya distribution with reduced gain to describe e.g. E1, and an exponential function to model suboptimal amplification at d1 due to field inhomogeneities and d1 asymmetry [301]. The present author notes that E1 should have far smaller charges than obtained from the best fit values, and that the fit supporting the phenomenological model of the exponential decay in [301] is tenuous. Albeit more refined, this is a phenomenological model.
[170] uses a different approach, as outlined on p. 106, where the effects from inelastic scattering are partially included. For data shown in [170; p. 221] this model matches

[62] The emission energy of the electrons can differ slightly.
[63] Any of the aforementioned spe fit functions can be seen as derived from Poisson distributions, be it as a special case, a concatenation, or by using a generalized Poisson distribution with more shape parameters.

the curve shape well, but cannot reproduce the complete magnitude of underamplified pulses, as it does not take into account all of the effects listed above. The present author, however, considers it superior to a simple phenomenological exponential fit, as it uses a physical model. If Wright's model is refined with regard to the mentioned additional inelastic effects and subtracted from the SER, the other undersized contributions should be better visible, which might allow to develop physical fit functions for them, too.

Overamplified pulses

Some of the effects described in this thesis and discussed in 2.3.5, 2.3.6 and 2.5.2.2 can enlarge the charge of a single photon pulse: If a photon produced by e.g. bremsstrahlung, cathodoluminescence or photoluminescence in the dynode stack (see 2.6.1) hits the cathode, another dynode or the same dynode, this can produce a secondary pulse (photonic afterpulse (γAP)) with delays ranging from a few tens of ps to the transit time plus a few ns (see also [264]). Even negative delays (earliness) of a few ns can occur, if a succeeding dynode is hit.
If the additional pulse falls within the charge integration window, this increases the collected charge. For a cathode hit effectively two photons produce pulses which feeds the 2 pe distribution. If a dynode is hit – this can be a previous, succeeding, or the same dynode – the resulting pulse has a charge according to the remaining amplification steps; about 5–10% for d1, around 1% for d2, 0.1–0.2% for d3, and negligible beyond. The same dynode can be hit either directly or after reflection on surrounding structures; in both cases the delay is extremely short, so the secondary pulse should merge with the primary pulse and essentially behave as if an additional, slightly delayed SE had been produced. If several γAPs are created, the charge is increased accordingly (compare 2.5.2.2).

Probability

The probability of photon production rises with the number of electrons in the cascade so it is largest at the anode. On the other hand, the probability for one of these photons to reach the cathode decreases with dynode number. It should be most likely for a photon to hit the previous or next dynode due to solid angle, reflection losses and obstacles, and increasingly less likely to hit dynodes further away, with the possible exception of the cathode due to its far larger area (discussed in 2.3.5.3). This can be compensated by the far larger number of trials for later stages.
The manufactures have known about anode glow for many decades and employ light shields around the later dynodes and in front of d1 (fig. 2.16) to prevent photons from leaving the dynode stack through slits and being reflected off the glass envelope onto the cathode. This shielding is not perfect though and cannot be used inside the dynode chain, where for the box-and-line dynodes (fig. 2.15) often used in astroparticle physics (APP) the geometry favors reflections onto previous dynodes or the cathode.
Measurements performed here for a box-and line R5912 (see p. 380) showed γAP rates of $> 0.3\%$ (3.7% predicted by time shift theory) of γnk (photon from dn hitting k), and a far larger value of 13.6% (13.2% predicted) for γn1 (dn photon hitting d1), both of which mainly originated from intermediate dynodes (see 2.5.2.2). The higher probability for γn1 might be due to the large area of d1, its high gain, which likely also leads to an improved quantum efficiency (QE) compared to other dynodes, and

that it is "operated" in reflective mode – the bremsstrahlung spectrum is strongest in the blue where the QE enhancement of a reflective cathode versus a transmissive one should prove effective (see p. 65). Photonic APs from d2 (γn2) unfortunately could not be unequivocally resolved due to their small charges, but a rate of 2.8% is predicted despite the smaller size and reduced QE of d2. The γAP rate is expected to quickly rise for later dynodes due to the high number of electrons in the cascade, so γAPs are part of every pulse and contribute to the trailing flank of the anode pulse shape (see 2.4).
According to the high measured γAP rates, frequently two or more γAPs will be observed simultaneously.

Contributions to the charge distribution

Therefore, γAPs constitute multiple additional delayed gain channels, which depend on dynode stack geometry, contribute to the gain at each dynode and change the spe peak distribution depending on charge window. The shape of the charge spectrum should be affected most by these additional gain channels at the first dynodes (see fig. 2.22a), furthermore these processes are likely non-Poissonian. The changes to the distribution will not be extremely large as the rising probability is offset by the dropping secondary pulse charge.
For the R5912 at d1 an additional charge of about $1/15.0\,\mathrm{pe} \cdot 13.6\% = 9.1 \cdot 10^{-3}\,\mathrm{pe}$ results, the cathode contributes with $> 2.9 \cdot 10^{-3}\,\mathrm{pe}$ in form of 2 pe pulses. The other dynodes should have comparable contributions, so these effects could add several percent of charge to the spe peak. The dependency on dynode geometry, material and integration time could possibly in part explain the variability of the applicability of spe peak fit functions observed in literature.
The contribution to the 2 pe peak can lead to deviations from the expected Poisson distribution for n pe populations. If such differences occur, a convolution of the Poisson n pe distributions with a Poisson distribution modeling the γAP probability with $\lambda_{\gamma\mathrm{AP}}$ as free fit parameter might help resolve the discrepancies.

Multi-photon distributions

The 2 pe distribution consists of events with two detected photons within the acquisition window, where both photons have a charge according to the spe spectrum. The sum of two independent random draws from the spe charge distribution mathematically is equivalent to a convolution of the spe function with itself. The n pe distributions can thus be obtained by consecutive convolutions. If these functions are used for a fit, this should be followed by *one* convolution with the pedestal for all n pe curves.
Due to the convolution the n pe distributions are broader than the spe curve. It can be shown (e.g. by using the convolution theorem and Fourier transformation [302]) that the sum of two Gaussian distributed random variables is also a Gaussian whose standard deviation σ observes $\sigma^2 = \sigma_1^2 + \sigma_2^2$ and with mean $\mu = \mu_1 + \mu_2$ (see p. 124). From this follows that for the n pe peak, assuming linear response, the mean μ_n and standard deviation σ_n read

$$\mu_n = n\,\mathrm{pe} \tag{2.45}$$
$$\sigma_n = \sqrt{n} \cdot \sigma_{\mathrm{spe}}. \tag{2.46}$$

While the mean scales linearly with n, the standard deviation grows broader with an increasing number of photons. Of course, this is only valid, if the spe distribution can be described by a normal distribution. Otherwise, it is only an approximation, but still demonstrates the broadening for a higher number of detected photons. This is also the reason why even for photosensors with extremely good spe resolutions such as SiPMs (see p. 628) at a large enough n adjacent peaks will start to overlap.

Pedestal

The pedestal is caused by the noise produced and picked up by the setup. This can usually be described by a Gaussian, but if strong artificial noise signals such as bursts are present, additional components extending beyond the normal distribution can appear. Needless to say, all noise sources should be eliminated as far as possible, as they cover underamplified pulses. This increases the systematic error of gain determination and requires higher hardware thresholds to avoid fake triggers which reduces the photodetection efficiency (PDE).

Dark count pulses[64]

Many dark count pulses are generated by thermionic emission from the cathode which produces pulses basically identical to a photon pulse, including possible underamplification (see p. 108). However, there is also a large contribution of small pulses from thermal emission from the dynodes, which is why the mean dark count charge is smaller than for spe events and the dark count charge distribution looks differently from the SER (fig. 2.24).
This knowledge can be used to suppress the background from dark counts and raise the signal-to-noise ratio (S/N) by choosing a high threshold [280]. For the purposes of a detector, however, the PDE must be maximized, so a low threshold must be set at the expense of a lower S/N.
Normally, the occupancy is set comparatively high (cf. (2.24)) to reduce the contribution from DN. Where this is not possible (or additionally), background subtraction can be performed, as it is in many other disciplines.

Seeing as an spe light source is quasi built in the PMT, it would be an elegant, tempting way to obtain the SER by simply measuring the dark count charge distribution. Alas, due to the large contribution of dynode dark counts (which in addition varies with HV) and high charges from iAPs and radioactivity this is not possible without introducing inaccuracies.
These sources of large charge dark counts will be discussed in the following.

Gamma rays

One contribution to the dark count spectrum – i.e. the PMT response in the absence of a light signal – is from gamma rays from either intrinsic or external radioactivity, which hit the PMT window or envelope. There they generate electrons by photoelectric effect, Compton scattering and pair production. In glass Compton scattering dominates compared to the photoelectric effect. Pair production can only occur for gamma energies above 1.022 MeV. The electrons in turn can produce light via ionization and excitation processes in the material and for high enough

[64] Content based on [170; pp. 259–280] [169]

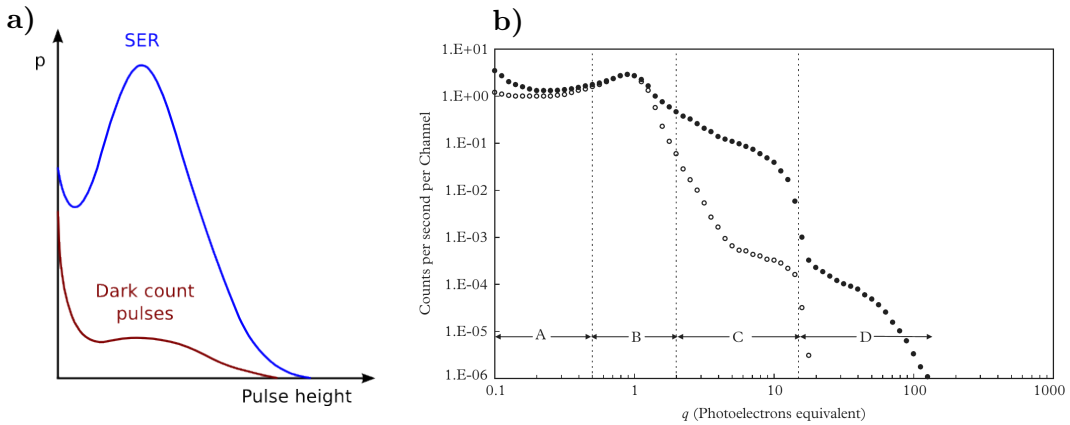


Figure 2.24: Dark count pulse height distribution (PHD) compared to the spe PHD (also called SER). PHDs are virtually identical to charge distributions inside the linear range. Dark counts are contained in the SER as well, but since the distribution is filled using a coincidence of light source trigger and PMT pulse within the acquisition window, they are suppressed compared to photon pulses.
a) Contribution of dark counts in the same recording window, linear abscissa and ordinate (probability), based on [184; p. 128]. Note the larger fraction of underamplified pulses.
b) Normalized to peak height, logarithmic abscissa and ordinate, dark counts filled circles, SER empty circles, for a fast 2" PMT. From [170; p. 260]. The relative contributions of thermionic emission from the dynodes (region A), ionic afterpulses (region C) and radioactivity (region D) are far higher in the dark spectrum.

kinetic energies through Cherenkov radiation. These photons can hit the cathode and generate photoelectrons and in turn pulses. Thus gamma rays cause pulses through indirect light production.
To quantify this phenomenon, the scintillation efficiency, defined as the fraction of incident particle energy converted into photons, can be used. This was measured by [303] and [170] for various window materials. Borosilicate glass produces 2 photons/MeV and has a scintillation efficiency of about 0.02–0.023%, which is very inefficient compared to plastic scintillators (3%).
Gamma rays with energy <175 keV cannot emit Cherenkov light in PMT glass and only produce photons through ionization and excitation with the aforementioned efficiency.
The gamma ray sources relevant for PMTs are mainly from naturally occurring ^{40}K, ^{232}Th and ^{238}U, and artificial sources. No gamma rays from natural sources were observed above 2.62MeV (Tl) by Wright [170].
The charge distribution measured by [170] for ^{57}Co (122 keV) is mostly identical to the SER, for ^{137}Cs (0.66 MeV) it extends up to 12 pe, and for ^{60}Co (1.17 and 1.33 MeV) up to 25 pe. The maximum expected charge can be calculated from the energy and the scintillation efficiency of 0.023%, which Wright obtained from these measurements, by assuming a photon energy of 3 eV, where the QE is still high, and a PDE of 20%. The obtained estimates are in very good agreement with all observations. According to this formula, for 2.6 MeV 48 pe are expected in borosilicate glass which sets the upper limit for the gamma ray contributions to the dark count charge spectrum.
The background rate due to gamma rays is essentially proportional to the window volume, since the solid angle and shielding reduce the contributions from the rest of the envelope.

Cosmic rays

An additional contribution comes from the cosmic ray flux. At sea level this is composed of 63% muons, 21% neutrons, 15% electrons and < 1% protons and pions. The dominant muon flux amounts to about $^1/_{\mathrm{cm}^2 \cdot \mathrm{min}}$ ($\approx {}^1/_4\,\mathrm{Hz}$ for a 2" PMT [171; p. 15]) and has an angular distribution of $\approx \cos^2(\theta)$.
Apart from ionization and excitation, muons produce Cherenkov light in the PMT window. The muon event rate is proportional to the cathode area and the produced charge is proportional to the window thickness. Since the Cherenkov angle is 48° in the glass ($\theta = \arccos({}^1/_{n\beta}) \approx \arccos({}^1/_n)$), but the total internal reflection angle of the glass–air interface is 42° ($\theta_c = \arcsin({}^{n_2}/_{n_1}) = \arcsin({}^1/_n)$), all produced photons are trapped if the PMT window faces in direction of the muon path, which results in a strongly enhanced QE – Wright observed an increase in charge by a factor of 2 for upward versus downward facing PMTs [170; pp. 277–279]. Therefore for muon events the charge depends on PMT orientation. For upward facing cathode the acceptance area is maximal, if it is facing down 25% less light is collected since the Cherenkov cone points away from cathode. If the cathode faces the horizon, it has the lowest cross-section, but the longest muon path lengths inside the window, which can produce very large pulses – up to 500 pe were observed for an ETE[65] 9390 (5" flat window) PMT.
Since the Cherenkov emission is biased towards the UV, the collected charge also depends on the window material; quartz windows have the largest charges per event [171; p. 15].

Regions in the dark count charge spectrum

Coming back to the dark count charge spectrum, in fig. 2.24b in region A an excess of undersized pulses is visible compared to the SER. This is mostly due to dynode dark counts, but also caused by iiAPs and light produced in the glass by ionizing and Cherenkov radiation, which is then collected by d1. [170; pp. 267–269] has performed measurements where first the cathode and then both cathode and d1 were open circuited (infinite resistance). The former shows that the excess of undersized pulses constitutes of pulses originating from d1, where the ratio of the peak position to the spe peak matches $\delta 1$. The latter demonstrates that pulses from d2 produce a peak with relative position to the d1 peak fitting $\delta 2$ and far higher density per bin. If these open circuit measurements are subtracted from the SER, still about 2/3 of the excess remains. Wright assumed this to be from inelastic scattering, but as shown above, there are more effects contributing to this region. In any case, only a fraction of the undersized events is temperature dependent, so there must be other contributions apart from thermionic emission.

Region B consists of spe background originating from the cathode. This can be through thermionic emission from the cathode, photoemission through ionization and excitation inside the window (e.g. low energy γ), or by a photon (γAP, Cherenkov light) or ion (iAP) which hits the cathode and produces a single pe. Only a fraction of events in the spe range is temperature sensitive.

The multi-pe events in zone C are from iAPs produced by cathode dark counts, but also from radioactivity (γ, β) in the PMT glass and dynode stack, as well as from outside.

[65] Electron Tubes Enterprises Ltd.

Finally, the very large charges observed in region D (15–200+ pe) stem from gammas and cosmic rays. Charge and event rate depend on window material, thickness, area and photocathode type [170, 171]. In these events an initial big pulse is followed by a series of up to 50 spe within 100 µs.

Ionizing radiation

Ionizing radiation can be considered part of the dark count charge spectrum as it is independent of the signal. This was covered in the last section starting on p. 112.

Afterpulses

Afterpulses are in part caused by photon-induced pulses and correlated to the signal, but also are produced by dark count pulses as a part of the dark count charge distribution. Please refer to p. 109 and 2.5.2.

Charge distribution properties

Regardless of the constituents of the SER spectrum, there are several properties by which the SER distribution can be characterized concerning its performance with regard to photon counting.
Before these properties can be addressed properly, however, some introductory remarks on gain statistics are helpful.
In the following the mean gain shall be denoted by g, its variance by σ_g^2 and its relative variance by v_g^2 with

$$v_g^2 = \frac{\sigma_g^2}{g^2} \tag{2.47}$$

Excess noise factor [66]

Photoelectron emission obeys Poisson statistics, so with λ_{pe} the mean number of photoelectrons and σ_{pe}^2 the variance, the relative variance v_{pe}^2 is

$$v_{\mathrm{pe}}^2 = \frac{\sigma_{\mathrm{pe}}^2}{\lambda_{\mathrm{pe}}^2} = \frac{1}{\lambda_{\mathrm{pe}}}. \tag{2.48}$$

In the subsequent cascade the photoelectrons are multiplied with a statistically fluctuating gain. Since pe generation and multiplication are independent processes, the mean number of electrons in the output anode pulse n_a is

$$n_a = \lambda_{\mathrm{pe}} \cdot g. \tag{2.49}$$

The Laplace theory of probability generating functions shows that the relative variance of the output v_a^2 equals the sum of the relative variances of each stage where each is divided by the mean from all previous stages. Photoelectron emission has no

[66] Content based on [267] – an excellent read.

preceding stage and thus is divided by 1. Using eqs. (2.47), (2.48) and (2.49) this leads to:

$$v_a^2 = \frac{\sigma_a^2}{n_a^2} = v_{\mathrm{pe}}^2 + \frac{v_g^2}{\lambda_{\mathrm{pe}}} = \frac{1}{\lambda_{\mathrm{pe}}}\left(1 + v_g^2\right) \tag{2.50}$$

$$\begin{aligned} \rightarrow \quad \sigma_a^2 &= n_a^2 \cdot v_a^2 = \lambda_{\mathrm{pe}}^2 \cdot g^2 \cdot v_a^2 = \lambda_{\mathrm{pe}} \cdot g^2 \cdot \left(1 + v_g^2\right) = \sigma_{\mathrm{pe}}^2 \cdot g^2 \left(1 + v_g^2\right) \\ &= \sigma_{\mathrm{pe}}^2 \cdot g^2 \cdot F \end{aligned} \tag{2.51}$$

$$\text{where} \quad F = 1 + v_g^2 = 1 + \frac{\sigma_g^2}{g^2} > 1 \tag{2.52}$$

From this follows that the PMT amplifies the input variance σ_{pe}^2 by the squared gain g^2 like an ideal amplifier does, but further increases it by the excess noise factor F (ENF, also called F-factor, F^2, or noise figure [244, 267]) due to gain fluctuations, where F is always > 1.
The ENF (also called noise factor or noise figure) is defined as the ratio of the squared signal-to-noise ratios (S/N) of input and output [170] (see (2.70))

$$F = \frac{\left(\mathrm{S/N_i}\right)^2}{\left(\mathrm{S/N_o}\right)^2} \overset{\mathrm{PMT}}{=} \frac{1/\sigma_{\mathrm{pe}}^2}{g^2/\sigma_a^2}$$

Here the input is governed by the Poisson statistics of photon hits.
F is ≤ 2 for most PMT types and close to 1 for high quality series. The ENF can therefore often be neglected.

In the derivation of (2.51) it was implicitly assumed that every pe is amplified and no losses due to CE and ME occur. If Poisson statistics are assumed for these processes as well, instead the mean number of anode pulses λ and the relative variance of the photoelectrons which successfully produce pulses $v_p^2 = 1/\lambda$ should be used, which results in

$$\sigma_a^2 \approx \sigma_p^2 \cdot g^2 \cdot F \tag{2.53}$$

$$v_a^2 \approx v_p^2 + \frac{v_g^2}{\lambda} = \frac{F}{\lambda} \tag{2.54}$$

Poisson statistics are *not* completely valid for these effects, however, for which reason the equations were marked as approximations, since measurements will show larger relative variances.

Contribution from the dynodes

The contributions of the dynodes, v_g^2, to eq. (2.54) shall now be examined closer.
The n-th dynode has a mean gain δn, a gain variance σ_n, and a relative variance $v_n^2 = \sigma_n^2/(\delta \mathrm{n})^2$. From the Laplace theory of PGFs follows:

$$v_g^2 = \frac{\sigma_g^2}{g^2} = v_1^2 + \frac{v_2^2}{\delta 1} + \frac{v_3^2}{\delta 1 \cdot \delta 2} + \ldots + \frac{v_n^2}{\delta 1 \cdot \delta 2 \cdots \delta(\mathrm{n}-1)}$$

This shows that d1 has a key importance for reducing the gain fluctuations, as v_1^2 contributes with its full weight and $\delta 1$ reduces the weight of all following terms. This justifies a higher voltage between k and d1 and using high-gain materials or special

treatment on d1. If one assumes that dynode gain can be described by Poisson statistics, it follows:

$$v_g^2 = \frac{1}{\delta 1} + \frac{1}{\delta 1 \cdot \delta 2} + \ldots + \frac{1}{\delta 1 \cdot \delta 2 \cdots \delta \mathrm{n}} \tag{2.55}$$

Assuming an equal gain δ on all dynodes except d1, this becomes

$$v_g^2 \approx \frac{1}{\delta 1}\left(1 + \frac{1}{\delta} + \frac{1}{\delta^2} + \ldots + \frac{1}{\delta^{n-1}}\right) = \frac{1}{\delta 1}\frac{1-(1/\delta)^n}{1-1/\delta} \approx \frac{1}{\delta 1}\frac{\delta}{\delta - 1} \tag{2.56}$$

In the derivation of eq. (2.55) and (2.56) the ideal case of Poisson multiplication on all dynodes was assumed. For real PMTs the multiplier noise always exceeds this and the true ENF, measured from a charge distribution, will be larger than $1 + v_g^2$.

If $\delta = 2$ and varying values for $\delta 1$ are plugged in (2.56) using (2.52), F results to 1.5 ($\delta 1 = 4$), 1.2 (10) and 1.1 (20). This illustrates that with a high $\delta 1$ a remarkable reduction of the ENF can be achieved even for modest values of δ.
If all dynode gains are assumed to be equal ($\delta 1 = \delta$), and a δ of 2 is taken ($\rightarrow g = 1{\cdot}10^3$ for 10 dynodes), $F = 2$. This can serve as a worst-case scenario and shows that modern large-gain PMTs with high-gain d1 will have a far lower ENF.
To give an example, for the R5912 a very good ENF of 1.075 results when the dynode gains from (2.41) are plugged into (2.55), thanks to the large $\delta 1 = 15.0$.

To achieve a small ENF, besides a large mean $\delta 1$ the uniformity of $\delta 1$, a good gain stability, high linearity as well as low AP rates are important [197].

Relative charge variance

The relative variance of the charge distribution is defined as

$$v_q^2 = \frac{\sigma_q^2}{\overline{q}^2}$$

where $\overline{q}$ is the mean charge and σ_q is the standard deviation of the charge distribution. The output charge q is simply the gain g times the elemental charge[67], so (2.50) and (2.51) can be directly applied. This is not sufficient to describe the measured relative charge (or gain) variance, however.

In the ideal case for a perfect photosensor, the photon flux arriving at the sensor follows Poisson statistics with a mean number of incident photons λ_{ph} and a variance $\sigma_{\mathrm{ph}}^2 = \lambda_{\mathrm{ph}}$. From this follows the relative variance to

$$\frac{\sigma_q^2}{\overline{q}^2} = \frac{\lambda_{\mathrm{ph}}}{\lambda_{\mathrm{ph}}^2} = \frac{1}{\lambda_{\mathrm{ph}}} \tag{2.57}$$

It can be shown using cascade theory that a Poisson distributed light flux with mean and variance λ_{ph} results in a Poisson distributed pe flux after conversion in the cathode [170; p. 154]. The pe conversion itself is governed by the binomial distribution where the mean is the QE and the variance is $\mathrm{QE} \cdot (1 - \mathrm{QE})$.

[67] This holds even for non-linearity as it can be considered a gain change.

Consequently, in a still rather ideal case, the sensor has a PDE $\neq 1$ and the pulses are produced by a Poisson process with

$$\lambda \approx \lambda_{\text{ph}} \cdot \text{PDE} \approx \lambda_{\text{ph}} \cdot \text{QE} \cdot \text{CE} \cdot \text{ME}. \tag{2.58}$$

$$\rightarrow \quad \frac{\sigma_q^2}{\overline{q}^2} \approx \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} \tag{2.59}$$

In this it was assumed that the PDE produces a Poisson distribution, which is valid for the QE, but not for CE and ME. Thus, the equations are only valid as approximations.

The same relation can be derived accounting for the gain variations described by the ENF. In doing so, the usual label $\sigma^2/E^2 \equiv v_q^2$ (relative energy variance) shall be introduced. The energy E denotes here simply the number of collected secondary carriers, i.e. the output charge produced by incident photons. Using (2.53) and (2.58),

$$\begin{aligned} & E \approx \lambda_{\text{ph}} \cdot \text{PDE} \cdot g \\ & \sigma \approx \sigma_p \cdot g \cdot \sqrt{F} \\ \rightarrow \quad & \frac{\sigma^2}{E^2} \approx \frac{\sigma_p^2 \cdot g^2 \cdot F}{\lambda_{\text{ph}}^2 \cdot \text{PDE}^2 \cdot g^2} = \frac{\sigma_p^2 \cdot F}{\lambda_{\text{ph}}^2 \, \text{PDE}^2} \approx \frac{\lambda \cdot F}{\lambda^2} = \frac{F}{\lambda} \end{aligned} \tag{2.60}$$

just as in (2.54).

Additional corrections are required to take into account the background rate from dark counts and photon background (e.g. stray light), which is described by its mean contribution λ_b, and the readout noise (shot noise) of the electronics characterized by its equivalent noise charge (ENC).
Combining the effects listed in [170, 197, 244, 267, 285] the following formula can be derived:

$$\frac{\sigma^2}{E^2} \approx \frac{F\,(\lambda + \lambda_b) + \left(\frac{\text{ENC}}{g}\right)^2}{\lambda^2} \tag{2.61}$$

The target of both photosensor manufacturer and experimentalist using the device is to minimize the relative charge variance, as this results in a better resolution (see p. 119). According to (2.58) the PDE is a constituent of λ, so for a λ_{ph} given by the experimental setup, λ can be increased by raising the PDE. Since the (2.61) scales roughly with $1/\lambda$, a large PDE reduces the relative charge variance. If the gain is high compared to the ENC, the last term in the numerator becomes negligible. The remaining terms then scales directly with the ENF which therefore should be as close to 1 as possible. Finally, the background rate λ_b should be far smaller than the signal rate λ.
Note that the contribution from the circuit noise scales with $1/g^2$, therefore for high gains near 10^7 the term is completely negligible compared to the background [267]. λ_b on the other hand consists dominantly of dark count pulses (if no light leaks are present) and is usually small as opposed to λ – this is the fraction of dark counts in the SER described by (2.24), which always should be kept as small as possible.

Thus it is usually a valid approximation to write [285]

$$\frac{\sigma^2}{E^2} \approx \frac{F \cdot \lambda}{\lambda^2} = \frac{F}{\lambda} \tag{2.62}$$

which is exactly what was obtained in (2.54) and (2.60).

Excess noise factor measurement

Since (2.62) is valid in most cases, the ENF can be easily measured directly.
This can be done with a measured SER using (2.52) and $v_g^2 = {}^{\sigma_g^2}/_{g^2}$ by calculating σ_g and g from the SER [171; p. 13].
Alternatively, a large $\lambda = 10$–20 can be used, which produces a rather Gaussian-shaped multi-pe charge distribution, and in this case (2.62) can be used to obtain the ENF from the standard deviation σ_q and mean $\overline{q}$ of the distribution through [197]

$$F = \lambda \cdot \left(\frac{\sigma_q}{\overline{q}}\right)^2 \tag{2.63}$$

Independent of the method, the measured values will be larger than expected from (2.55) due to non-Poissonian corrections and the fraction of underamplified pulses; in the derivation of (2.54) it was assumed that CE and ME show Poissonian behavior, which is certainly not the case.
For the R5912 a theoretical ENF of 1.075 was calculated above, while the value measured in this work with 1.25 is far higher (see table 3.2).

Resolution

Also called energy/charge/single electron resolution, this is the most relevant quantity for characterizing the SER as it relates directly to the ability to distinguish a pulse with n photoelectrons from one with $n \pm 1$.

Unfortunately, for the resolution (R) two definitions coexist, which both are frequently used.

One uses the square root of the relative variance, i.e. the standard deviation of the spe distribution σ_1 divided by the mean $\overline{q}$:

$$R_\sigma = \frac{\sigma}{E} = \frac{\sigma_1}{\overline{q}} \tag{2.64}$$

which has to be calculated from the SER.
Sometimes $\sigma_{q,\mathrm{spe}}$ is extracted from a Gaussian fit to the spe peak instead of through numerical integration. This artificially shrinks the resolution, because the under-amplified pulses are left out, and should be regarded with great skepticism, as the ability to resolve the peaks is still affected by the ignored effects.
Using eqs. (2.64), (2.52) and (2.55), R_σ can be related to the ENF F and the dynode gains:

$$R_\sigma = \sqrt{F-1} = \sqrt{\frac{1}{\delta 1} + \frac{1}{\delta 1 \cdot \delta 2} + \cdots + \frac{1}{\delta 1 \cdot \delta 2 \cdots \delta \mathrm{n}}} \tag{2.65}$$

This shows that a high first stage gain $\delta 1$ is of paramount importance to achieve a low charge resolution and enable photon counting.

In the other definition, the full width at half maximum (FWHM) of the spe peak (w_1) is divided by either the peak position q_{p1} or the mean $\overline{q}$ (more correct):

$$R_F = \frac{\text{FWHM}}{E} \quad \rightarrow$$

$$R_{F,p} = \frac{w_1}{q_{p1}} \tag{2.66}$$

$$R_{F,m} = \frac{w_1}{\overline{q}} \tag{2.67}$$

If the peak position is used, the resolution can be obtained directly by eye from the distribution without numerical integration, which explains the popularity of (2.66). Values obtained using (2.67) are larger than from (2.66) since due to the large fraction of undersized pulses the mean lies below the peak. Therefore (2.67) gives more realistic values.
When it is not clear, if $R_{F,p}$ or $R_{F,m}$ is used in a publication, R_F can be used.

All three definitions are usually quoted in percent. For a Gaussian distribution σ and FWHM can be related by [287; p. 28]

$$\text{FWHM} = 2\sqrt{2\ln(2)}\,\sigma = \sqrt{8\ln(2)}\,\sigma \approx 2.35\,\sigma \tag{2.68}$$

This can be used to estimate R_F from R_σ and vice versa. However, since the SER differs from a Gaussian, especially due to the underamplified pulses, this conversion should be used with caution.

To give a few examples, R_F amounts to 70.0±7.9% for the R12860 20" PMT [236], and for the newest prototypes of the 20" Hamamatsu HPD R12850 (p. 684) an R_σ of only 10% ($\rightarrow R_F \approx 24\%$) was measured [165].
SiPMs are even better (e.g. [304]) and allow non-statistical photon counting, compare fig. 2.25. One reason why SiPMs have such good resolutions is that due to the low applied voltages and their small effective atomic number Z_{eff} there is only little backscattering and thus few underamplified pulses filling up the valleys [170; p. 254].

Care should be taken to ascertain which definition is used in a publication, as the choice of the respective variant tends to be too "obvious" to the authors to need to be stated explicitly. Often, the choice of definition is also used to euphemize the performance of a sensor: Def. (2.64) gives values which are about 2.35 times smaller than for (2.67) or (2.66), which normally helps to distinguish if FWHM or σ was used based on realistic expectations.

At its heart, the resolution is nothing different than the ability to resolve two close peaks smeared by statistical effects occurring during measurement of the respective variable (e.g. the creation of secondary particles) – i.e. the capability to distinguish two close lying values. Two overlapping identical peaks are typically considered unresolvable if their distance is smaller than the FWHM [195; p. 117]. This directly leads to the FWHM-based definitions.
Alternatively, the resolution can be seen as the uncertainty with which the true value can be reconstructed, and in this case a definition based on the standard deviation σ seems more natural.

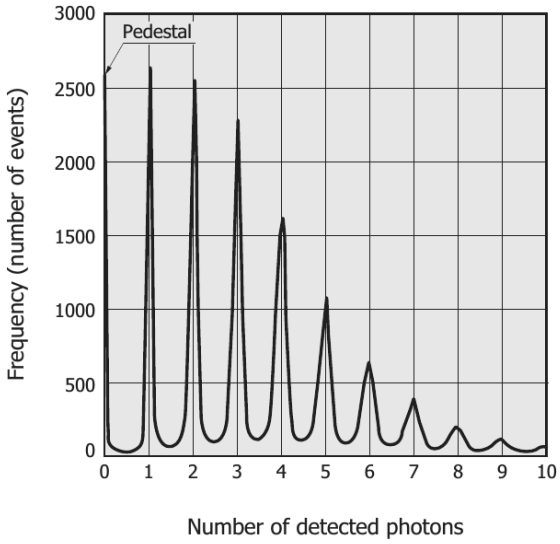


Figure 2.25: Charge spectrum for a Hamamatsu silicon photomultiplier (SiPM)[68] for multi-pe illumination, from [305; p. 3]. The extremely small resolution leads to an excellent separation of the n pe distributions. With increasing n, the statistical broadening predicted by (2.46) becomes visible.

In any case, a smaller resolution allows to better separate the n pe peaks, since they overlap less resp. the overlap starts at higher n pe. From (2.46) (or using (2.30) as an approximation) it can be seen that $\sigma_n \propto \sqrt{n}$, so even for a sensor with an extremely low resolution at some point the peaks will start to overlap and unambiguous photon counting is no longer possible (compare fig. 2.25).

As was discussed above, in order to have a small resolution, it is beneficial to maximize the PDE and gain and have a small ENF and background rate. The ENF itself is minimized by a large and uniform $\delta 1$, good gain stability and linearity, low AP rates, and a small fraction of underamplified pulses. Of these properties the QE and the gain on d1 considered most crucial by [197]. Besides improved photon counting capabilities, a smaller resolution also reduces the required dynamic range of the electronics [232]. The dependence on HV and PMT series is covered on p. 497 based on measurements performed in this work.

Poisson limit

Determining R_σ for a Poisson light source (e.g. from (2.59)) yields the familiar formula for the energy resolution of detectors known from literature:

$$\frac{\sigma}{E} = \frac{1}{\sqrt{\lambda}} \tag{2.69}$$

This is the Poisson limit imposed by basic physics, which is valid for any photosensor [170; pp. 8, 20] [244, 285].

In principle, performance *better* than the Poisson limit is possible through a Fano factor[69] $F_F < 1$, as for Si and Ge detectors. Of course, this does not circumvent

[68] Called MPPC (multi-pixel photon counter) by this manufacturer.
[69] This is usually *also* abbreviated as F, like the ENF, and thus the Fano factor is denoted by F_F here.

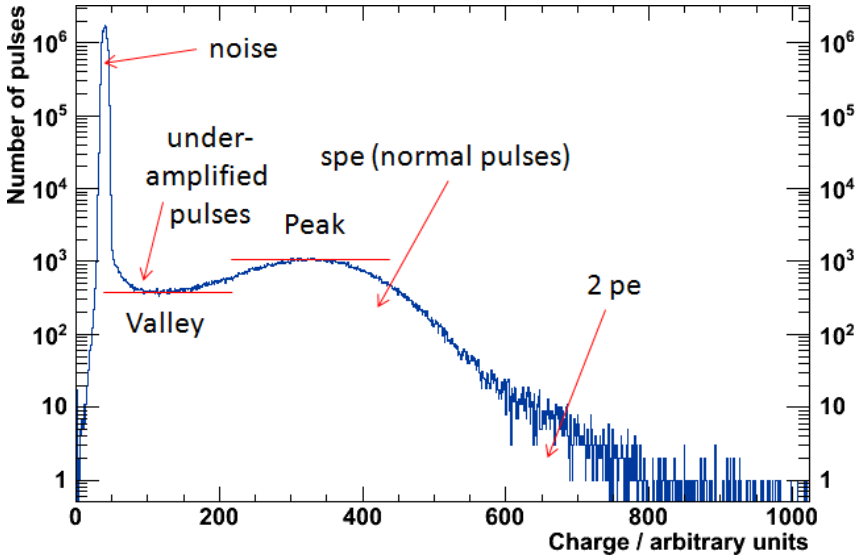


Figure 2.26: Definition of the peak-to-valley ratio (P/V) in the spe charge distribution (SER). Measurement for a Hamamatsu R5912 at $g = 1.3 \cdot 10^7$.

the laws of probability. Instead, a greater number of interactions occurs through other effects such as phonon scattering, which does not contribute to the creation of charged particles (which are detected) and causes only minute energy losses, but effectively increases N in $\sigma/E = 1/\sqrt{N}$ and thereby lowers the resolution [195; p. 118].

For visible/near-UV light detection, however, to the best knowledge of the author no photosensor with $F_F < 1$ exists due to the small photon energies, and the best sensors only come close to the Poisson limit [170; pp. 8, 20] [244, 285].

Peak-to-valley ratio

The peak-to-valley ratio (P/V) is a secondary quantity associated with the charge resolution. It is defined as the ratio of the maximum value of the spe peak (peak) to the minimum between spe peak and pedestal (valley), compare fig. 2.26.

Both the resolution and the P/V are used as variables to assess the photon counting ability of a PMT, but the resolution is a more direct criterion. A high P/V indicates a small fraction of undersized pulses and a well-defined spe-peak, and therefore will result in a small resolution, which is why the P/V can be used to assess the quality of the resolution. However, no conversion formula between resolution and P/V exists. The P/V measures how much the valleys in between the n pe peaks are filled and gives an indirect measure for the separability of n pe peaks.
For sensors with extremely good resolution, the valley is devoid of events, so the P/V is simply ∞ and stops being useful for comparison. On the other hand, it should be noted that both $R_{F,m}$ and $R_{F,p}$ are only defined if P/V > 2, whereas the P/V can already be employed if it is > 1. Thus the P/V is a useful quantity only for good (with regard to photon counting), but not for very good photosensors.

Good PMT specimens can reach P/V values of above 5 (Hamamatsu R7081 [160]), for currently available large area PMTs, the series average ranges from 2.3 (ETE D784 [306]) to 3.8 (Hamamatsu R12860 [307]); see also section 3.1.

This feature has seen remarkable improvements throughout the last few decades; not so long ago, even good PMTs had no valley at all and the spe "peak" was just a bump in a monotonically decreasing charge distribution.
On p. 498 the dependence on HV is shortly addressed for PMT series measured in this work.

In order to obtain reliable values for peak and valley, unaffected by fluctuations, it is recommendable to perform local fits of both. For the valley e.g. a third degree polynomial can be used, for the peak one of the spe peak functions mentioned above – failing that, a 3rd degree polynomial can be used as well [190, 306].

Signal-to-noise ratio

The signal-to-noise ratio (S/N) is simply the inverse of R_σ [244, 267]:

$$\mathrm{S/N} = \frac{E}{\sigma} \tag{2.70}$$

and all observations made for the relative charge variance can be directly applied. Using (2.61) with (2.70) yields

$$\mathrm{S/N} = \frac{\lambda}{\sqrt{F\left(\lambda + \lambda_b\right) + \left(\frac{\mathrm{ENC}}{g}\right)^2}} \tag{2.71}$$

For a derivation of the signal-to-noise ratio from shot noise and electronics parameters, please refer to [267], [169; pp. 75–78] and [171; pp. 12–13]; the results are the same as obtained here, but with less effects taken into account.

Regarding (2.71), some points should be stressed.
The noise (the denominator) is *signal-dependent*, since the photocurrent noise rises with the signal (λ); a direct consequence from Poisson statistics. Nevertheless, all in all S/N *increases* with the signal (as it should be). The background noise (DCR and photon background) and the anode shot noise are independent of the signal.
The S/N for low light levels is < 1 due to Poisson statistics. Using (2.69) with (2.70) yields $\mathrm{S/N} = 1$ for $\lambda = 1$.
Since the background noise constitutes a constant contribution, when λ approaches λ_b the S/N deteriorates with falling λ, in addition to the decrease caused by the Poisson limit (see [267; p. 38]). Again, this calls to mind the importance of a low dark count rate.

Background subtraction [70]

The background rate λ_b from DCR etc. can be measured in the absence of an optical signal in order to statistically subtract it. The aim of this subtraction may be to improve the S/N and the determination of λ, as well as to correct the values for the occurrence rate of timing effects and afterpulses. Since an additional measurement with its own uncertainties has to be performed for the subtraction, this can, however, only *increase* the variance and *reduce* the S/N.
Assuming a background measurement with a total recorded time window t_b, compared to the actual measurement with a total signal window t_s, according to [267]

[70] Content based on [267].

(2.71) can be extended with a correction factor $1 + {}^{t_s}/_{t_b}$:

$$\mathrm{S/N} = \frac{\lambda}{\sqrt{F\left(\lambda + \lambda_b\left(1 + \frac{t_s}{t_b}\right)\right) + \left(\frac{\mathrm{ENC}}{g}\right)^2}} \tag{2.72}$$

If both windows have the same length, i.e. the dark noise measurement was as long as the original measurement, the correction factor amounts to 2; only if the background window is far longer it approaches 1.

Measurement

The charge can be measured using either a charge-to-digital converter (QDC) or digitally with an oscilloscope or FADC. In all cases, in order to obtain the charge the PMT waveform (in form of voltage values) is integrated over a chosen time window.

Integration window

The integration window should be chosen not too large or the obtained value will be subject to statistical broadening from baseline noise and shifts due to imperfect subtraction of a baseline offset determined from a part of the waveform, or could include EAPs and possibly dark counts if the DCR is high.

Noise contribution

The contribution from noise can be seen as a sum of N random picks from the (typically Gaussian) noise distribution, with N the number of datapoints. This is equivalent to the convolution of N identical Gaussian distributions. It can be shown [302] that the convolution on N Gaussian functions with standard deviation σ_i and mean μ_i has a mean of

$$\mu_c = \sum_i \mu_i \tag{2.73}$$

and a standard deviation of

$$\sigma_c = \sqrt{\sum_i \sigma_i^2} \tag{2.74}$$

In the case of baseline background, a (remaining) baseline offset V_b to zero is then linearly amplified with N when determining the charge, and the noise standard deviation σ_n is broadened to $\sqrt{N}\sigma_n$, which is the statistical broadening mentioned above.
From this can be concluded that the integration window should be minimal so as not to increase the error introduced from noise.

Baseline subtraction

Another consequence is that the baseline should be subtracted before integration for *each* waveform, since the offset can vary with time, due to electronics drift (temperature changes etc.) and low frequency noise coupled into some part of the setup, so it is not sufficient to determine this value once for all waveforms. Even the long-term mean baseline value can lie at a voltage different from zero, if not

everything connected to the signal line is decoupled correctly using capacitors (which can introduce bandwidth limitations and ringing).
Already a minute average baseline value $\overline{V_b}$ can introduce large charge errors when integrating over a pulse, let alone a large pulse window: Each datapoint contributes a charge error δq (see eq. (2.86)) of

$$\delta q = \overline{V_b} \cdot \delta t / R \tag{2.75}$$

with δt the time increment between points and R the load resistance. The pulse height is of course affected as well, but the error is "only" $\overline{V_b}$. It is therefore good practice to subtract the baseline from each waveform before determining charge, gain and height.
This can be achieved by determining the baseline mean from a window $[t_{b,s} \dots t_{b,e}]$ (n_b datapoints) containing no correlated signals, and it is best to plan such a window, e.g. before the $\gamma 1$ peak, already when setting the acquisition window. The mean baseline value then can be calculated through

$$\overline{V_b} = \frac{\sum_{t=t_{b,s}}^{t_{b,e}} V(t)}{n_b} = \frac{\sum_{t=t_{b,s}}^{t_{b,e}} V(t)}{\frac{t_{b,e}-t_{b,s}}{\delta t} + 1} \tag{2.76}$$
$$\text{where} \quad n_b = \frac{t_{b,e} - t_{b,s}}{\delta t} + 1$$

$\overline{V_b}$ can then be subtracted for all datapoints of this waveform.

Occasionally, dark count events will lie inside this window, or the falling flank of a dark count pulse lying before the window can be partially contained, which alters the calculated $\overline{V_b}$ and thus introduces errors. These compromised waveforms should be excluded by an upper cut on the standard deviation of all datapoints inside the window, and ideally also upper and lower [71] cuts on the baseline mean value.

The method presented in (2.76) is equivalent to fitting a constant to the baseline. Naturally, this only produces reliable results, if the baseline *can* be approximated by a constant. This is the case if the time scale of baseline fluctuations is much larger than the acquisition window. Alternatively, the baseline can be fitted with a higher order polynomial, which is then extrapolated over the whole waveform. This, however, is numerically far more costly than the simple algorithm of (2.76). To decrease the error of such extrapolations, especially near the other end of the waveform, it can be reasonable to use an additional baseline window at the end of the waveform and interpolate between both.

The error σ_b on $\overline{V_b}$ from noise due to a finite averaging window length can be estimated from n_b. Similar to the noise contribution to charge, the averaging of the baseline mean contains the sum Σ_b over n_b draws from a Gaussian noise distribution, which is equivalent to an n_b-fold convolution of the Gaussian function with mean μ_n and standard deviation σ_n. From (2.73) and (2.74) follows for the mean μ_c and standard deviation σ_c of the convolved Gaussian

$$\begin{aligned} \Sigma_b \equiv \mu_c &= n_b \, \mu_n \\ \sigma_c &= \sqrt{n_b} \, \sigma_n \end{aligned} \tag{2.77}$$

[71] Overshoots can cause too low values.

From this the baseline mean can be calculated to

$$\overline{V_b} \equiv \mu_n = \frac{\Sigma_b}{n_b}$$

as is intuitively clear (compare (2.76)). The statistical error σ_{μ_n} of μ_n then follows to

$$\sigma_b \equiv \sigma_{\mu_n} = \sqrt{\frac{1}{n_b^2}\,\sigma_c^2} = \frac{\sigma_c}{n_b} = \frac{\sigma_n}{\sqrt{n_b}} \tag{2.78}$$

which again is not unexpected.
If e.g. a relative error of 10% is desired for $\overline{V_b}$, $n_b = 100$ is required. For an acquisition rate of 8 GS/s (gigasamples/s) follows δt =0.125 ns, and then the baseline window needs to span at least 12.5 ns.

Now the influence of this error on the integrated charge of a pulse can be derived. For a pulse with length t_p including $n_p = (t_p/\delta t) + 1$ datapoints [72] and with a charge q_p follows the relative charge error $\delta_{q,b}$ introduced by a finite baseline window length from (2.75), (2.77) and (2.78):

$$\delta_{q,b} = \frac{\sigma_b\, n_p}{q_p} \cdot \frac{\delta t}{R} = \frac{n_p}{\sqrt{n_b}} \frac{\sigma_n}{q_p} \cdot \frac{\delta t}{R} \tag{2.79}$$

To reduce this error, a low noise environment, long baseline window, short time increments, large gain, and short PMT pulse shapes are beneficial.
For a pulse with $q_p = 1.6\,\mathrm{pC}$ ($g = 10^7$), $t_p = 9\,\mathrm{ns}$, $\delta t = 0.125\,\mathrm{ns}$ and $R = 50\,\Omega$, assuming $\sigma_n = 0.3\,\mathrm{mV}$ this results in a relative charge error of 0.34% for $n_b = 100$ datapoints. The error improves with $1/\sqrt{n_b}$; however, setting a much larger value than 20 or 50 ns typically is not possible, since this limits the usable acquisition window, reduces sampling rate, or increases measuring time.
At first glance this error seems acceptable; however, in the example, the mean spe charge at $g = 10^7$ and a low-noise baseline were used. If effects with smaller charges, such as E1 or γ1, shall be studied, the error quickly becomes large, since $\delta_{q,b}$ scales with $1/q_p$. For a 0.1 pe charge it amounts to 3.4%, for 0.01 pe already resp. 34.2%. To analyze small charges the baseline window thus should be made as large as reasonably possible.

Online analysis

If NIM [73] resp. VMEbus electronics or digital oscilloscopes are used, the charge integration window should be as short as possible. This should always include the complete pulse shape, especially the longer falling flank, and contain at least the main peak and the time shift effects from EP and LP in the TTD (see 2.3), as they comprise a large fraction of events. Depending on which effects are important for the desired application of the PMT, the window should also encompass γ1, and possibly also L1L1. These are L1 which are elastically scattered on d1 towards the cathode *again* (an L1 of an L1) and frequently are visible in the TT distribution due to the large backscatter coefficients. As a drawback, extending the integration window to account for L1 will also include most EAPs.

[72] Assuming t_p is an integer multiple of δt
[73] Nuclear Instrumentation Module

Offline analysis

If an FADC is used, this yields the benefit of offline analysis, as one can select the integration window *event-based*, e.g. using a Schmitt trigger: A small start threshold $V_{\mathrm{thr,s}}$ lying just above the expected noise spikes from the baseline (see (2.89)) is used to find the start time of the pulse, whereas the end threshold $V_{\mathrm{thr,e}}$ for determining the pulse end can be set at a lower value lying *within* the baseline fluctuations in order to use as much of the trailing flank for integration as possible and reduce the systematic error.
Care must be taken if overshoot or oscillations occur, as in this case it is not sufficient to stop integration when the pulse has descended into the baseline again. In this case a fixed time window can be used which can be set base on the average pulse shape (APS, see p. 339). Alternatively, the integration can be stopped if the absolute values of all datapoints within a test window are below a threshold. As a third option, the integration can be ended at the zero-crossing when the resulting charge value is multiplied with a correction factor obtained from the APS – this introduces a small error from the extrapolation, but reduces statistical smearing and excludes EAPs.
Another benefit of an offline analysis is that one can also veto pileup from e.g. EAPs or even 2 pe-pulses (which *reduces* the 2 pe contamination) using pulse shape parameters.

Numerical integration

Simpson's rule[74]

When performing the integration offline, this is best achieved using algorithms optimized for numerical integration of equally spaced subdivision, e.g. one of the various forms of Simpson's rule. For analyzing the charge contained in a voltage pulse on a PMT waveform, an integration method including the boundaries (closed form) is required. Using Simpson's rule as a building block, adding up the individual integrations allows integration over the complete range (extended/composite Simpson's rule). This yields undesired oscillatory behavior in the summands, which can be avoided by an alternate derivation of the integration formula by fitting cubic polynomials through successive groups of four points. The integrated value $W_{0,n-1}$ for n datapoints from the first to the last datapoint $V_0 \dots V_{n-1}$, which span the time range $t_0 \dots t_{n-1}$ including both boundaries, then results to

$$\begin{aligned} W_{0,n-1} &= \int_{t_0}^{t_{n-1}} V(t)\,dt \\ &\approx \left(\frac{3}{8}V_0 + \frac{7}{6}V_1 + \frac{23}{24}V_2 + \sum_{i=3}^{n-4} V_i + \frac{23}{24}V_{n-3} + \frac{7}{6}V_{n-2} + \frac{3}{8}V_{n-1} \right) \cdot \delta t \end{aligned} \quad (2.80)$$

where δt is the time increment between two datapoints. The coefficients $\neq 1$ close to the boundaries result from approximating the waveform with a cubic polynomial. This has the consequence that integration is compromised if the pulse starts or ends within three datapoints from the obtained waveform. As an alternative, the trapezoidal rule or a plain sum can be used in such cases. Note that in (2.80), only the outermost six data points have weights differing from 1, so for $n < 6$ integrated data point using the plain sum produces similar values.

[74] Content based on [308; pp. 155–162]

Charge

Using (2.80) the charge q can then be obtained (compare p. 132) from

$$q = \frac{\int_{t_0}^{t_n} V(t)dt}{R} \approx \frac{\sum_i V_i \cdot \delta t}{R} = \frac{W_{0,n}}{R} \tag{2.81}$$

where R is the load resistance, typically $50\,\Omega$. t_0 and t_n are chosen such that the pulse is completely contained (offline analysis), resp. the pulses of all studied effects are included (online).
Optionally, a negative polarity of the voltage pulse can be corrected to produce positive values for q.

Setup gain correction

In addition, the total gain g_{setup} of the setup without the PMT needs to be taken into account. To obtain this value, one has to measure the gain of used amplifiers, as well as the attenuation through cables, splitters and electronics. In general, cables should be as short as possible as they both attenuate the pulse and broaden it since the capacitance of a cable in combination with the load resistance acts as a low pass. This leads to a correction term g_{setup} which when combined with (2.81) leads to the corrected charge q_c:

$$q_c = \frac{W_{0,n}}{R \cdot g_{\text{setup}}} \tag{2.82}$$

Occupancy correction in the charge distribution

In order to obtain the mean charge of an spe pulse, the contributions from pulses with 2 pe or more have to be eliminated, exactly as for the gain (see p. 97 and 137).

2.2.3 Pulse height

The pulse height is the voltage difference between the peak of a PMT pulse and the baseline. For typical PMT gains around 10^6–10^7 and photon counting, usually the pulse height is quoted in mV. Since a pulse consists of a bunch of electrons leaving the anode and ultimately being sent to ground over a resistance of normally $50\,\Omega$ after passing the electronics, the anode pulse is *negative* in height. Nonetheless, for ease of treatment often the absolute value is employed as the pulse height, since no positive pulses are generated (except for overshoot in the wake of a negative pulse).

Pulse height distribution

Analogous to the SER, the pulse height distribution (PHD) shows the spread of pulse heights in response to a light source (spe or multi-pe). The spe PHD is very similar in shape to the charge distribution, and often both are used interchangeably [75] to analyze the behavior of a PMT and determine characteristics such as the resolution. However, there are subtle differences and both distributions are *not* identical when scaled by a conversion factor. Therefore, for variables obtained from the PHD

[75] In fact the spe PHD is sometimes also called SER in literature.

the prefix "pulse height" should be placed in front, e.g. "pulse height resolution". Otherwise the same definitions as introduced in the charge section can be used. Due to the differences between SER and PHD it is therefore unavoidable to measure both the charge and height distribution, since the threshold setting is often based on the *height* distribution.

Differences between height and charge distributions

These differences are caused by effects which modify the pulse shape compared to the "standard pulse" resulting from regular amplification (regular pulse, RP).

Statistics

Statistics alone is sufficient to induce small changes, since pulse shapes can slightly differ for the same charge. For example, if one SE from d1 is slightly delayed through a longer trajectory to d2 (e.g. because of a different emission angle), the pulse will be slightly broader and the pulse height a little smaller.

Noise

There is always noise present on the baseline from various sources. This is typically normally distributed around the mean voltage offset $\overline{V_b}$ of the baseline, so on average it mostly cancels out when integrating over a window to obtain the charge, save for statistic effects (see p. 124). The case is different for the pulse peak, as this equates to the extremum found during the pulse duration, so random noise fluctuations in the large majority of cases will artificially increase the measured quantity above the true value.

Smoothing

In order to reduce the effect from noise, one can use smoothing [76].
The simplest way to do this is by using the moving average of datapoints within a small window. This should always be avoided, as it admittedly preserves the charge under the curve (0^{th} moment) and its mean position in time (1^{st} moment), but *not* the width (2^{nd} moment) and therefore also not the height – the filter increases the width and reduces the pulse height.
A better alternative are Savitzky-Golay filters (also called least-squares or digital smoothing polynomial filters). These can be seen as extensions of the simple moving average filter, which preserve higher order moments of a pulse such as the variance, skewness and so on. They are based on the following function, which replaces a datapoint f_i with a smoothed value $f_{i,s}$:

$$f_{i,s} = \sum_{n=-n_l}^{n_r} c_n \cdot f_{i+n} \tag{2.83}$$

where n_l and n_r are the number of points lying to the left resp. right of f_i, and c_n are the weighting factors of each datapoint.

[76] The following passage is based on [308; pp. 766–772]

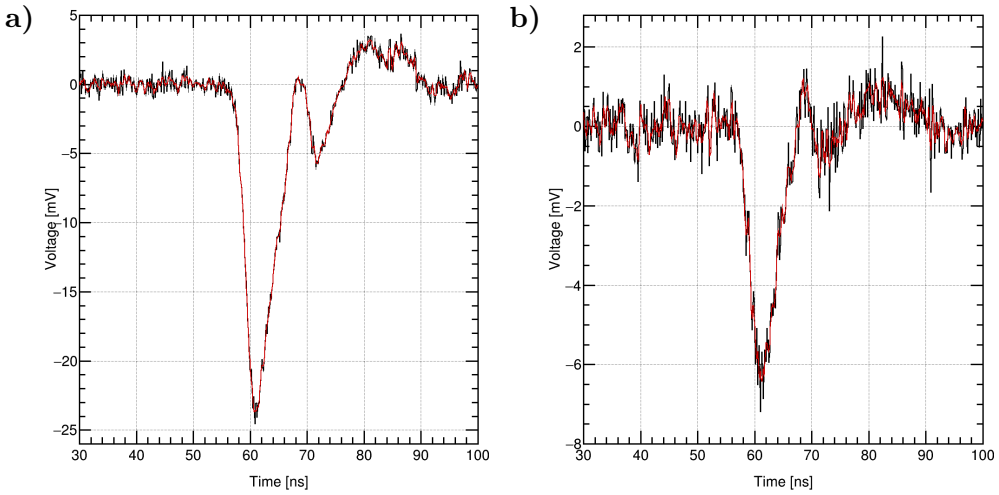


Figure 2.27: Effect of using a Savitzky-Golay filter (red) with a window of 11 datapoints and a 4^{th} order polynomial on PMT waveforms (black) with a) regular charge and b) low charge. Smoothing reduces the noise spikes, while preserving the PMT pulse shape, which gives better estimates for the true pulse shape parameters such as height, rise time, fall time and width. The pulses were recorded for a Hamamatsu R5912 at $g = 1.3 \cdot 10^7$, $o = 2.9\%$ with a sampling rate of 8 GS/s.

The moving average filter can be obtained as a special case for constant $c_n = 1/(n_l + 1 + n_r)$.
For higher-order filters, one basically performs a least-squares fit of a higher-order polynomial (the moving average uses $f(x) = \text{const}$) for each datapoint using the surrounding points within a chosen window. Luckily, the fitting can be performed in advance using a dummy function, which results in a set of coefficients c_n for the chosen window and order of the fitted polynomial. A very good introduction to Savitzky-Golay filters including the derivation of the coefficients and a script to calculate them is given in [308; pp. 766–772].
Examples of using a Savitzky-Golay filter on large and small PMT pulses are shown in fig. 2.27. Since the standard deviation of the baseline is similar during a measurement, the effect of noise enhancement of the pulse height is most severe for small pulses.

Smoothing should only be used carefully, however, and just to obtain pulse shape parameters such as height, width, rise time and fall time; variables such as the charge should be determined from raw data as even the best smoothing algorithm biases the data to some extent. Smoothing is justified only under the condition that the measured curve varies slowly compared to the overlying *random* noise, as it performs some sort of local averaging of data points.
An alternative to smoothing is the use of optimal filtering (also called Wiener filtering) via fast Fourier transform (FFT) [308; pp. 649–652, 673–675].

Pulse branching and broadening

Many of the time shift effects introduced in 2.3.5 can divide the charge into two (or more) separate pulses, e.g. BL1, BE1 and SE scattering on later dynodes. Depending on the delay between both pulses, they will either overlap and be identified as one pulse by a threshold–threshold pulse-search (pileup, leading to a broader pulse), or

be visible as two independent pulses (pulse branching). In both cases the maximum height of the primary pulse will be reduced with the severity depending on the effect.

Pulses with reduced number of amplification stages

In principle γ1, E1, and dynode dark counts will have pulse shapes somewhat different from the regular pulses since the high gain from d1 is exchanged by the smaller gain on d2 (confer eq. (2.40a)) and the cascade uses less stages. This should broaden the average pulse shape and reduce the height-to-charge ratio since the number of emitted SEs is lower and the pulse on average will be more statistically smeared in time due to different SE trajectories. In addition, the lower first gain should heighten the resolution (smaller mean gain but larger relative value of σ in the charge distribution). The transit time is smaller, which should lowering the TTS slightly, but the stronger fluctuations in pulse shape will increase it, so it is difficult to say which effect will dominate.

Satellite pulses

Independent of the charge splitting effects, if a primary pulse is shortly followed by a second pulse (e.g. EAP or a second photon pulse), it will have a reduced height compared to the total charge, if the integration window encompasses both pulses.
This is often observed for muon events, which in addition in most cases saturate the attached amplifier, so the pulse height is clipped at the maximum output voltage.

Height calculation from charge

In general, it is not possible to directly derive the height from charge or vice versa, since the height depends on the pulse shape of the PMT and is furthermore altered by the abovementioned effects.
The pulse shape alone already has a tremendous influence on the resulting height for a fixed charge: For a fast, compact pulse shape the height will be larger than for a slow, broad pulse. A higher transit time typically leads to a larger TTS and a broader waveform.
As a rule of thumb, a short rise time indicates a short TTS (see 2.4). If a manufacturer does not quote the TTS for a PMT, one can use this approximation to *roughly* estimate the TTS quality of a PMT.
Of course the pulse shape also depends on the number of dynodes and their type. A linear focused PMT – designed to compensate differences in SE drift times – will have a sharper pulse shape than a venetian blind PMT.

With some crude assumptions one can, however, estimate the expected pulse height.
The height $V_{\max}$ (voltage) is connected to the charge q via $R = V/I$ since the pulse flows to the ground over a resistance R of $50\,\Omega$. It is furthermore smeared in time due to the effects occurring during amplification, so instead of a delta function in time, a pulse is characterized by a rise time t_r, fall time t_f and a pulse width t_w (FWHM).
The easiest approach is to approximate the pulse shape with two triangles for the rising and falling flanks. For this method the targeted gain (or charge) needs to be known, for the rise and fall time sophisticated guesses can be made or average values for the PMT series can be taken.

From

$$R = \frac{V}{I}$$

$$\text{and} \quad I = \frac{dq}{dt}$$

$$\text{follows} \quad q = \frac{\int V\,dt}{R} \approx \frac{V_{\max}\,(t_r + t_f)}{2\,R} \tag{2.84}$$

$$\rightarrow \quad V_{\max} \approx \frac{2\,q\,R}{t_r + t_f} \tag{2.85}$$

If values measured for a R5912 are plugged in ($Q = 2.11\,\text{pC}$ corresponding to $g = 1.3 \cdot 10^7$, $R = 50\,\Omega$, $t_r = 2.65\,\text{ns}$, $t_f = 4.34\,\text{ns}$), this results in $V_{\max} = 30.1\,\text{mV}$ compared to the measured value of 19.8 mV. If no prior knowledge besides the gain is assumed and average values for this series are used ($t_r = 3.15\,\text{ns}$ and $t_f = 4.05\,\text{ns}$ [309]), a value of 29.3 mV is obtained. The predicted values are off by a factor of 1.5, which is due to not taking into account the pulse width, since the peak is surrounded by a rather flat plateau.

The pulse shape approximation can be adopted accordingly to improve the prediction by assuming that rise and fall time still can be described by triangles, but now their half height points have a distance of the pulse width t_w. This leads to

$$q = \frac{\int V\,dt}{R} \approx \frac{V_{\max}\left(\frac{t_r}{2} + \left(t_w - \frac{t_r}{2} - \frac{t_f}{2}\right) + \frac{t_f}{2}\right)}{R} = \frac{V_{\max} \cdot t_w}{R} \tag{2.86}$$

$$\rightarrow \quad V_{\max} \approx \frac{q\,R}{t_w} \tag{2.87}$$

Interestingly, this yields the same results as the simplest imaginable pulse shape model: a rectangle with height $V_{\max}$ and width t_w [170; p. 251]. Please also refer to p. 341 for a more sophisticated phenomenological model of the pulse shape.
If the values are entered with units mV, ns and Ω in (2.86), the resulting charge already is in pC; the same applies for (2.87) with pC, Ω and ns in, which lead to a height in mV.
Using the measured values (no average value for t_w was reported by [309]), this produces a $V_{\max}$ of 20.5 mV, which is only 3.5% larger than the real value, which shows that the improved formula (2.87) is a valid approximation.

Of course, using (2.86), also the charge can be estimated from a measured pulse shape.

From (2.86) one can also estimate the minimum gain required for photon counting by demanding a minimum pulse height $V_{\min}$ for reliable distinction of pulses from noise and using $g = q/e$:

$$g_{\min} = \frac{V_{\min}\,t_w}{e\,R} \tag{2.88}$$

If a typical threshold of 0.25 pe is chosen, a minimum threshold of 1 mV is assumed and the mean t_r and t_f of the R5912 are used, this leads to a required mean pulse height of 4 mV and a $g_{\min}$ of $2.6 \cdot 10^6$. This is the bare minimum for this PMT, and the gain should be set higher than that. Nevertheless, this simple estimate already explains why photon counting is usually performed with gains in the range of $5{\cdot}10^6$ to $2{\cdot}10^7$ [170; p. 207] – for too large values of g the problems listed in 2.2.4 start to become important.

Choosing the threshold value

In order to set the electronics threshold for a PMT, the PHD should be known.

Lower limit from baseline noise

An estimate for the required value can be calculated from the baseline noise under the assumption that this follows a normal distribution. For this, for at least one pulse-free window (better more) the standard deviation σ of the datapoints contained within has to be calculated with a digital oscilloscope or offline after FADC acquisition. The probability $p(h)$ that a single datapoint lies above a threshold value h follows from the cumulative distribution function (CDF) $\Phi(x)$ of the Gaussian noise distribution to

$$p(h) = 1 - \Phi(h) = 1 - \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{h-\mu}{\sqrt{2}\,\sigma}\right)\right)$$

where a mean value $\mu = 0$ can be assumed after baseline subtraction.
For the probability p_w to observe a noise trigger in a waveform, the number of trials equals the number of datapoints N in the window Δt, with

$$N = \frac{\Delta t}{\delta t}$$

where $\delta t = 1/$ sampling rate is the time increment. p_w then is 1 minus the probability to observe no noise trigger, from which follows:

$$p_w(h) = 1 - (1 - p(h))^N = 1 - \left(\frac{1 + \operatorname{erf}\left(\frac{h}{\sqrt{2}\,\sigma}\right)}{2}\right)^{\Delta t/\delta t} \tag{2.89}$$

The probability to observe fake noise pulses therefore depends on the threshold h, the baseline standard deviation σ, the window, and the sampling rate.
The expected number of fake triggers from noise spikes can then be calculated by multiplying $p_w(h)$ with the number of recorded waveforms.
For example, using $\sigma = 0.3\,\mathrm{mV}$, a long acquisition window of $\Delta t = 250\,\mathrm{ns}$ and a sampling rate of $8\,\mathrm{GS/s}$ leads to $p_w < 0.1\%$ for $4.9\sigma \mathrel{\hat{=}} 1.47\,\mathrm{mV}$. For 10^6 recorded waveforms this still produces 1000 fake triggers. If a contamination of < 1 fake events is desired, a threshold corresponding to $6.2\sigma \mathrel{\hat{=}} 1.86\,\mathrm{mV}$ is required.

From analysis of the height distribution

While this yields a lower limit for the threshold, which should not be undercut, this can only serve as an estimate. In the end, the threshold has to be selected by eye from the PHD, as there can also be non-Gaussian noise components.
Usually, one should set the threshold at the valley or if possible lower, just above the strong onset of the pedestal. It is often difficult to distinguish the pedestal from small pulses due to dynode dark counts, $\gamma 1$, E1 and other underamplified time shift effects. The choice of value also depends on the used electronics and whether or not postprocessing can be done. With e.g. a FADC one can perform offline analysis and reliably remove noise spikes, but this comes at the cost of more recorded events and consequently longer measurement time and requires more storage space; one could also say that the "data signal-to-noise ratio" is reduced. Without the possibility of offline waveform analysis it is better to set the threshold conservatively to avoid fake triggers.

2.2.4 Gain

There are two different types of gain relevant in photomultipliers [170; p. 414]. One is the multiplier gain g, defined as the average charge $\overline{q}_1$ of a pulse leaving the anode, which is induced by a single photon, divided by the elementary charge:

$$g = \frac{\overline{q}_1}{e} \tag{2.90}$$

The other is the PMT gain G, which is the ratio of anode current I_a to cathode current I_k:

$$G = \frac{I_a}{I_k} \tag{2.91}$$

$G < g$, since G includes CE and ME losses, whereas g is obtained only for photoelectrons which have successfully produced detectable cascades – it denotes the amplification of the electron multiplier. As already stated in (2.9d) the pulse efficiency PE is simply the ratio G/g.
For photon counting the multiplier gain g is most relevant and will be used in this work. Manufacturers, however, prefer to use the PMT gain G [285] as it is easy to measure and is of practical relevance in industrial applications [170; p. 203].

From (2.90) can be seen that g is proportional to the mean spe charge; thus, many effects relevant for g were already discussed in the charge section.
The peak gain (corresponding to the peak in the SER) and the mean gain are typically within 10% from each other [170; p. 253]. Thus it is often sufficient to use the peak gain and avoid computation of the mean.

Setting the gain

The PMT is a low noise amplifier, which is subject only to only shot noise and *not* to Johnson (thermal) noise, as it incorporates no resistor in the amplification process [170; p. 6]. Naturally, a resistance will always be present before termination on ground, but this is part of the electronics and the accompanying Johnson noise will occur here. Therefore one always want to *maximize* the amplification done *inside* the PMT, i.e. g; setting a lower gain and using preamplifiers instead will only increase the total noise. Since for photon counting most times it is unavoidable to use preamplifiers, the devices employed need to have high bandwidths of at least some 100 MHz to avoid broadening the PMT pulse, and so-called 50 Ω GHz wideband amplifiers should be used [310].

In addition, one wants to avoid unstable operation of the PMT (breakdown) at high voltages (HV), and to minimize unwanted disturbing effects such as the DCR, field emission [77], APs (since signals in a detector typically are randomly distributed in time, APs are an unavoidable background), and light feedback, all of which increase drastically with operating voltage. This sets a maximum limit to the gain.

In order to find the optimal operating point for the gain, measurements of signal rate and DCR vs. operating voltage such as shown in fig. 2.28 have to be carried

[77] This can be considered a contribution to DCR and dark current.

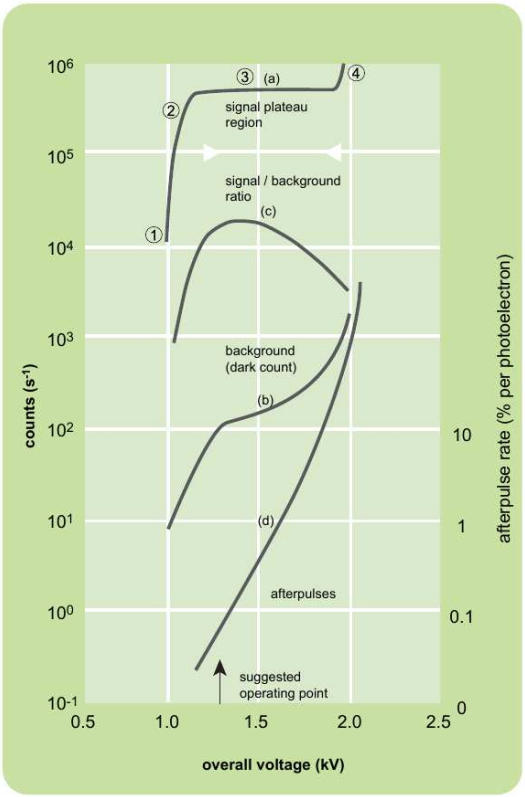


Figure 2.28: Selection of optimal working point in high voltage (HV) resp. gain for photon counting based on signal rate, DCR and AP rate, from [171; p. 26].

out [171; pp. 25–26]. The DCR is most critical for the S/N, but the AP rate should not be neglected either, as this becomes increasingly problematic for higher voltages. When a fixed threshold is used, the signal rate (a) first steeply increases with voltage, since due to small gain only the highest fraction of pulses in the PHD exceed the threshold (region 1), then levels out in region 2 as the bulk of the spe peak slips over the threshold, afterwards reaches a plateau with small slope (region 3) since basically all events save a few underamplified pulses are detected, and finally shows a strong rise when the PMT enters breakdown (region 4) due to light feedback and afterpulses inside the acquisition window. Note that during the signal plateau the gain continues to rise with HV following a power law.
If the DCR (b) is recorded, the S/N in dependence of HV can be plotted (c). The DCR first shows a similar behavior as the signal for low HV as the spe peak slips above the threshold, but then continues to rise as more and more dynode dark count events become detectable. Therefore the S/N is a flat peak close to the onset of the DCR "plateau". Since APs also constitute background and their rate increases dramatically with voltage (d), this suggests to choose a HV tending to the left of the broad S/N peak or at the maximum.
This is not always possible. Often it is necessary to use a higher gain to draw more underamplified pulses out of the pedestal and increase the PDE, or to use extremely high gains due to long cable lengths between PMTs and electronics. For example, in the case of AMANDA, Hamamatsu R5912-2 PMTs with 14 dynodes were operated at a gain of 10^9 to drive the signals through 2 km of cable without amplification [311]. For unstable power supplies it might be best to operate on the flattest point of the

signal plateau instead in order to maximize the stability against gain changes, since the gain follows a power law with HV.

For the interested reader, [312] elaborates on how to set a predefined gain (e.g. 10^7) in a PMT based on the λ and charge measurement methods introduced in [286] (see p. 94). This is important for experiments incorporating many PMTs, whose charge behavior is supposed to be comparable. In this case, the optimal working point is determined for a select few specimens as above, yielding a target gain. Since the dynode gains vary from device to device even if from the same series, the individual HV required to obtain the target gain must be determined for each PMT, for example by using the method detailed in [312]. [285] mentions that for different PMTs the δ of each stage can vary by $\pm 10\%$ at the same HV, which for a 10-dynode device can lead to variations by roughly one order of magnitude.

Measurement [78]

In the following, several methods for measuring the multiplier gain g and the PMT gain G are detailed. These measurements are subject to unavoidable systematic errors and obtaining high accuracies can proof rather difficult.
Since g and $\overline{q}_1$ are connected by (2.90), the methods explained in the following can also be used to measure the mean charge.
It should be noted that the gain measurement depends on the wavelength, angle of incidence, illuminated area and the used voltage divider (VD) configuration. When measuring gain it is therefore important to illuminate the PMT in the same way and use the same VD as in the experiment.

Measurement of PMT gain G

All methods to obtain G revolve around measuring I_k and I_a. The problem is that one cannot measure I_k and I_a at the same time for high gains of about 10^7: When a light flux is applied which results in a measurable photocurrent $\gg$ pA in order to minimize errors from leakage currents, the anode current is deep in the non-linear regime, if not in breakdown already. So both values have to be measured separately. This is the same problem as in the PE measurement (see p. 86).

Gain scaling

In the first method initially a HV corresponding to $g = 10^4$ is applied, at which both I_k and I_a can be measured at the same time using a high light flux. The light intensity is then reduced until the anode current is 10 times smaller, followed by an increase in HV until the same anode current is reached. This is repeated until the desired gain is reached. As a benefit, the gain–voltage curve can be obtained as a side-product. The errors of this method originate from gain scaling uncertainties.

Light attenuation

Alternatively, the light flux can be reduced by a known factor using neutral-density filters (NDFs) (also [285]). If after measuring I_k with a high photocurrent, an NDF

[78] Content based on [170; pp. 241–254]

with optical density D is put in front of the PMT and I_a is measured, G can be calculated from

$$G = \frac{I_a \cdot 10^D}{I_k} \tag{2.92}$$

The errors now derive from light attenuation uncertainties and are manifold, similar to the methods for the PE measurement (see p. 86).

Test sheet

The simplest way is by using the sensitivity data from the manufacturer's test sheet of the respective PMT. There, the cathode luminous sensitivity S_k and the anode luminous sensitivity S_a are quoted. Both quantities are related by

$$S_a = S_k \cdot G \tag{2.93}$$

$$\rightarrow G = \frac{S_a}{S_k} \tag{2.94}$$

Typically two voltages are listed at which certain values of S_a are reached, and from these the gain–voltage curve can be reconstructed as described on p. 140.
Since the manufacturers also use optical filters, the same problems as for the previous method apply. Furthermore the illumination properties used when obtaining the values might not be the same as in the final application.

As a word of warning: Wright [170] considers the accuracy of all three methods as order of magnitude only.

Measurement of multiplier gain g

Single electron response

The first method is based on the measured SER (also [285]). This can be achieved either using a digital oscilloscope, a FADC or a multichannel analyzer (MCA) incorporating a charge sensitive preamplifier. The gain is then calculated directly using (2.90).
The accuracy of this method depends on the inclusion of the low charge underamplified pulses; according to [170] this is the best method. While one could fit the pedestal with a Gaussian and subtract it to reveal the covered underamplified pulses, this will not help much due to the high *absolute* statistical error of the bin counts in the noise peak: Subtracting a large bin count (noise fit) from a large bin count (noise + signal) results in enormous *relative* error bars for the resulting small bin content (signal). There is simply no way around minimizing setup noise for high-quality gain (and other) measurements.

As already explained on p. 97, the gain obtained from (2.90) should be corrected to account for the mean number of detected pulses. Using (2.35), (2.15) and (2.17) results in the corrected gain g_c:

$$g_c = \frac{\overline{q}_1}{e} \cdot \frac{o}{\lambda} = \frac{\overline{q}_1}{e} \cdot \frac{1 - e^{-\lambda}}{\lambda} = \frac{\overline{q}_1}{e} \cdot \frac{o}{-\ln(1-o)} \tag{2.95}$$

The MiniBooNE collaboration arrived at the same formula [313; p. 100].

Single electron response and anode current

Alternatively, one can measure the SER and at the same time I_a. The SER is then integrated between a charge q and ∞, and the result is plotted in dependence of q. Following this, the curve should be extrapolated from the lowest values to $q = 0$ to obtain the total count rate f per second. The gain can then be calculated from

$$I_a = f\, e\, g. \tag{2.96}$$

As for the last method, the inclusion of underamplified pulses through the extrapolation is a source of uncertainty. In addition, a dead time correction is needed for f (see (2.196)), and APs contribute to I_a but not necessarily to f, which is why this method overestimates the gain.

Shot noise

This method employs the S/N of the cathode photocurrent. As it turns out, this is identical to the S/N for the anode current S_a/N_a, which can be easily measured with a multimeter through

$$\frac{S_a}{N_a} = \frac{V_{\mathrm{DC}}}{V_{\mathrm{rms}}} \tag{2.97}$$

with V_{DC} the direct current (DC) voltage and V_{rms} the root mean square (RMS) voltage.
It can be shown that [170; pp. 245–246]

$$\left(\frac{S_a}{N_a}\right)^2 = \frac{I_a}{2\, e\, B\, g\, F} \tag{2.98}$$

where B is the quadratically combination of the bandwidth of the multimeter used for the measurement and $1/(4RC)$, with R the parallel combination of load resistor and multimeter resistance, and C the total capacitance of the circuit, obtained with a capacitance meter.
If $(S_a/N_a)^2$ is plotted over I_a, the gain g can be extracted from a fit of the slope.
While this may sound complicated, it can be performed with a common multimeter *alone* with any software capable of performing fits or even a manual calculation using the two outermost points and (2.97) and (2.98).

Pulse counting and anode current

Here, simply the count rate in dependence of the voltage $f(\mathrm{HV})$ is obtained with a scaler for a fixed threshold, alternating with measurements of I_a. The gain is then derived from (2.96).
The benefit of this method is that no knowledge of the SER is required. However, for the same reason, it is difficult to determine which f corresponds to the true gain, as the signal plateau observed in $f(\mathrm{HV})$ has a slope > 0, and it is not clear at which point all underamplified pulses are included but noise and APs are still excluded. This results in large errors in the determination of g of about 30%.

Quick estimation of multiplier gain g

When it is sufficient to merely quickly estimate the gain, and the methods listed above would require too much time, one of the following procedures can be applied instead.

Pulse shape

g can be easily estimated from observing a few pulses on an oscilloscope and applying (2.86) for the obtained average pulse variables.

Pulse counting and anode current

Alternatively, the number of pulses n in a recording window Δt can be counted to obtain a rate $f = n/\Delta t$. By measuring I_a the gain can be calculated through

$$g = \frac{I_a \cdot \Delta t}{e\, n} \tag{2.99}$$

The inclusion or rejection of small pulses requires individual judgment.
This is exactly the same method as described in "Measurement of multiplier gain g", but with an oscilloscope in place of a scaler.

Integrating capacitor

Lastly, g can be appraised by using a self-made integrator. For this purpose, the average pulse height V produced by an spe pulse integrated by a capacitor with known capacitance C is measured. If the time constant of the circuit $RC \gg t_w$, the pulse is completely integrated and the gain can be retrieved using

$$\begin{aligned} q &= C \cdot V \\ \rightarrow g &= \frac{C\,V}{e} \end{aligned} \tag{2.100}$$

To realize this circuit, a coax cable connecting the PMT to an oscilloscope with a resistor added in parallel is sufficient. From the oscilloscope output, the mean pulse height can be determined and substituted into (2.100).
This method has a far better accuracy than the previous ones, limited only by the inclusion of small pulses.

Measurement of first dynode gain

This was already covered on p. 71.

Dependence on temperature

The gain g decreases with rising temperature, probably due to increased phonon scattering of the secondary electrons, which reduces the SEY [173; p. 49]. While the temperature coefficient for the gain of each dynode is very small (about $1.67 \cdot 10^{-2}$%/K between -60 and 20 °C), the effect on the total gain g is greatly amplified since $g = \delta 1 \cdot \delta 2 \cdots \delta \mathrm{n}$, and amounts to approx. -0.2 %/K from -20 to $+30$ °C, for both CsSb and BeCu dynodes [170; pp. 66–67] [171].
This is still a small effect, which makes an accurate measurement difficult, especially considering that the temperature *also* changes the QE. [314] describes a method for independent measurement of the temperature dependence of both QE and gain via

photon counting: The QE change affects the count rate[79] while the change in gain shifts the peak in the SER or PHD, which is how both effects can be disentangled.

Dependence on voltage[80]

Dynode gain

The gain of a dynode can be described by a semi-empirical formula for SEY(E).
If no inelastic scattering losses occur, the SEY δ can be assumed to be proportional to the primary electron energy:

$$\delta = aV$$

where a is a constant.
Scattering losses can then be taken into account by an exponent $\alpha < 1$:

$$\delta = aV^{\alpha} \tag{2.101}$$

Of course this is only correct for the low-voltage range of fig. 2.10; for high voltages the primary electron will penetrate deeper into the material and the SE emission probability drops.
Measurements of [170] nevertheless show that (2.101) is a valid approximation below the maximum of SEY curve, even for d1. Wright obtained $a = 0.172$ for CsSb and 0.09 for BeCu, an $\alpha = 0.72$ for both materials, and for 100 V interdynode voltage a δ of 2.5 for BeCu and 4.7 for CsSb. The difference in stage gain between both compounds thus comes from the different values for a.

Total gain

An PMT with n dynodes requires a voltage divider with $n+1$ stages, since the anode has to attract the electrons emitted from the last dynode.
For a linear VD, which has uniform interdynode voltages V and a total voltage V_t, one obtains:

$$V = \frac{V_t}{n+1}$$

Using (2.101), the total gain g then follows to

$$g = \delta 1 \cdot \delta 2 \cdots \delta \mathrm{n} = a_1 \cdot a_2 \cdots a_n \cdot \left(\frac{V_t}{n+1}\right)^{\alpha_1+\alpha_2+\cdots+\alpha_n} = c\, V_t^{\beta} \tag{2.102}$$

$$\text{where} \quad \beta = \alpha_1 + \alpha_2 + \cdots + \alpha_n$$

$$\text{and} \quad c = \frac{a_1 \cdot a_2 \cdots a_n}{(n+1)^{\beta}}$$

c and β are constants. (2.102) shows that the gain–voltage dependence follows a power law. Note the rapid increase with voltage.
The power law function of an individual PMT[81] can be obtained by measuring g for at least two voltages and either solving the equation system produced by substituting the measured values in (2.102) (for two datapoints) or fitting the function.

[79] This might be further improved by also using the change in occupancy which alters the fraction of 2 pe pulses to spe pulses.
[80] Content based on [170; pp. 203–208]
[81] Stage gains vary between specimens, so this has to be measured for *each* PMT separately.

Power supply stability

It is crucial that the high voltage supplied to the PMT is extremely stable. From (2.102) follows [170; p. 319] [195; p. 186]

$$\frac{dg}{g} = \beta \frac{dV_t}{V_t} \tag{2.103}$$

So a 1% fluctuation in V, assuming that all $\alpha = 0.72$, results in a 7.4% change of g – for a typical HV of 1500 V this is no less than a gain change of 0.5%/V. Conversely, to achieve a gain stability of 1% the HV must be stable to within 0.14%. With modern ultra-stable power supplies this is not a problem anymore [170; p. 448]; however, it was in the past, which resulted in the recommendation to operate at the flattest point of the signal plateau (see p. 135) – nowadays, this advice can safely be ignored.

Number of stages [82]

It is crucial to select a PMT with an appropriate number of stages n for the respective application. A too low number n results in either insufficient gain, or, if the HV is raised to compensate for this, excessive DCR and AP rates, reduced lifetime and possibly breakdown of the device. A too high n on the other hand results in low interdynode voltages, which gives rise to poor linearity (δ is smeared by the low mean SEY, which roughly complies to Poisson statistics) and timing properties, since the drift times between stages is increased which also increases the TTS.

Dependence on dynode material

As can be seen from fig. 2.10 and the results of [170; pp. 203–208] quoted on p. 140, BeCu has lower stage gains for the same voltage than CsSb, which is why a higher number of dynodes or a larger total voltage V_t is required to reach similar total gains.

Dependence on light level

For the low light levels of photon counting depending on dynode material 10 to 14 stages are required, whereas for higher light fluxes, the output linearity (see 2.2.5) comes into play and limits the number of stages. In extreme cases, in high-energy physics experiments with high light levels, only three, two or even one dynode [83] are used [170; pp. 203–204].

Dependence on voltage divider

Tapered voltage dividers have a higher dynamic range compared to linear dividers (see p. 148), but also a reduced gain, typically by a factor of 3–5 [169; pp. 99–100]. The reason for this is that some stages are supplied with raised interdynode voltages, so the voltage between other stages is reduced. Since the SEY(E) levels out towards high voltages, this deteriorates the overall gain.
When using a tapered divider, the applied high voltage thus needs to be increased compared to a linear one to obtain the same gain.

[82] Content based on [171; pp. 22–23]
[83] Strictly speaking, a PMT with one dynode only is called vacuum phototriode instead.

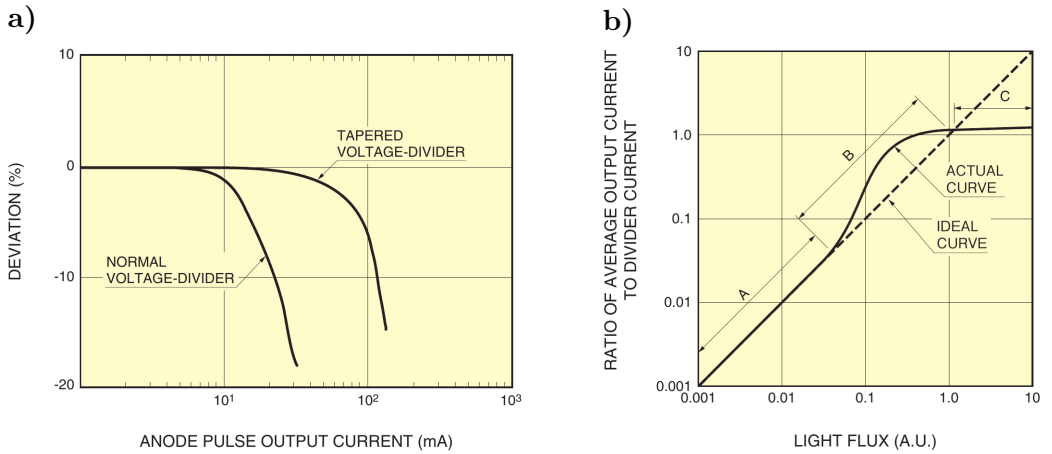


Figure 2.29: Output linearity of a PMT. Discussion see text.
a) Pulsed linearity (pulsed mode) for linear and tapered voltage divider (VD), from [169; p. 100]: The output current refers to the peak current at the pulse peak. Only negative deviations caused by space charge effects are observed, the linear and dynamic range are larger for the tapered divider.
b) Direct current (DC) linearity (current mode), from [169; p. 88]: After the linear range (A), a region of over-linearity (B) is observed due to voltage redistribution from the anode, before the output levels out and goes into saturation (C), which sets the dynamic range. The ordinate is plotted as fraction of the anode current I_a to the divider current I_d.

2.2.5 Linearity [84]

The linearity is the measure of proportionality between the number of incident photons and the resulting PMT output in form of charge, gain, height or current. At a certain photocurrent deviations from the ideal linear response will start to occur, whose magnitude will grow with increasing input (fig. 2.29a). Both a decrease or an increase of response can occur, but ultimately the output will saturate (fig. 2.29b), which limits the maximum obtainable output.

Linear and dynamic range

The region where response is still linear and no notable non-linearities occur is called *linear range*, while the maximum possible output including the non-linear region is called *dynamic range*.

Non-linearity

The deviation δ_l from linear behavior can be obtained from

$$\delta_l(x) = \frac{y(x) - a \cdot x}{a \cdot x} \cdot 100\% \tag{2.104}$$

where x is the input value (light flux), y the PMT response (e.g. current), and a the linear proportionality constant between both variables obtained from a fit in the linear range. When δ_l is plotted over x, this results in a plot as in fig. 2.29a.

[84] Content based on [169; pp. 56–57, 86–103] [171; pp. 10, 23–24]

Linearity values from datasheets

In manufacturer's datasheets often the linearity is quoted in form of the peak currents [mA] at which certain deviations (e.g. 3, 5 or 10%) occur, measured on an oscilloscope. The peak current is the current flowing at the time of the pulse peak and can be converted to pulse height or charge if the pulse shape is known. If this data is not available, the numbers quoted in PMT datasheets are of little use for photon counting applications aside from estimates using the pulse shape approximations in eqs. (2.84)–(2.87) or (2.192).
Measurements undertaken by physics experiments in this respect are usually more helpful as they typically quote the linearity in pe.

Operating range

For photon counting one wants to operate in the linear range, but in principle if the linearity of the PMTs is known and tabulated, the charge of a pulse can be reconstructed – with an error growing with the deviation δ_l – during a large fraction of the dynamic range.

Anode and cathode linearity

Deviations from linearity are primarily caused in the stages near the anode (anode linearity), but for large photocurrents or low temperatures also the cathode linearity plays a role due to the limited conductivity of the cathode, which also depends on photocathode size. For the light levels and temperatures considered here, this plays no role, however, so in the following the focus is put on anode linearity.

Pulsed and direct current linearity

Linearity is relevant both for pulsed mode, i.e. photon counting (pulsed linearity), and direct current operation (DC linearity). However, there are different causes limiting the performance for both types of linearity.

The linear and dynamic range are actually defined not only by an upper, but also by a lower limit of detectable input (number of photons or flux), which for both linearity types is set mainly by the DCR resp. the dark current (see pp. 91 and 123, and 2.5.1).

The upper limit for pulsed linearity mainly derives from space charge buildup due to high current densities of the SEs emitted from the last dynode, if basic recommendations for voltage divider design are observed.
For DC linearity the upper limit is set by the voltage divider configuration (divider current), since in the last stages large currents are withdrawn from the divider chain, which leads to interstage voltage (V_{dd}) redistribution.
Due to the different effects, for the DC mode there is a region of initial over-linearity ($\delta_l > 0$, the output is paradoxically *larger* than for normal operation) before δ_l drops below 0 and saturation occurs, whereas for the pulsed linearity only a decrease ($\delta_l < 0$) is observed (compare fig. 2.29a and 2.29b).

In order to derive the reasons for this behavior, first the layout of voltage dividers (VDs) has to be discussed, since the VD type determines the anode linearity of a PMT (but also influences the cathode linearity).

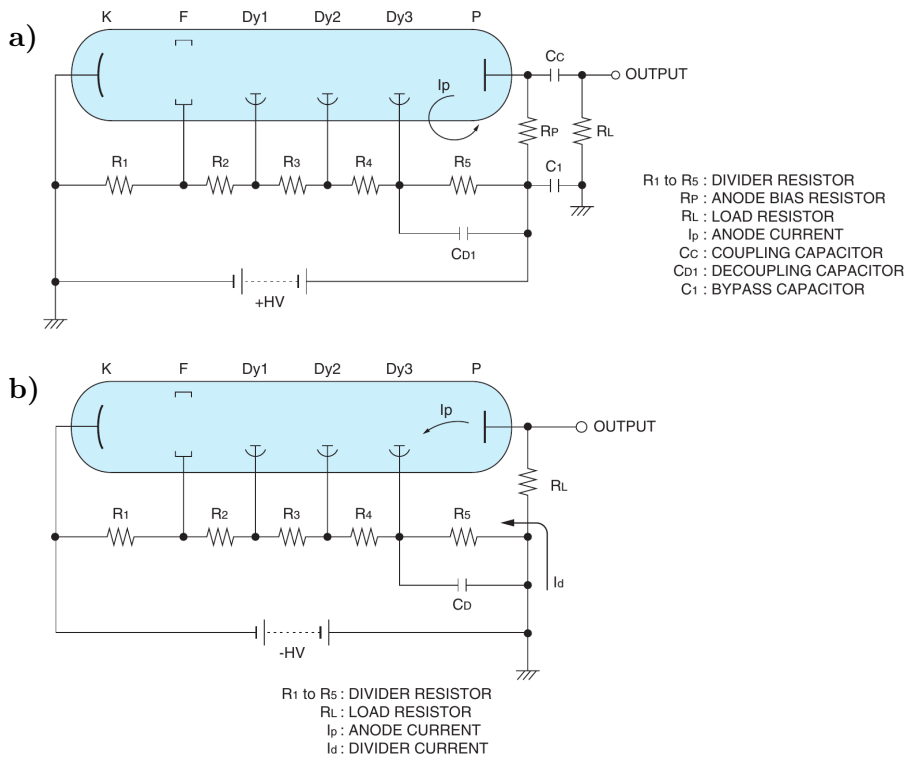


Figure 2.30: Schematic voltage divider layout:
a) Positive high voltage (HV): Cathode at ground, anode at positive HV; from [169; p. 87]. The signal must be separated from the HV through a coupling capacitor C_c.
b) Negative HV: Anode at ground, cathode at negative HV; from [169; p. 86]. Care must be taken to sufficiently isolate the cathode from the environment or else extreme DCRs can arise and the cathode may be damaged.

Dependence on voltage divider type

A voltage divider is a resistive divider network supplying voltages to the cathode, focusing electronics, dynodes, and anode in order to accelerate and focus electrons on d1 resp. the succeeding dynodes, and to collect the SEs emitted from the last dynode on the anode (see fig. 2.30 and 2.31). This configuration allows to only use a single power supply instead of one for every interstage voltage.
The divider is attached to the leads protruding from the PMT stem, which are connected to the respective parts inside the PMT, and largely affects the attainable performance of the PMT; not only with regard to anode and cathode linearity, but also for transit time, TTS, EP, LP, CE, ME, g, δ of all dynodes ($\delta 1$ is especially important), resolution, AP rates, DCR, and susceptibility to magnetic fields. This list actually contains *almost all* PMT properties, since most characteristics depend on the distribution of the total voltage between stages. This highlights the importance of using an optimized VD design.

Linear voltage divider

In the simplest case, the so-called linear voltage divider, the resistances between cathode, dynodes and anode (see fig. 2.30a) all have the same value R.
However, the linearity and timing of linear dividers is not optimal, as will become clear in the following, and they are mainly used for DC applications.

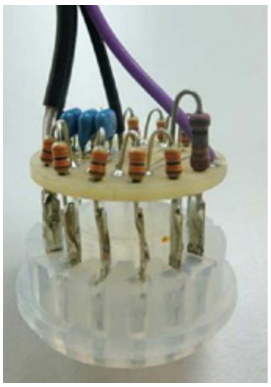


Figure 2.31: Photo of a voltage divider, from [169; p. 86]. Note the resistors and capacitors between the leads, which are plugged onto the pins emerging from the PMT stem.

Negative high voltage

Regardless of the polarity of voltage applied to the PMT, the anode must always be at a higher potential than the cathode so that the electrons of the cascade are drawn away from the cathode and towards the anode.
When using a negative HV, this must therefore be applied to the cathode, while the anode it connected to ground (fig. 2.30b). In this case one can directly connect electronics to the anode for measurement. However, care must be taken that no conductive grounded materials such as a metal holder, housing, or magnetic shield touch the cathode or come too close to it (usually a minimum distance of 4 mm is recommended [169] [170; pp. 459–460]). If this is the case, electrons in the PMT are accelerated towards this ground potential, strike the glass and can cause light emission which raises the DCR significantly. In addition a (small) current can flow between cathode and ground which can damage the photocathode (see p. 60). Only sufficiently good insulators may touch the window.

Positive high voltage

A positive voltage on the other hand is applied to the anode, whereas the cathode is grounded (fig. 2.30a). This configuration avoids the problems mentioned for negative polarity, and is required when the PMT is in contact with e.g. a scintillator. However, a decoupling capacitor is needed to separate signal from HV on the anode, and only pulsed signals can be extracted. This can cause baseline shifts for large count rates, and leakage currents in the capacitor can produce noise.

Origin of direct current non-linearity

The non-linearities of DC operation are connected to the divider current I_d, which flows through the resistor series even in the absence of a signal due to $I = V/R$. I_d equals approximately [85] the total HV V_t divided by the sum of resistances:

$$I_d \approx \frac{V_{ka}}{\sum_i R_i} \tag{2.105}$$

When there is no light incident on the PMT, the dark currents between and from the dynodes as well as on the anode typically are negligible compared to I_d.

[85] There are also leakage currents in the divider and the PMT.

Therefore the current through all resistors is identical and equals I_d, and the voltage distribution on the dynodes is as expected from the division ratios.

With incident light, the current which flows between dynodes and correspondingly into the dynodes from the resistor chain is not negligible anymore. Thus the current flowing through the resistors decreases with dynode number, as more and more current is branched into cascade production. The remaining divider current is smallest for the resistor between the last dynode dn and the anode a, as the full cascade flows from dn to a. Since $V = R \cdot I$, the interstage voltage V_{dd} drops progressively with dynode number compared to the value without signal. This voltage drop is redistributed to the other dynodes which raises their V_{dd} and δ. V_{k1} is raised most since it benefits from all other dynodes and its small photocurrent. V_{na} drops most but raises the overall gain, since it is only used to collect the SEs from dn and the redistributed voltage increases all other δn, which causes the over-linearity. If the photon flux increases that much that V_{na} approaches zero, the anode SE collection efficiency drops significantly, which results in the saturation of the output.

Choice of divider current and total resistance

This voltage drop at later stages is the reason for choosing a large I_d compared to the expected anode current I_a, as this directly affects the linearity, which can be improved by simply increasing I_d.
The divider current can be set through the resistance values R in the divider chain using (2.105). It can be calculated that for a target linearity of 1% I_a/I_d should be $< 1.4\%$. The exact ratio, for which 1% deviation occurs, however, depends on the applied HV and the dynode gains δ which vary from PMT to PMT, so for actual operation the value should be at least twice as low, leading to the recommendation [169; p. 91]

$$\frac{\bar{I}_a}{I_d} < 0.7\% \tag{2.106}$$

where $\bar{I}_a$ is the mean anode current. When I_a exceeds $I_d/10$, serious non-linearity will arise.
However, setting a high I_d also increases the power consumption, and the HV power supply possibly cannot provide the required currents. Furthermore, the dissipation also produces heat in the VD, and if this raises the temperature of the PMT it will lead to a higher DCR and reduce the stability. These difficulties are aggravated for high count rates and correspondingly large I_a. I_d can be reduced by incorporating Zener diodes or transistors in the last stages (see pp. 149 and 149).
Typical values for R range from 10 kΩ to 10 MΩ.

Origins of pulsed non-linearity

Voltage drop due to peak current

In pulsed operation, even if I_a is small the peak current at the pulse peak can be large, in extreme cases hundreds of times higher than I_d if a signal containing many photons is detected. In this case, the VD cannot supply sufficient *peak* current to the last dynodes which reduces voltages and thereby linearity. The benefit of gain increase observed in DC linearity is lost, since the pulse already passed the previous

stages where a voltage gain could be used, therefore this effect *decreases* gain and causes negative δ_l.

Space charge effects

This is not the only effect limiting pulsed linearity, however. Even when the voltage drops due to peak load are compensated by using decoupling capacitors (see below), at some point output saturation occurs, because the high electron density near the last dynode causes space charge effects on the dynodes which prevent the current flow to the anode. The SEs can, however, be effectively withdrawn from the last dynode if the extraction field strength at the dynode surface (from the potential applied to the anode) is high, and/or the dynode emission area is large (this lowers the charge density). This is the reason why e.g. Mesh-PMTs with their large surfaces and close distances (higher field gradients) have far better pulsed linearities [169; p. 57].
This effect will reduce the output charge for high fluxes. As this was also the case for the voltage drop induced by large peak currents, in pulsed operation no over-linearity but only negative deviations from linearity are observed.

Decoupling capacitors

If decoupling capacitors are connected to the last stages (fig. 2.32), they can supply sufficient interdynode current during pulse production and reduce the voltage drop between stages, efficiently eliminating one source of pulsed non-linearity. They can be connected to the anode either in serial or parallel; however, due to capacitor breakdown voltages the serial configuration is used more often.
The required capacitances can be calculated from the expected peak output voltage V_p by first determining the expected peak charge q_p using (2.86):

$$q_p = \frac{\Delta t\, V_p}{R_l}$$

where Δt is the expected pulse width (this is not necessarily a spe pulse width, e.g. for scintillators) and R_l is the load resistance between anode and ground.
For a linearity of the peak charge of better than 3%, the last capacitance C_n between dn and anode should be able to store a charge q_n for which holds [169; p. 97]

$$q_n \geq 100\, q_p.$$

From $C = q/V$ then follows

$$C_n \geq 100 \frac{q_p}{V_{na}} \tag{2.107}$$

To determine the required capacitances for the previous stages assumptions about their δ must be made. From this follows

$$C_{n-1} = \frac{C_n}{\delta \mathrm{n}}\,, \quad C_{n-2} = \frac{C_n}{\delta \mathrm{n} \cdot \delta(\mathrm{n}-1)}\,, \quad \dots \tag{2.108}$$

Usually, decoupling capacitors are used for the last three stages, but more can be added if required. For real VDs the capacitances should be larger by a safety factor of 10 than from these simple calculations. One should further bear in mind that this method can only maintain voltages during short timescales ($\ll R\,C$).

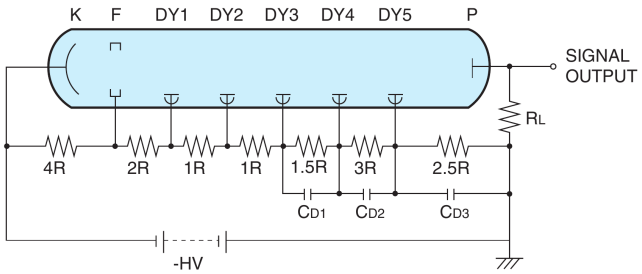


Figure 2.32: Tapered voltage divider layout, from [169; p. 101]. Tapering is used for the last as well as the first stages. Decoupling capacitors are included for the last stages to further improve linearity.

High count rates can lead to high anode currents I_a and in this case using decoupling capacitors alone is not sufficient for good pulsed linearity. In addition, as for DC operation, one should employ (2.106) and configure I_d to be at least a few ten times higher than I_a. As an alternative to using such high I_d, an active VD or the booster method can be used (see pp. 149 and 151).

Extraction fields

Regarding space charge effects, PMTs of any dynode type shows improved linearity if V_{dd} is increased since the extraction field rises. It can be observed that the linear and dynamic range grow if the total voltage V_t is raised [169; p. 57] [170; p. 388].

Tapered divider

It is, therefore, beneficial to increase the V_{dd} for the last stages to overcome the space charge effects by using larger resistances in the divider chain. This *tapered* voltage divider (fig. 2.32) results in a far better linearity (5–10 times larger linear range) compared to a linear VD where all resistances have equal values (fig. 2.29a).
However, the gain is reduced by a factor of 3–5 compared to a linear divider, because the other stages receive less V_{dd} and since SEY(E) levels out for large impact energies this deteriorates the overall gain. Thus the HV needs to be increased to compensate this. Also, the CE may be lower due to the changed voltage distribution if this is not considered in VD design, and care has to be taken not to exceed the breakdown voltage of the components placed between the last stages. Lastly, it is complicated to determine the resistance ratios yielding the optimum extraction fields.

R_{na} is usually chosen slightly smaller than $R_{n-1,n}$ due to the small distance from the anode to the last dynode, which allows to reach a voltage gradient sufficient for SE collection on the anode with a comparatively low V_{na}.

The voltages applied between cathode, focusing electrode and d1 affect the CE, ME and $\delta 1$ and through this the resolution, g, TTS, cathode linearity, and susceptibility to magnetic fields. These potentials are therefore crucial for optimal operation and it is beneficial to use a tapered design for these stages as well. A high V_{k1} improves CE and $\delta 1$ and reduces shot noise for photon counting.
In order to keep these critical voltages fixed at the optimum values (depending on application), it can be reasonable to use Zener diodes in place of resistors (see also p. 149).

To give two examples, the front-tapered divider offered by Hamamatsu for the R12860 (20") uses 11.5 times the interdynode voltage of the last dynodes between cathode and d1. The new, fully-tapered divider for the 8" R5912 (E7694-03) uses even $16.8R$ between k and d1 [213] – this layout is heavily optimized (compare the half-tapered previous design E7694-01 [315]), and the standard divider resistance of $1R$ is only used for one stage anymore, between d5 and d6.

For tapered dividers, resistance values typically lie between 10 kΩ to 1 MΩ.

Low-pass filter

To limit the noise picked up from the power supply (ripple), a low-pass filter can be employed in the VD (not shown, see [169; p. 101]). In so doing, one should use a resistance which is small compared to the resistors in the divider chain so as not to reduce the HV supplied to the stages of the PMT.

Ringing reduction

Ringing in the anode pulse shape can be reduced by placing small damping resistors of 10–200 Ω directly on the supply line of the last one or two dynodes[86]. Otherwise, for small pulse widths the SE current emitted from the last dynodes can cause compensating currents to flow from adjacent dynodes, which can lead to resonances from the RC elements. The damping resistance should be as small as possible to not impair the time response, and care should be taken to use non-inductive resistors to avoid generating noise [169; p. 102] [316].

Zener diodes

Using Zener diodes instead of resistors for the last stages (fig. 2.33a) stabilizes the interstage voltages and improves linearity.
This, however, limits the adjustable range of the total voltage V_t, because for too low or too high V_t the ratios of the V_{dd} of non-stabilized dynodes to Zener-stabilized ones become imbalanced, which affects optimal PMT operation. Zener diodes also can generate noise if the supplied voltage or the current flowing through them is too low. This can be avoided by setting a high I_d or placing noise-absorbing capacitors in parallel, and the bandwidth of the processing electronics should be kept as narrow as possible.

Active voltage divider

In active voltage dividers field-effect transistors are used instead of resistors for the last 2–4 stages (fig. 2.33b). Their emitter is connected to the dynode, whose voltage shall be set. Since the emitter-base voltage drop is fixed to about 0.6 V and is basically independent of the flowing current, the base voltage determines the dynode voltage and can be set by another high-resistance resistor chain connecting the transistor bases.
This provides good linearity up to 60–70% of I_d for DC mode (fig. 2.34), since V_{dd} is not affected by interdynode currents. When I_a approaches I_d, however, the over-linearity and saturation occur within a shorter region than for a purely resistive divider and in this range the gain changes are therefore stronger.

[86] An example configuration is shown in fig. 5-29 in [169; p. 102].

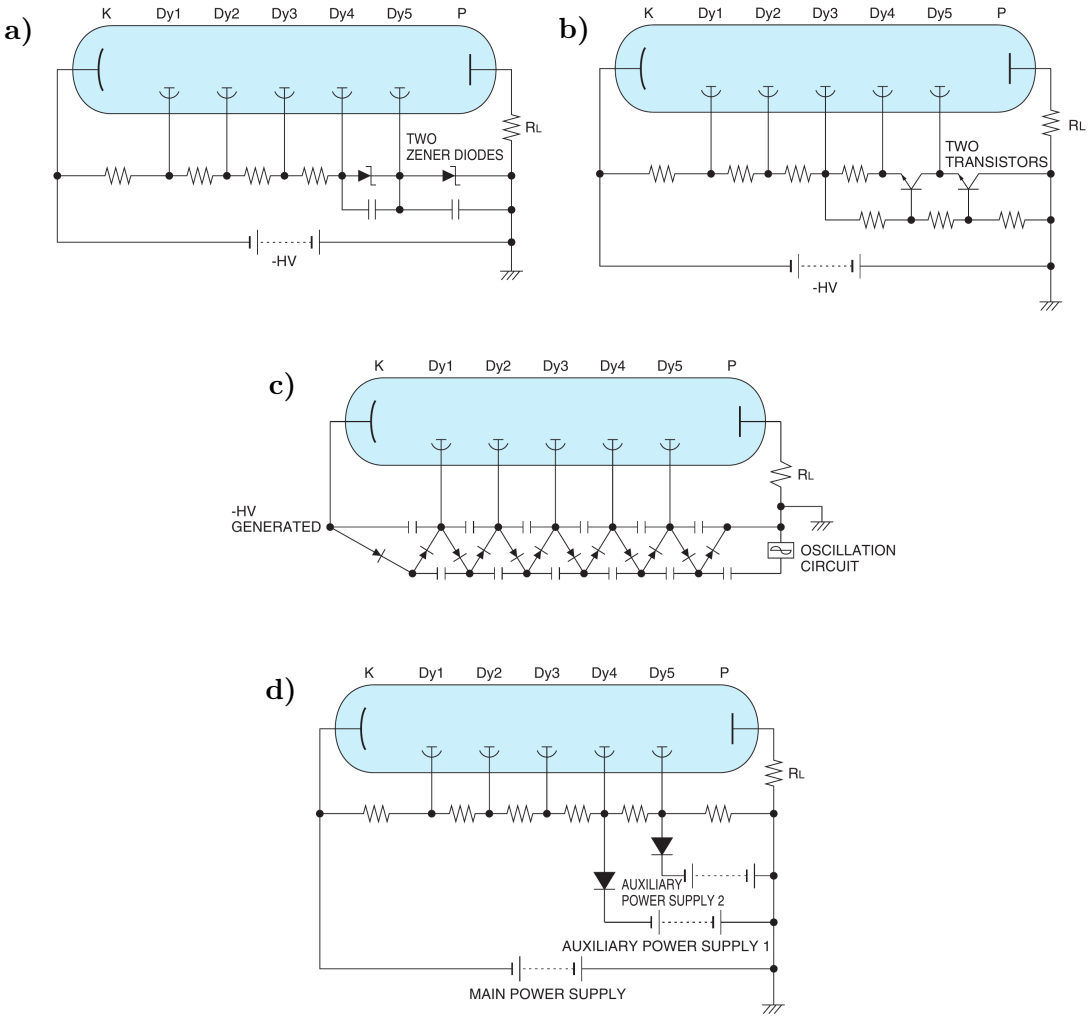


Figure 2.33: Voltage divider circuits, from [169; pp. 92–95]:
a) Zener diodes with noise-absorbing capacitors at the last stages.
b) Active voltage divider with transistors at the last stages, whose base voltage is set by an additional resistors chain.
c) Cockcroft-Walton circuit of diodes connected in series, which are powered by an alternating current source.
d) Booster method using auxiliary power supplies for the last stages.

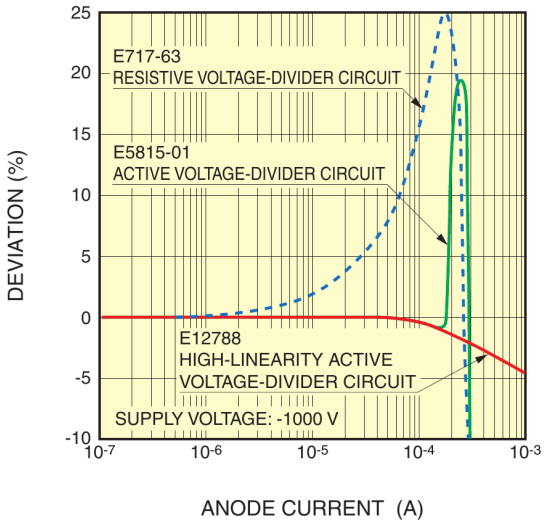


Figure 2.34: DC linearity of an active voltage divider and a high-linearity active VD compared to a resistive voltage divider. From [169; p. 96].

High-linearity active voltage divider

There is a high performance variant called high-linearity active VD, which uses transistors for *all* dynodes. In this case, I_a is not limited by I_d anymore, which results in excellent linearity (no over-linearity in DC mode) up to the maximum allowed anode current (fig. 2.34), and low power consumption of typically 100 mW since I_d is nearly zero, which also minimizes heat production.
There is one drawback, since the used power supply must be very stable to be able to deal with the fluctuating current, which is mostly determined by the momentary signal level.

Cockcroft-Walton circuit

A Cockcroft-Walton circuit (fig. 2.33c) consists of a voltage multiplier ladder network of capacitors and diodes connected in series, driven by an alternating current power source with peak voltage V_p, which results in a constant HV at the dynodes dn equal to $n \cdot V_p$.
This circuit has good linearity, very low power consumption and correspondingly low heat production. In addition, no HV power supply is needed.
However, the produced values are integer multiples of V_p, so it is difficult to produce a non-linear VD this way. Also, care must be taken to avoid coupling noise from the alternating current power source into the output.

Booster method

A straightforward approach to improving the VD linearity is by using auxiliary external HV power supplies to directly supply the voltages for the last dynodes, which draw large currents (booster method, fig. 2.33d). That way the voltage – and thus the linearity – is maintained independent of current flow even for pulsed operation, as long as the respective power supply can provide the required current. This method is mainly used for large pulses and high count rates such as in high-energy physics. It provides good linearity, but requires multiple power supplies per

PMT which is costly, and when a change of HV is desired, all power supplies must be adjusted accordingly.

Dependence on dynode material

As discussed on p. 141, BeCu has a lower gain δ than CsSb for the same interdynode voltages V_{dd}, which requires to use more stages and/or higher V_t to reach the same g. On the other hand, BeCu allows the application of higher interdynode voltages V_{dd} before disturbing effects become strong, which reduces the transit time and the TTS – this is offset by the higher required n, however. BeCu also has a better pulsed linearity than CsSb at the same V_{dd} [171; p. 10], but this shall be critically examined in the following.

Comparing two large PMT types with BeCu (ETE 9823, 5" [317]) and CsSb dynodes (ETE 9854, 8" [263]), the following can be found:

- The 9823 has 14 linear focused BeCu dynodes, a plano–concave window, $g_{\text{typ.}} = 8 \cdot 10^7$, a TTS (FWHM) of 2.4 ns, 55 ns transit time , $\text{HV}_{\text{typ.}} = 2400\,\text{V}$, rise time 2.7 ns, pulse width 3.6 ns, DCR 1.5 kHz, and a pulsed linearity using a tapered VD of -5% deviation at 150 mA. This yields $\overline{\delta} = 3.67$ and $\overline{V}_{dd} = 171\,\text{V}$.

- The 9354 has 12 linear focused CsSb dynodes, a hemispherical window, $g_{\text{typ.}} = 7 \cdot 10^6$, a TTS (FWHM) of 2.7 ns, 65 ns transit time, $\text{HV}_{\text{typ.}} = 1300\,\text{V}$, rise time 3.5 ns, pulse width 4.5 ns, DCR 4 kHz, and a pulsed linearity using a tapered VD of -5% deviation at 100 mA, which results in $\overline{\delta} = 3.72$ and $\overline{V}_{dd} = 108\,\text{V}$.

- Both PMTs have virtually the same δ due to the adapted, higher V_{dd} for the 9823. The smaller TTS for the BeCu PMT can be attributed to the smaller transit time (despite the higher number of stages) caused by a significantly larger HV and a slightly smaller cathode time spread (CTS, see p. 179) due to the smaller diameter. The DCRs are similar when scaled with diameter. Most important, the peak current at 5% linearity is higher for BeCu.

Regarding the linearity, it must be taken into account that the gain of the 9823 is far higher, which reduces the linear range if measured in pe instead of mA. Also, the pulse shape should be narrower for BeCu due to the shorter transit time and the smaller TTS, so a higher peak current is expected for the same output charge. In addition, due to the higher interstage voltages the extraction field is stronger. So the higher peak current of BeCu at a fixed δ_l can be at least partially explained by raised interdynode voltages.
Nevertheless, [171; p. 10] states that peak anode currents for $\delta_l = 5\%$ at *identical* interdynode voltages are higher for BeCu (at $V_{dd} = 300\,\text{V}$ for linear-focused dynodes 100 mA for CsSb and 150 mA for BeCu). This is in tension with [170; p. 387], who observed that there is little difference in linearity for PMTs with both dynode materials when using the same HV and I_a (which requires higher light levels for BeCu).
This discrepancy might be resolved when considering that for the same V_{dd} the dynode gain δ is lower for BeCu and the pulse is expected to be more statistically

smeared in time, i.e. broader. This lowers the peak charge. However, it also might improve SE extraction dynamics since the charge cloud is emitted over longer times, and so the mean space charge produced by the cascade could be lower. Furthermore, CsSb has a far larger Z_{eff} than BeCu (see p. 73), so for equal δ a larger fraction of SEs is backscattered, which also should contribute to space charge effects.
From statistic effects no large influence is expected, since the charge variation (which could affect the linearity) is mainly set by the first stages and quickly converges (see p. 100). This would in fact favor the linearity of the 9354 CsSb PMT due to its higher V_{k1} and $\delta 1$.

The present author thus comes the conclusion that the the increased peak current linearity of BeCu compared to CsSb is in large part – if not completely – due to the higher applied voltage. A small remaining difference could be due to space charge effects from the charge cloud of backscattered electrons, and possibly also a higher conductivity of the BeCu dynode substrate compared to the usually employed stainless steel or nickel for CsSb, which might improve current supply to the dynode surface and reduce space charge effects.

Measurement methods [87]

The pulsed linearity can be measured using an LED (light emitting diode), which alternately illuminates the PMT with a low and a high amplitude pulse (double-pulsed mode) through a diffuser. The LED driver pulses are adjusted such that the PMT output pulse charges (or heights) differ by e.g. a factor of 4 at very low light levels, which can be achieved through either NDFs or a large distance from the PMT. The initial charge ratio $\eta_i = q_{h,i}/q_{l,i}$ is determined with e.g. an oscilloscope or an FADC. Following this, the light level is stepwise increased through changing the NDFs or reducing the distance between LED and PMT, and the corresponding charge ratios $\eta = q_h/q_l$ are measured. At some point, η will differ from the initial value due to non-linearities. The deviation from linearity δ_l can be obtained from (compare (2.104))

$$\delta_l = \frac{\eta - \eta_i}{\eta_i} \cdot 100\% \tag{2.109}$$

and the values can be plotted over q_h, resulting in a curve as in fig. 2.29a. Knowing the exact ratios of flux increase between measurements is not required, since the measured charge determines the position on the abscissa.

Other methods to obtain the pulsed linearity can be found in [172; pp. 4-28–4-32] [170; pp. 384–408].

[87] Content based on [169; pp. 59–60] [171; p. 10]

2.3 Time response

2.3.1 Transit time

The transit time T (sometimes also called transient time) is the time between photon arrival at the photocathode and the resulting anode pulse leaving the output.

Stop signal

In order to determine the time difference T, both a start and a stop time have to be measured. For this it is necessary to select a reference point on the pulse shape (see 2.4) which defines the time of pulse output, i.e. the stop signal.

Definitions

Three definitions are commonly used (compare fig. 2.49): the peak of the pulse, its centroid, and the point of 50% pulse height on the rising flank [170; pp. 353–354]. The latter can be restated in a more generalized form as the point at a fixed height fraction on the initial flank. A specialization of this definition is to pick the point of the steepest slope on the rising flank.
While it initially appears irrelevant which definition is chosen, since all seem to be identical but for a fixed offset, this is only valid for ideal pulses without variations in height and width, and without noise. In reality there are marked differences in the resulting scatter of the obtained T values depending on which definition is selected. In the literature, a preference is found for choosing the peak as stop time [169; p. 52] [267], but all definitions have their pros and cons, which shall be discussed in the following.

Peak

Choosing the peak to characterize transit time is a natural choice.
However, the peak is by definition surrounded by a flat region, which makes is particularly susceptible to small voltage changes from noise spikes, which can strongly move the position of the absolute maximum and introduce a large systematic jitter. The author therefore suggests to use *any* other definition for measurement and then correct the offset between definitions using an average pulse shape (APS, see p. 339).

Constant fraction of height

A point on the rising flank, especially at the steepest slope, on the other hand is far less affected by noise jitter (discussion on p. 335), which helps to reduce systematic errors.

Centroid

The centroid is a logical quantity to assess transit time, as it describes the average time the electrons in the cascade need to reach the output. If the transit time between stages is analyzed instead, the centroid yields the average interstage transit time.

In order to determine the centroid, one has to record the pulse and calculate its mean time.
The advantage of this method is that the resulting value is determined from many datapoints and not only one or a few, which suppresses statistical variation from noise and in part also from the amplification at the first dynodes.
The disadvantage is its sensitivity to time shift effects (TSEs) which split the cascade, and early afterpulses (EAPs) sitting on the falling flank in form of pileup. As will be seen in 2.3.6 and 2.5.2.2, there is a multitude of effects which alter the pulse shape from event to event (see fig. 2.52), which explains the great diversity of photomultiplier (PMT) pulse shapes observed on an oscilloscope.
While in principle choosing the centroid as stop signal could allow to reduce the transit time spread (TTS), this is offset by the variation in pulse shape. A measurement based on pulse shape analysis which is able to recognize pileup could allow such an improvement, which would merit further studies. Selecting another integral-based quantity for the stop signal, such as e.g. the mean of all datapoints on the rising flank, might also be more stable than the centroid, but in this case care must be taken to recognize pileup on the rising flank, e.g. through a cut on the standard deviation of the included points.
Note that since the fall time is longer than the rise time, the centroid always lies *after* the peak (see fig. 2.49).

Comparison

In practice, it is always necessary to first determine if a pulse is present by setting a threshold to separate noise from signals. If a pulse is found, the discriminator sends a logical output pulse signaling the time of threshold crossing. Depending on discriminator type, this time either simply corresponds to when a fixed voltage (the threshold) was surpassed (leading edge discriminator, LED), or when a fixed fraction of the pulse height was exceeded (constant fraction discriminator, CFD). The first type is subject to time walk (see p. 335), the latter not, which is why CFDs are the better choice when measuring the transit time. Normally a trigger point on the rising flank is chosen (the first value exceeding the threshold), but in principle also a point on the falling flank can be chosen (the first value falling below the threshold) – the discriminator can distinguish between both from the polarity of the signal slope.
As one can see, a CFD automatically returns a stop time equivalent to the constant-fraction definition. The times corresponding to other definitions have to be obtained by recording and analyzing the pulse, or as mentioned by using an APS, but this introduces errors, since the variation in time differs between definitions and the APS does not compensate that. These, however, are smaller than the error of determining the peak time directly, and arguably smaller than the uncertainties introduced by pulse shape variations when using the centroid.
For the results presented in this work, therefore the constant fraction at the steepest point is selected as stop point, and the commonly used peak time is determined afterwards using the offset obtained from the APS.

Start signal

The start signal is equally as important, and ideally the photon arrival time distribution would have the shape of a delta function, as its dispersion cannot be separated

from the photosensor response [88] and fully enters the measured distribution of transit times (see 2.3.8). Therefore it is important to employ a single-photon source, whose photon emission time distribution is as sharp in time as possible, and which has a trigger output with minimal jitter.
Even if the photon *emission* time from the light source is known with great accuracy, in order to obtain the *arrival* time at the sensors, the delays introduced by photon flight time have to be taken into account, which can also add sources of jitter.

To calculate an absolute value of T, in addition the offsets between the start and stop signals introduced by transit times in the electronics have to be considered. This is discussed on p. 336.

Contributions

The transit time in every photosensor is > 0 due to effects which introduce delays in the pulse production. For PMTs these are the following, in order of occurrence:

1. **Photoemission delay:** Generally, it is assumed that photoemission from atoms occurs instantly. This is *not* the case [318, 319], due to e.g. Wigner delay or a tunneling time; however, the observed delays lie in the order of a few tens of attoseconds, which is negligible compared to the time scales relevant here.

2. **Photoemission lag:** Due to the finite values of the photocathode thickness and the photoelectron (pe) energy, the photoelectron production and emission does not occur instantaneously after the photon enters the cathode [170; pp. 322, 370, 373–374]. The delay t_{pe} is very small and lies in the order of ps. It increases with wavelength, as smaller photon energies translate to lower pe energies and drift velocities. [320] reports an increase to 100 ps at 860 nm. For wavelengths in the region of interest of this work, the delay is very low compared to the other contributions to T and can be neglected.
   The case is different for negative electron affinity (NEA) materials: Here, the photoelectrons typically thermalize to the conduction band bottom and thus have far slower drift velocities, resulting in photoemission lags of around 100–150 ps [203, 205].

3. **Photoelectron transport to the multiplier:** The drift time t_{k1} from the cathode to d1 [89] constitutes a substantial fraction of the total transit time due to the small extraction fields near the cathode, which only yield modest initial acceleration. For large PMTs the dimensions of the cathode further increase the drift distances and t_{k1} typically lies in the order of $T/2$.

4. **Amplification:** Secondary electron (SE) emission and drift to the next dynode introduce delays at all dynodes. The SE emission is similar in timescale to photoemission and this delay can usually be ignored. The drift time $t_{\mathrm{n(n+1)}}$ from one dynode to the next is small compared to t_{k1} due to the higher extraction fields and shorter distances between dynodes. Nevertheless, since

[88] Unless a deconvolution of the light source time distribution is performed, see p. 167.
[89] dn denotes the n-th dynode.

there are typically around 10 dynodes, for large PMTs the total time spent for amplification is about $T/2$.

5. **Anode readout:** Finally, the electron cloud produced on the last dynode dn is accelerated onto the grid-shaped anode a, where it induces a current spike and is collected. Since the distance between dn and a is small, the delay t_{na} introduced by readout is short, and therefore often ignored [170; p. 370].

6. **Signal transit to output connector:** The pulse produced at the anode still has to travel through the anode wire and voltage divider to the output connector, where the pulse is passed to subsequent electronics for processing. Since this electromagnetic pulse travels at the speed of light divided by the refractive index of the cable $c/n \approx 0.67c$ and covers about 20 cm per ns, the resulting delay t_{ao} is a fraction of a nanosecond and usually ignored when calculating T.

If all contributions to the transit time are summed up, one obtains

$$T = (t_{pe}+)\ t_{\mathrm{k1}} + t_{12} + t_{23} + ... + t_{(\mathrm{n}-1)\mathrm{n}}\,(+t_{\mathrm{na}})\,(+t_{\mathrm{ao}}) \tag{2.110}$$

In 2.3.4 formulas will be derived, which allow to calculate estimators for the respective quantities.

Typical values

In table 2.1 (p. 172) the complete range of the mean transit time values T_0 (definition in 2.3.2) for currently available Hamamatsu [90] PMTs is listed. The values range from merely 2.7 ns for a proximity-focused metal channel PMT (R9880U-210) up to 115 ns for a 5” PMT with plane window, box-and-grid dynodes and linear voltage divider (R877).
For modern large area PMTs with diameters of 5” or above the transit time ranges from 45 ns (5”, Hamamatsu R6594) to 95 ns (20”, Hamamatsu R12860) [213].

Transit time versus diameter

In fig. 2.35, T_0 is plotted vs. the cathode diameter for fast Hamamatsu PMTS (box-and-line or linear-focused dynodes) from 3” to 20” (R6091, R6594, R5912, R7081, R11780, R12860; see table 3.1 for values). A linear dependence on size is found (red. χ^2 1.33), which is exactly what is expected from theory (sec. 2.3.4). All values in the plot are for a total voltage $V_t = 1500$ V except for the R6091 and R12860 (both 2000 V). The higher voltage is expected to reduce T_0 in the order of 10% for both PMTs from theory. The T_0 offset of the fit function can be attributed to the dynode chain, which is essentially similar in dimensions for all PMTs. Altogether, the agreement of device values with scaling expectations from analytical calculations performed without the aid of a 3D electron-optics simulation is remarkable.

[90] Hamamatsu Photonics K. K.

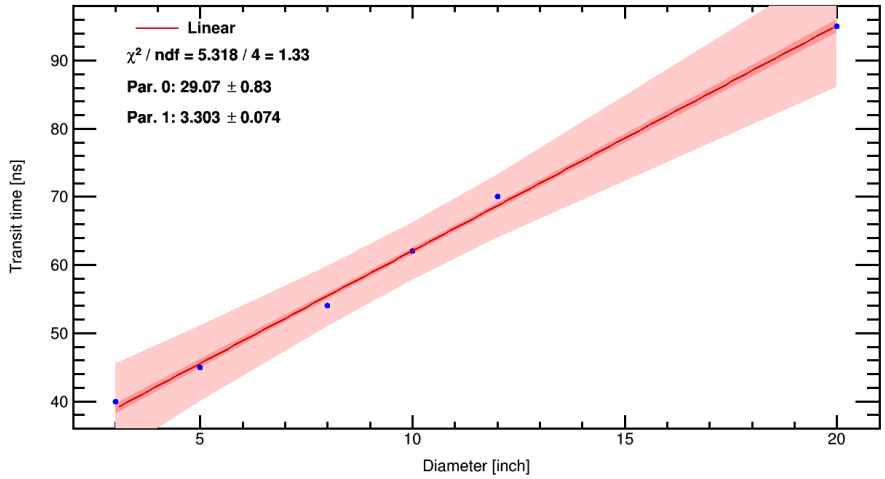


Figure 2.35: Mean transit time T_0 vs. diameter D for fast Hamamatsu PMTs from 3" to 20". Values from [213]. Fitting with a linear function $T_0 = m \cdot D + n$ gives good agreement with data (red. χ^2 1.33) for m =3.3 ns/inch and n = 29.1 ns. 1σ and 3σ error bars from fit in red, resp. light red.

2.3.2 Transit time distribution

The transit time distribution (TTD) shows the dispersion of the transit times of single-photon (spe) primary pulses due to statistical spread and time shift effects (fig. 2.36). A primary pulse (PP) is the first output pulse produced in response to a single photon. Subsequent pulses which can be produced through internal processes are called afterpulses (APs).
The TTD describes the photosensor time response to a delta function light source [170; p. 379]. Therefore the emission jitter of the used light source itself should be as small as possible to not artificially increase the jitter attributed to the sensor.

Effects

The most notable feature in the transit time distribution is the main peak (centered around 41 ns in the example). This consists of regular pulses (RPs) which undergo normal amplification in the PMT without any notable effects, just as delineated in the section on gain 2.2.4. The peak can be typically approximated with a Gaussian function close to its center, with deviations increasing with distance. The main peak constitutes the bulk of events and all other effects together merely represent a few percent of events in the TTD.
It is, therefore, the properties of this peak which mostly determine the uncertainty of reconstructing the photon arrival time. This jitter is the transit time spread (TTS), which is obtained from the full width at half maximum (FWHM) or standard deviation of the main peak.
The bulk of the pulses in the main peak consists of RPs, which undergo normal amplification in the PMT, i.e. a photon produces a pe in the cathode, which is subsequently multiplied on the dynodes. This pulse type is called γk in time shift theory (TST) (see 2.3.5.1 and 2.3.5.2) and its mean transit time is labeled T_0, the

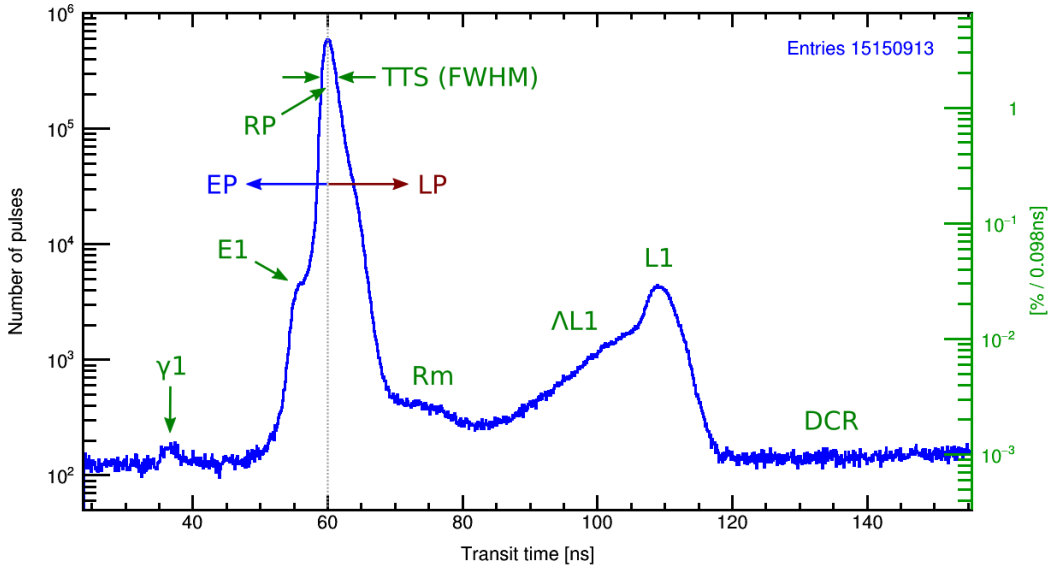


Figure 2.36: Example of a transit time distribution (TTD), measured for a Hamamatsu R5912 with front-tapered voltage divider (Hamamatsu E7694-01) at +1425 V, $g = 1.3 \cdot 10^7$, 0.2 pe threshold, illuminated with occupancy o =1.8%, classic measurement at the LNGS (see p. 457).
The dominant feature is the main peak ($\approx$ 94.6% of all pulses) with the regular pulses (RPs, ca. 59.9% of pulses), for which the transit time spread (TTS) is determined. Early pulses (EPs; 1.5%) occur before the center of the main peak (gray dotted line), late pulses after it (LPs; 2.9% after peak, 37% when subtracting a fitted Gaussian).
Some notable time shift effects (TSEs) are marked, a detailed discussion of the TTD follows in 2.3.6. E1 occurs a few nanoseconds early (0.7% of pulses), γ1 is early by about the photoelectron drift time t_{k1} from the cathode to the first dynode. Since both event types have underamplified charges, the size of their populations strongly depends on threshold; at 0.2 pe threshold the γ1 pulses are just barely visible (0.011%). L1 (1.5%) occurs ca. $2\,t_{k1}$ later than the RPs and has about normal charges. The range between RPs and L1 is mostly filled by ΛL1 (1.0%) whose charge decreases towards earlier times. Rm is also underamplified and strongly suppressed.
Random triggers from dark counts (DCR, ca. 1.3% of pulses) occur during the complete acquisition window and set a background floor covering more seldom effects.

regular transit time. Some effects create pulses through the same processes as for γk and with the same mean transit time T_0 but with reduced charge, since not all of the impact energy on e.g. d1 is used to create SEs (such as for BE1 and BL1). These pulses are considered (undersized) RPs as well.
There are also classes of events, however, which are centered around transit times T different from T_0. Their time shift $T - T_0$ compared to the regular pulses can be negative, when the pulses arrive earlier at the anode than the RPs. This event class is thus termed early pulses (EPs). When the time shift is positive, i.e. the pulses leave the PMT later than T_0 they are labeled late pulses (LPs). The absolute value of a negative time shift is called earliness, a positive time shift is dubbed lateness. The time distribution of many of these event types can be approximated by a normal distribution as well. The physical origin of these effects will be discussed in detail in 2.3.5 and 2.3.6.
Finally, there is a contribution from dark counts, which is uncorrelated to the photon source and therefore produces a constant background, which covers all smaller effects.

Number of pulses for an effect

For an individual effect (e.g. L1, see 2.3.5 and p. 316), the number of events n in the TTD can be calculated from the effect probability p that a PP in the TTD belongs to that effect and the likelihood $p_d(0)$ that no dark noise (DN) pulse is present before it in the same waveform (Poisson statistics with mean λ_d, see (2.12)) [276]:

$$n = n_l \cdot o \cdot p \cdot p_d(0) = n_l \cdot o \cdot p \cdot e^{-\lambda_d} = n_l \cdot o \cdot p \cdot e^{-f_d \Delta t_T} \tag{2.111}$$

where n_l is the number of light source triggers, o the occupancy, f_d the dark count rate, and Δt_T the length of the acquisition window up to the time T of the effect. Only dark count pulses which lie before a pulse of the studied effect prevent it from being detected as PP. To obtain the number of pulses of the effect, for which the waveforms contain no dark counts (also after the event), substitute Δt_T by the total acquisition window Δt_g.
The DN correction factor (last term) is usually very close to one, except for very high dark noise rates (e.g. silicon photomultipliers, SiPMs) or long waveforms (ionic afterpulse (iAP) measurements for PMTs).

Required number of pulses for target accuracy

The TTD is obtained by filling the measured transit times of pulses in a histogram until the desired level of detail is reached. The required number of recorded events for this can be estimated from Poisson statistics.
For an effect constituting a fraction η of all events, corresponding to n_η events, which shall be determined with a target statistical error of δ_η, let n be the total number of events required to achieve this. From Poisson statistics follows

$$\delta_\eta = \frac{1}{\sqrt{n_\eta}}$$

Combining this with

$$n_\eta = n \cdot \eta$$

leads to

$$n = \frac{1}{\eta \cdot \delta_\eta^2} \tag{2.112}$$

For an 1% effect which shall be measured with 1% accuracy, this results in $n = 10^6$, so for observing small effects with good accuracy very high statistics are required.

If the recorded number of events is fixed, the obtainable precision of an effect with probability η can be determined from

$$\delta_\eta = \frac{1}{\sqrt{N_\eta}} = \frac{1}{\sqrt{\eta \cdot N}} \tag{2.113}$$

For the same $N = 10^6$, a 0.1% effect could still be determined with an accuracy of 3.2%, a 0.01% effect only with 10% uncertainty. Looking at fig. 2.36, effects with 0.01%

probability can in fact be observed for PMTs with low dark count rates (DCRs) – e.g. $\gamma 1$ here has a probability of 0.011% after subtracting the DCR – and reveal a great deal of information about the occurring processes, so it is reasonable to assess the statistics required to observe them using (2.112) and determine the error of the measurement with (2.113).

According to [170; p. 369] about 100 pulses are already sufficient to reasonably determine the TTS. In order to resolve the TTD, however, far higher statistics are needed. From eq. (2.112) follows that to observe the most dominant effects (E1 and L1, probabilities $\mathcal{O}(1\%)$) with reasonable error (10%), at least 10^4 pulses are required. To measure the fine structure of the TTD, at least 10^6 acquisitions are recommended.

Dark noise floor

The features in the TTD, however, cannot be resolved with arbitrary precision if statistics is improved by longer data taking. This is due to the random background of dark counts which constitutes a signal floor. Luckily, this background level can be tuned by adjusting the occupancy using (2.24) and (2.15), at the cost of increased 2 pe contaminations (confer (2.18)).
Dark counts are not only caused by the PMT itself. Light leaks or unwanted light sources inside the setup can lead to uncorrelated photon background ("dark photons", so to speak). These contributions can become larger than the PMT-intrinsic dark noise. Thus it is mandatory to render any photosensor setup light tight, for example by furnishing the lid with photon traps or covering it with high-grammage molleton, taping cable-lead-throughs, and filling slits with black silicone. These measures can be tested by the dark-accustomed eye of the experimentalist (if the setup is large enough to allow this) or with a PMT operated at a voltage where dark noise pulse can just be counted in combination with a flashlight. A quick test can be performed by varying the integral outside illumination of the setup (by turning the lights on and off or drawing the curtains) while observing if this introduces a change in the dark count rate.

Number of dark noise events

The number of dark noise pulses n_d recorded in the TTD follows from the probabilities that in response to incident light no pulse is produced but one dark count pulse occurs (using (2.13)), multiplied by the number of trials, i.e. the number of incident light pulses n_l [276]:

$$\begin{aligned} n_d &= (1-o)\cdot p_d(1)\cdot n_l \\ \text{where}\quad p_d(1) &= \lambda_d \cdot e^{-\lambda_d} \\ \text{with}\quad \lambda_d &= f_d \cdot \Delta t_g \end{aligned} \tag{2.114}$$

f_d was the dark count rate, Δt_g the acquisition window length.

This formula, however, does not account for the events where both a DN pulse and a correlated pulse lie in the acquisition window. In this case, the DN pulse enters the TTD as primary pulse if it occurs earlier, which yields an additional

contribution $n_{d,\mathrm{PP}}$ to the total number $n_{d,t}$ of DN pulses in the TTD. This can be calculated by integrating the differential contributions within the time dT over the complete acquisition window from $t_{g,l}$ to $t_{g,h}$. The contribution within dT equals the probability $p_{\mathrm{PP}}(0) = e^{-s(T)}$ that no PP occurs before, where $s(T)$ is the integral over the TTD up to T, times the probability $p_d(1+)$ that one or more DN pulses are recorded within dT, which reads

$$p_d(1+) = 1 - p_d(0) = 1 - e^{-f_d dT} \overset{dT \to 0}{=} f_d \, dT$$

$p_{\mathrm{PP}}(0)$ also has to be normalized to the fraction $e^{-s(t_{g,h})}$ of the TTD lying within the acquisition window. Combining all of the above leads to:

$$\begin{aligned} n_{d,\mathrm{PP}} &= o \cdot n_l \cdot \int_{t_{g,l}}^{t_{g,h}} \frac{e^{-s(T)}}{e^{-s(t_{g,h})}} \, f_d \, dT \\ \text{where} \quad s(T) &= \int_{t_{g,l}}^{T} \mathrm{TTD}(t) \, dt \\ \rightarrow n_{d,t} &= f_d \, n_l \cdot \left((1-o) \cdot \Delta t_g \cdot e^{-f_d \, \Delta t_g} + o \cdot \int_{t_{g,l}}^{t_{g,h}} \frac{e^{-s(T)}}{e^{-s(t_{g,h})}} \, dT \right) \end{aligned} \tag{2.115}$$

The correction through $n_{d,\mathrm{PP}}$ is small compared to n_d. For a PMT with 1 kHz dark noise, $o = 5\%$, $\Delta t_g = 100\,\mathrm{ns}$ (0 to 100 ns) and $n_l = 10^6$ results $n_d = 95.0$. Assuming the TTD as a delta function at $T = 50\,\mathrm{ns}$ leads to $n_{d,\mathrm{PP}} = 2.50$, for a boxcar function from 0 to 100 ns a value of 3.16 results. For a partially contained TTD (boxcar from 50 to 150 ns) the value increases to 7.39, since in the first half of the recording window only DN pulses can trigger acquisition.

For the delay distribution of afterpulses (DD), the number of contained DN pulses $n_{d,\mathrm{AP}}$ in a window Δt_g can be estimated from the mean number of dark counts per window ($f_d \, \Delta t_g$) times the number of primary pulses n_{PP}, without which no AP can be observed:

$$n_{d,\mathrm{AP}} = f_d \cdot \Delta t_g \cdot n_{\mathrm{PP}} \tag{2.116}$$

Since dark counts can only be recorded as APs if a PP has occurred before, analogous to $n_{d,\mathrm{PP}}$ a corrected formula for all DN counted as APs in the waveforms can be derived to

$$n_{d,\mathrm{AP},t} = n_{\mathrm{PP}} \cdot f_d \cdot \int_{t_{g,l}}^{t_{g,h}} \left(1 - \frac{e^{-s(T)}}{e^{-s(t_{g,h})}} \right) dT \tag{2.117}$$

where the result of the integral is the effective window.

The number of dark noise pulses per bin in a histogram (TTD or DD) can be obtained by calculating $n_{d,t}$ or $n_{d,\mathrm{AP},t}$ and dividing it by the number of bins in the trigger gate.

Fitting the dark count rate

However, the exact length of the gate usually has some non-trivial uncertainties. At the gate brink, the acceptance does not abruptly drop to zero, and large pulses can still trigger acquisition. Furthermore, the DCR is a function of time. For PMTs it can notably decrease with the time in the dark, especially shortly after illumination

by room light, and even after long times the DCR still fluctuates due to gain drift. Therefore, while the above equations are helpful, it is usually better to fit the dark noise in an (unshadowed) distribution (see p. 164), since this automatically takes into account the true average values of f_d and Δt_g. This is described on p. 350.

Dark noise subtraction

Conveniently, if the acquisition window is chosen large enough to accommodate a short time range before the $\gamma 1$ events, the mean dark count rate present during measurement is automatically measured as well (in form of the dark count entries per bin). This allows to obtain and subtract the mean DCR without an additional measurement and the uncertainties of comparability introduced hereby. The mean dark noise entries per bin for an unshadowed TTD (p. 164) can be determined from a fit on the dark count floor. Note that unless unshadowing is performed on the TTD first, the fit function is *not* a constant (see (2.130)), as it might seem at first glance. The error on the signal bin content, which results after dark noise subtraction can be calculated from error propagation, assuming the fit error of the noise bin content is negligible. The signal bin content b_s can be reconstructed by subtracting the dark noise bin content b_d from the measured counts b_m in a bin by

$$b_s = b_m - b_d$$

The standard deviation σ_s of b_s can be obtained through error propagation from the standard deviations σ_m and σ_d of measured data and noise assuming normal distributions:

$$\sigma_s = \sqrt{\sigma_m^2 + \sigma_d^2}$$

Since for a bin count A according to Poisson statistics $\mu_A = A$ and $\sigma_A = \sqrt{A}$, and $b_m = b_s + b_d$, it follows

$$\sigma_s = \sqrt{b_m + b_d} = \sqrt{b_s + 2b_d} \tag{2.118}$$

which for $b_s \to \infty$ becomes $\sqrt{b_s}$, while for $b_s \to 0$ the error $\sigma_s = \sqrt{2b_d} > \sqrt{b_d}$ since a noise measurement is subtracted from noise, which increases the error. Since

$$\sigma_m = \sqrt{b_s + b_d}$$

$\sigma_s > \sigma_m$; the error on each bin only increases with subtraction, as was already seen on p. 123.

Nevertheless, the subtraction is reasonable, as is intuitively clear, since it better reveals small features and reduces the error of dark noise contamination when determining the probability of an effect.
This can be seen on the example of a simple square-shaped signal sitting on top of a flat noise background. Let the bin content from signal, noise and their sum be b_s, b_d and b_m, and the width of the signal span w bins. The total number of events lying within the signal range for signal (n_s), noise (n_d) and the measured sum of both (n_m) then amounts to

$$\begin{aligned} n_s &= b_s \cdot w \\ n_d &= b_d \cdot w \\ n_m &= b_m \cdot w = n_s + n_d \\ \to \quad n_s &= n_m - n_d = w\,(b_m - b_d) \end{aligned} \tag{2.119}$$

If the total number of signal entries n_s is obtained through subtraction from (2.119), this has an error σ_{sub}. Using (2.119), Poisson statistics and (2.118), this follows to

$$\sigma_{\text{sub}} = \sqrt{\sigma_m^2 \cdot w^2 + \sigma_d^2 \cdot w^2} = w\sqrt{\sigma_m^2 + \sigma_d^2} = w \cdot \sigma_s$$

This can be compared to the error σ_{raw} on the signal entries, if no subtraction is performed. In this case the full number of noise events n_d must be added to the uncertainties resulting from counting fluctuations:

$$\sigma_{\text{raw}} = n_d + \sqrt{\sigma_m^2 \cdot w^2} = w \cdot b_d + w \cdot \sigma_m$$

Finally, the improvement of the signal error through subtraction can be assessed from the ratio

$$\frac{\sigma_{\text{sub}}}{\sigma_{\text{raw}}} = \frac{w \cdot \sigma_s}{w \cdot b_d + w \cdot \sigma_m} = \frac{\sqrt{b_s + 2b_d}}{b_d + \sqrt{b_s + b_d}} \tag{2.120}$$

For not too small b_d (above $\left(\sqrt{2}-1\right)^2 \approx 0.17$) this ratio is < 1, which justifies the subtraction of dark counts.
This also means that for low noise count in combination with low signal count, if

$$b_d < 0.17 \tag{2.121}$$

$$\text{and} \quad b_s \leq \frac{1}{4}\left(b_d^2 - 6b_d + 1\right) \tag{2.122}$$

the subtraction *increases* the error and is *not* justified. However, this will seldom be the case, as the measurement can be simply prolonged until (2.121) or (2.122) is no longer valid.
As one would expect, if no noise is present ($b_d = 0$) or $b_s \to \infty$, the ratio in (2.120) becomes 1, as subtraction yields no benefit.

Unshadowing

One problem, which always emerges, when only the first event is taken to fill a distribution, is that early effects shadow later effects.
This is best explained using an extreme example, while comparing with the TTD in fig. 2.36. For an occupancy of 99% there will be on average 4.6 detected photoelectrons produced per single light pulse. These photoelectrons will produce anode pulses statistically independent from each other, which is equivalent to random draws from the TTD. Now, if this would be used to measure a TTD, only the first pulse would be used to fill the distribution, since all others are indistinguishable from afterpulses and have to be discarded. That means that from the requirement to include only the first pulse, the early part of the TTD is favored over the latter part, which is equally represented when *creating* pulses, but not when *selecting* them. A bias is introduced which grows with transit time. For example, the latest percentile of the TTD is not assigned a probability of 1% anymore, but instead of $1\% \cdot (1 - 0.99 \cdot 0.99) \approx 0.02\%$ since for $o = 0.99$ in $0.99 \cdot 0.99 \approx 98\%$ of cases there was already a primary pulse present at an earlier time.
This argument still holds when scaling o down to usually employed values $< 10\%$, and introduces a bias towards earlier transit times. Since the shape of the TTD is not known a priori, the measured distribution is difficult to compare with expectations.

However, the effect becomes clearly visible for the normally flat dark noise background, which for growing o increasingly becomes sloped towards larger T. As mentioned, shadowing always occurs as soon as only the first events are selected and the rest is discarded. It is therefore also a problem when measuring the AP distribution, since this is often done using only the first observed APs, because otherwise one observes a trail of afterpulses created by afterpulses (and so on), which is statistically difficult to treat [288].

In order to remove the distortion of the TTD introduced by shadowing, the distribution has to be unshadowed. This procedure is described in [288] and [276] for afterpulses, but the treatment is exactly the same for primary pulses in the TTD and shall be summarized here.
The TTD is a histogram with bins. The probability p_i of measuring pulses in bin i stretching between transit times T_i and T_{i+1} can be obtained from the measured number of events b_i in the bin and the total number of light source triggers n_l and eq. (2.10) through

$$p_i = \frac{b_i}{n_l} = \frac{b_i \cdot o}{n} \tag{2.123}$$

$$\text{with} \quad \sum_i p_i = o \tag{2.124}$$

where n is the number of events in the distribution and o the occupancy.
Using Poisson statistics, p_i can also be written as the probability that no pulse occurred before T_i (see (2.12)) multiplied with the probability to observe at least one pulse within the bin (eq. (2.15)):

$$p_i = e^{-s_{i,t}} \left(1 - e^{-p_{i,t}}\right) \tag{2.125}$$

$$\text{with} \quad s_{i,t} = \sum_{j=1}^{i-1} p_{j,t} \qquad s_{1,t} = 0 \tag{2.126}$$

Here $p_{i,t}$ is the true, unshadowed bin probability, and $s_{i,t}$ is the true summed probability of all pulses up to T_i.
$p_{i,t}$ and the corrected number of bin entries without shadowing $b_{i,t}$ can then be obtained from (2.125) and (2.123) to

$$p_{i,t} = -\ln\left(1 - p_i \cdot e^{s_{i,t}}\right) \qquad p_{1,t} = p_1 \tag{2.127}$$

$$b_{i,t} = p_{i,t} \cdot n_l = \frac{p_{i,t} \cdot n}{o} \tag{2.128}$$

$p_{i,t}$ and its sum $s_{i,t}$ can be calculated iteratively starting at the first bin.

For afterpulses instead of n_l the number of primary pulses n_{PP} should be used. This is equivalent to replacing o with the probability $p_{\mathrm{AP}}(1+) = \frac{n_{\mathrm{firstAP}}}{n_{\mathrm{PP}}}$ to observe one or more APs (see 2.5.2.1 for the definition of AP probability).

After the unshadowing is performed, the TTD (or DD) can be processed in the usual way, that is, by fitting and subtracting the now flat dark noise background.

Shadowed dark count rate fit function

Not only photon pulses, but also dark count pulses contribute to shadowing since both are contained in the TTD. The unshadowing, however, takes this into account automatically.

The influence of the dark count rate r_d on itself, i.e. without signal, can be calculated assuming a constant rate r_d and Poisson statistics [170; p. 184] [321]. The probability to observe n dark counts within the time T is

$$p_d(n,T) = \frac{(r_d\,T)^n}{n!}\,e^{-r_d\,T} \tag{2.129}$$

from which the likelihood to observe no events follows to $e^{-r_d\,T}$. Since the probability to observe a pulse within dT is $r_d\,T$, the probability to observe no pulse within T and one pulse within T and $T+dT$ is

$$dp_d(T) = e^{-r_d\,T}\cdot r_d\,dT \tag{2.130}$$

This can be directly applied to a histogram with bin width dT, which will show a curve following $r_d \exp(-r_d\,T)$.
This occurs for example in the TTD after the time shift effects, or in the delay distribution of the first ionic afterpulses after the afterpulses.

Coincidentally, eq. (2.130) also describes the distribution of the time between subsequent dark count pulses (or any effect with a constant but random rate), which shows a peak at $T = 0$, meaning that the most frequent interval is actually zero [170; p. 184]. This is readily observed in the waveform of a PMT without incident light in form of pulse clustering.

Afterpulse contamination

Early afterpulses

If the instrumental threshold is too high (this can already be the case for 0.1 pe), small primary pulses can be missed, in which case larger early afterpulses can be mistakenly detected as PPs and show up in the TTD. This is especially frequent for BL1 (see 2.3.5 and p. 318), for which an anti-correlation between the charges of PP and AP occurs due to cascade branching.

Ionic afterpulses

If the light pulse rate f_l is too high, afterpulses from primary pulses created during previous illumination cycles can extend into the next acquisition window. This has to be avoided by either setting a low f_l such that the period $T_l = 1/f_l$ is larger than the largest expected afterpulse delays, or by using an exclusion window. In case of the latter, only events are used for which no pulse occurred within a veto window T_v before the acquisition window, where T_v again should be larger than the maximum afterpulse delay.
While the iAP background is not tremendous, it can produce contaminations of the TTD in the order of 1% for unfavorable values of T_l.
Note that for very short T_l also EAPs can enter the next acquisition window.

Afterpulses lying in the TTD window can also be caused by dark noise pulses occurring before the window (all AP types). Normally this produces a flat background, since dark noise pulses are randomly distributed in time, and APs initiated by DN are regarded as just another constituent of dark noise.
This, however, only is valid, if the period between acquisition windows is large compared to the AP delay times. Otherwise this can lead to complicated non-constant background functions, which have to be calculated from the convolution of the AP delay distribution, which is cut at the light pulse period, with the acquisition window gate function (a step function with values 1 inside the acquisition window and 0 outside). This becomes even more complicated, if more than one acquisition window lies within the AP delay range. Once more, this leads to the recommendation of choosing a sufficiently large illumination period T_l.
It should be noted that the effect from DN-induced AP background is small – only a fraction of the DCR – and normally plays no role.
For determining the AP probability – a small effect of some percent size – DN causing afterpulses can be a problem and should be eliminated by using a veto window before the acquisition window [276].

Deconvolution of the light source emission time distribution

As an alternative to using a fast (and expensive) light source, one can measure the light source jitter with a fast photosensor and unfold it from the TTD[91].
Note that the true signal cannot be recovered perfectly, because the measurement of the light source response function introduces errors, too, not the least of which is the jitter of the photosensor employed for this purpose. Also, care has to be taken regarding the implementation of the deconvolution, as many algorithms are prone to oscillatory behavior introducing artifacts. Using a fast light source therefore yields superior results.

Deconvolution of the photosensor response function[92]

By measuring the TTD, it is possible to obtain the true photon arrival times in an experiment through deconvolution of the TTD from the signal. This is analogous to a background subtraction, since the finite time response of the sensor (background) is folded with the signal; no mere subtraction as for the DCR background in the TTD is sufficient.
To eliminate all timing uncertainties of the experiment and not only the ones from the photosensor, actually not the TTD, but the detector response function has to be measured, which includes the jitter from other components such as a light source, optics and electronics. The detector response function can be measured by replacing the measured object with a single-photon light source with very low jitter. Care must

[91] This is not the same as the deconvolution of the photosensor response function (which in fact *is* the TTD) from a signal consisting of many photons – discussed in the next paragraph – which is performed for each pulse and greatly suffers from low o. The light source deconvolution is performed *on* the TTD, and the occupancy with which the latter is obtained can be arbitrarily low.
[92] Content based on [170; pp. 439–440], which quotes [322].

be taken to use the same threshold for both measurements, since this can change the detection probability of transit time shifting effects tremendously (see 2.3.5 and 2.3.6). This method allows dramatic improvements in time uncertainty for certain cases but requires a signal with large statistics per photosensor and event to work – one cannot deconvolve a TTD from a single photon. It is thus of limited use for liquid scintillator neutrino detectors (LSNDs); only for the tracking of GeV events, this technique might improve performance slightly.

2.3.3 Transit time spread

For experiments, a constant offset between photon arrival on the sensors and pulse output due to a transit time $T > 0$, while possibly requiring the use of delay lines for compensation, usually is unproblematic. The jitter (the fluctuations in time) of this delay, however, is a different matter altogether, as this introduces errors on the reconstruction of event times, positions and tracks, and affects pulse shape discrimination.
The transit time spread (TTS) gives a measure of the variation of transit time of a photodetector for spe pulses and is thus the most important timing characteristic. It is defined as the FWHM of the main peak T_{FWHM}, its standard deviation σ_T, or, less commonly, its variance [170; p. 325].
The FWHM can easily be determined graphically, while the standard deviation requires calculation. For a Gaussian peak with standard deviation σ, the FWHM $\approx 2.35\sigma$, see (2.68). However, the distribution usually only can be approximated with a Gaussian function close to the peak, and then the simple relation from (2.68) is not valid and T_{FWHM} and σ_T have to be determined individually. The standard deviation is often obtained from fitting a Gaussian function near the peak center, which then gives lower values than the standard deviation of the complete main peak. T_{FWHM} is a more realistic estimate of the true fluctuations, as at half height of the peak non-Gaussian effects typically already set in and are partially considered in the FWHM.
In the datasheets of PMT manufacturers, T_{FWHM} is usually termed "single electron jitter fwhm" (ETE[93]) or "Transit Time Spread (FWHM)" resp. "T.T.S. (FWHM)" (Hamamatsu).
Care must also be taken to ascertain which definition is used in a publication, as this is often neglected to be mentioned and usually the lower σ_T value is then used to euphemize sensor properties. Normally, the definition can be recovered from common sense and expectable behavior, or graphically if a TTD is displayed.

If the TTS is determined from multi-pe events instead of spe pulses, the TTS is reduced statistically through averaging of the TTD to [173; p. 65] [197, 323]

$$\mathrm{TTS}_{\mathrm{npe}} = \frac{\mathrm{TTS}_{\mathrm{spe}}}{\sqrt{\lambda}} \tag{2.131}$$

where λ is the mean number of pulses per light pulse. This is also sometimes used to make the TTS of a sensor look smaller.

[93] Electron Tubes Enterprises Ltd.

The variation of transit time actually is exactly what the transit time distribution displays, and why its measurement is that important. However, the TTS is normally defined as the properties of the main peak *only*, and effects like late pulses and to a large part early pulses are ignored. When using the complete TTD instead, this has no impact on the FWHM, the standard deviation $\sigma_{T,t}$ of the complete distribution, however, is far larger than σ_T, since outlying effects enter quadratically. The average absolute deviation might therefore represent the mean fluctuation about the mean transit time more realistically.
Usually, however, T_{FWHM} or σ_T are used to represent the timing fluctuations caused by the photosensors, and possible problems introduced by EPs and LPs in form of early triggers or changes to the summed pulse shape have to be taken into account separately in the analyses of a detector. These effects can also be excluded to a point through higher thresholds or time cuts. Particularly the EP-types can often be ignored due to their small charges. This threshold-based discrimination, however, can fail for large occupancies for a single PMT, in which case a premature trigger can be caused by multiple EPs.

Contributions [94]

The following discussion for the moment sets aside the contributions from the various time shift effects – these are covered in detail in section 2.3.6 – to focus on the contributions which already cause transit time dispersion for *regular* pulses:

1. **Photon statistics:** As mentioned, the TTS scales with the mean number of concurrently detected photons as $1/\sqrt{\lambda}$. This follows simply from Poisson statistics (compare (2.30) [324], the multi-photon TTS is a time resolution). Incidentally, the same behavior is observed for the charge resolution, and this is the reason that the energy resolution of scintillators improves with rising energy.

2. **Photoemission lag:** The delay of pe emission from the cathode also introduces a jitter, since photoelectrons created near the inner surface will have shorter drift times until emission. Again, this is mostly relevant for long wavelengths and NEA cathodes [170; p. 370].

3. **Photoelectron transport to the multiplier:** The path length and thus the transit time between cathode and d1 depends on the starting position, initial energy and emission angle of the photoelectron [170; pp. 322–341, 370–373, 378–380].

    a) **Initial energy:** Since photoelectrons are emitted with varying energies, this introduces a fluctuation in the drift time to d1, because the extraction field and thus the acceleration near the cathode is small and a slight variation in initial velocity can already lead to a rather large change in drift time.

    b) **Cathode time spread:** The path length of a photoelectron from cathode to d1 and therefore its drift time depends on its starting position. This

[94] See [170, 197]

introduces a systematic variation in transit time depending on photon hit position, the so-called cathode time spread (CTS), which for large area PMTs or non-hemispherical cathodes can constitute the dominant contribution to the TTS.
For plano–concave or hemispherical cathodes, d1 is located at the center of curvature of the photocathode to obtain equal trajectory lengths from all points on the cathode and minimize the CTS. Varying initial energies and emission angles, however, lead to drift time variations even for perfectly identical path lengths from the complete cathode surface.
A plano–planar window is less costly, but the photoelectrons from the brink take far longer to reach d1 which increases the TTS.
From the existence of the CTS follows that in order to obtain the mean timing characteristics in a measurement, a complete and homogeneous illumination of the cathode is required to uniformly take into account all trajectories, e.g. from the front, or even better with varying incident angles. Alternatively, using the same illumination distribution as in the foreseen experiment might provide more appropriate values.

c) **Lateral momentum:** An emission angle different from the normal in combination with an initial energy > 0 causes electron trajectories to fan out even if starting from the same position (fig. 2.18). This leads to different hit positions on d1 (spot size), which causes variations in the drift time from the cathode to d1, as well as in the trajectory and transit time from d1 to d2.

Altogether, the variation of the transit time from cathode to d1 is the major contribution to the overall TTS. This is greatly influenced by high voltage and field configurations, as will be seen. The higher the field gradient at the cathode, the lower the jitter will be.

4. **Electron multiplier**[95]**:** Each dynode contributes to the transit time spread due to statistics and electron-optics. The sources of jitter for SE emission are the same as for pe ejection, namely emission delay, initial energies and angular spread, but the emission energy is higher (up to tens of eV and increasing with interdynode voltage [267], the mean energy is about 5 eV [170; p. 323]) and the angular distribution is Lambertian.
In order to minimize the jitter from the electron multiplier, the early stages are crucial, just as for the variation of the gain.
In the following, the contributions to jitter and their design countermeasures are discussed for linear focused dynodes; however, they are largely valid for box-and-line PMTs as well.

a) **Lateral momentum:** In order to minimize the pe hit area due to lateral momentum and improve collection efficiency (CE), the first dynode d1 is angled at 45° compared to the rest of the dynodes (fig. 2.15). This reduces path lengths differences. d2 and d3 are also different in shape and orientation than the later dynodes (compare fig. 2.15 with fig. 2.16 and 2.20) to equalize transit times and optimize collection.

[95] Content based on [170; pp. 337–339, 323]

b) **Extraction fields:** By design the photoelectrons are directed onto the part of d1 close to d3 (cf. fig. 2.20), since here the extraction field and the initial acceleration is larger. Note that due to the proximity of d3, the extraction field is determined mainly by d3 and *not* d2. This reduces the drift time between dynodes, since the interdynode transit time depends to a large degree on initial acceleration, and thus decreases statistical fluctuations in the drift time due to varying SE emission angles and energies.
This design principle is applied to all dynodes, except the last, which is a plate close to the anode.

c) **Transit time balancing:** The crossover of SE trajectories between d1 and d2 (fig. 2.20) serves to balance the total transit time for varying hit positions on d1, which can not be avoided due to lateral pe momentum. Between d1 and d2 path 1 has a long transit time compared to path 5 due to low initial acceleration; however, between d2 and d3 the trajectories are reversed, which compensates the difference.
This compensation is incorporated between all dynodes and also occurs in the later stages of box-and line PMTs [325] (compare fig. 2.15).

d) **Focusing:** In addition, the dynodes also focus the electron cloud more and more from stage to stage, which reduces its cross-section and variations introduced from hit positions. This is actually the reason why the design is called "linear-focused". This focusing is further improved, if the dynodes are equipped with side cheeks (terminations on both sides, which are parallel to the paper plane in all figures shown) [170; pp. 233–234, 339].

5. **Anode readout [96]:** The collection of the electron cloud on the anode also introduces a small jitter. Typically, a reflex anode is used, for which the last dynode (dn) is a flat plate and the anode (a) is a fine mesh between dn and d(n−1), about 1 mm in front of the last dynode. Thus the electrons from the penultimate dynode must pass the anode on their way to dn, where they are amplified. The produced SEs then drift the short distance to the anode where they are collected. Transit-time balancing cannot be applied since there is no subsequent stage anymore. However, the high field gradients in this region due to the small distance dn–a, which can further be enhanced by use of a tapered voltage divider, together with the small path lengths lead to short drift times compared to interdynode transit times, below ≈1 ns. So the jitter resulting from drift time fluctuations is also smaller than for other dynodes, and often is neglected.
However, there is another effect contributing to the jitter originating from reflex anodes which can pose a problem. The electron cloud moving from d(n−1) to dn induces an anode current (Shockley-Ramo theorem), which flips polarity once the cloud passes the mesh; the resulting anode output current first is positive and then negative [170; p. 340]. The actual anode signal itself follows after 1–2 ns. Normally, this just causes an earlier, shallow onset of the anode pulse, but for very fast PMTs, the induced component can be visible at the very beginning of the anode pulse, which can cause early triggers – especially for large pulses containing many photons – and increase TTS. For

[96] Content based on [170; pp. 339–340, 438–439]

Series	Diam. [mm]	Window shape	Dynode type	No. of dyn.	HV [V]	Voltage divider type	g	T_0 [ns]	T_{FWHM} [ns]
R13478	25	Plano-concave	Linear focused	8	1500	Front-tapered	$5 \cdot 10^5$	9.1	0.13
R5990U-00-L16	30×30	Proximity-focused	Metal channel	10	800	Linear	$4 \cdot 10^6$	7.4	0.18
R9880U-210	30×30	Proximity-focused	Metal channel	10	1000	Linear	$2 \cdot 10^6$	2.7	0.2
R7761-70	39	Proximity-focused	Fine mesh	19	2000	Slightly front-tapered	$1 \cdot 10^7$	7.5	0.35
R12860	508	Hemi-spherical	Box-and-line	10	2000	Heavily front-tapered	$1 \cdot 10^7$	95	2.4
R877	127	Plano-planar	Box-and-grid	10	1250	Linear	$4 \cdot 10^5$	115	18.5

Table 2.1: Extreme examples for transit time and transit time spread for discrete dynode PMTs from Hamamatsu. Values from [179, 213].

this reason, the Photonis XP20D0 (very fast 2" PMT) incorporated a screening grid between d(n−1) and a, which reduced the induced current and decreased the TTS by 20%. This is also why extracting the signal from dn or d(n−1) instead of the anode can improve timing.

Typical values

For currently available large area PMTs with diameters of 5" or above, the TTS ranges from 1.5 ns (Hamamatsu R6595, 5" [213]) to 4.7 ns (ETE D784, 12" [306]); see also table 3.1.

In table 2.1 the TTS range of all – also smaller – currently available Hamamatsu PMTs is exemplarily summarized together with parameters relevant for timing, taking into account the different existing PMT subtypes.
The jitter ranges from merely 130 ps for the R13478 to 18.5 ns for the R877.
Larger diameters lead to higher TTS values, just as they increase the transit time (fig. 2.35).
Curved cathodes minimize the CTS; proximity-focusing is even better since the drift times to d1 generally are small, visible in form of low values for T – only 2.7 ns for the R9880U-210.
Naturally, the dynode type has great influence. Surprisingly, the dynode structure of the fastest PMT available from Hamamatsu now is linear focused and not fine mesh or metal channel anymore, which have the advantage of said short total transit times. This goes to show the potential of a well-optimized design. The slow box-and-grid layout results in enormous TTS values. Clearly, this design is not optimized for timing and is used for purposes where this plays no role, such as in spectroscopy.

A lower number of dynodes is beneficial for timing, since less jitter is introduced by the electron multiplier. However, this comes at the cost of a lower gain.
Larger high voltages decrease the transit time and jitter and can compensate this gain loss partially. See p. 507 for the high voltage (HV) dependence of the TTS which was observed in measurements performed in this work.
A tapered voltage divider also reduces TTS by applying higher voltages to the front stages and increasing the acceleration close to the cathode; tapering the rear end is more important for linearity than for timing. Metal channel and fine mesh PMTs not necessarily need tapered VDs, since the distances between stages are small and the field gradients thus are high anyway. The R7761-70 uses a voltage between k and d1 which is twice as high as between the rest of the stages to improve CE [326], the listed metal channel PMTs use completely linear voltage dividers. By contrast, the 20" R12860 uses 11.5 times the interdynode voltage of the last stages between cathode and anode [213] in order to minimize t_{k1} and the CTS, which account for the bulk of the TTS. This pays off, as T_{FWHM} is only 2.4 ns for a 20" cathode. Of course, this would be impossible without optimized electron optics and other improvements compared to the predecessor type R3600, which still used venetian blind dynodes and had a T_{FWHM} of 5.5 ns [228].
The R13478 achieves its short TTS by employing all secrets of the trade: a plano–concave cathode with small diameter (1"), linear focused dynodes with only 8 stages, a front-tapered voltage divider (VD), large HV of 1500 V resulting in high interdynode voltages and short total transit time, and an accelerator electrode at d8 potential placed between cathode and d1, which dramatically increases the extraction field and reduces t_{k1}. The result is a TTS of 130 ps; as a side-effect T amounts to just 9.1 ns. The cost of this timing optimization is a rather small gain of $5 \cdot 10^5$.
As opposed to this, the R877 lacks all of the above: it has a flat 5" window, slow box-and-grid dynodes with 10 stages, a linear VD, and small interdynode voltages. This leads to a transit time of 115 ns and a TTS of 18.5 ns.
As a side note, one of PMTs listed above, the R5990U-00-L16 even is position-sensitive in one dimension with 16 channels (suffix L16).

2.3.4 Calculation of transit time and transit time spread [97]

In the following, the electron movement in the electric fields of the photomultiplier shall be calculated analytically, from which the transit time and its jitter, the transit time spread can be derived.

Electron-optical simulations

An analytic treatment of the complex potential layout necessarily requires to make some approximations. For a fully correct calculation of the mean T and T_{FWHM}, a 3D electron-optical simulation would be needed, using a program as e.g. Maxwell or SIMION. For such a simulation, the electrode layout and the applied voltages are provided as inputs, from which the 3D potential distribution can be calculated by

[97] Content based on [170; pp. 322–341]

solving the Laplace equation. The electron trajectory, including landing position on d1 (or not) and the associated transit time, is then determined by ray tracing from points on the cathode to the hit position. By using a set of starting points distributed on the whole cathode one can in this way verify if designed TTS or CE requirements are met. However, the program only allows to check the performance of a layout; the design must be created and adapted by the user.
Apart from the costs for the license fees of a software, there is another problem: The electrode layouts of relevant PMTs are not public knowledge but trade secrets, and without accurate knowledge of the geometries no precise values can be obtained from simulations.

Analytical treatment

However, the voltages applied to the PMT stages *are* known – this can be easily calculated from any VD layout, presented in the brochures and datasheets of all manufacturers – and so are the rough properties of typical field shapes. From this the electron trajectories can be calculated analytically reliably enough to allow direct comparison with measured values. Relativistic effects can be ignored since the velocity of electrons and ions produced in the PMT is always $\ll c$. The following analysis is therefore based on 2D classical Newtonian laws of motion to simplify the treatment. According to Wright, crosschecks with actual 3D electron-optical simulations confirm the obtained principal results.

The electric potential will be assumed to vary only in the direction x of electron drift and shall be denoted as $U(x)$. The distance between electron emitting and receiving surface is termed x_1. Let the initial velocity of the electron be v_0 and its emission angle θ.
Equating the gained kinetic energy with the traversed potential difference yields

$$\frac{1}{2}m\left(v(x)^2 - v_0^2\right) = e\,U(x)$$

This follows from energy conservation alone and thus is valid *independently* of electron trajectory or field geometry. The velocity at x then reads

$$v(x) = \sqrt{v_0^2 + \frac{2e}{m}U(x)} \tag{2.132}$$

and again is independent of the path taken and the field geometry.
It is useful to introduce an effective potential U_0, which corresponds to the initial velocity, as auxiliary variable:

$$U_0 = \frac{\frac{1}{2}m\,v_0^2}{e} \tag{2.133}$$

This is connected to the initial energy E_0 by $E_0 = e\,U_0$.

Based on these initial considerations, the transit time can now be calculated for specific field distributions between two subsequent stages. The linear and quadratic potential best approach real field distributions, and therefore are analyzed closer in the following.

Linear potential

In the simplest case a uniform linear electric field can be assumed. This is the field shape encountered in proximity focused devices between cathode and multiplier region (MCP-PMT, LAPPD, ...; see ch. 4), which allows one-to-one mapping and thereby good position sensitivity [327].
In a linear field the electron experiences a constant force towards increasing potential U: dU/dx is constant. The potential thus can be described as

$$U(x) = U(x_1)\frac{x}{x_1} \tag{2.134}$$

The time t for an electron to move from $x = 0$ to x_1 can then be obtained using (2.132), (2.134) and (2.133): :

$$t = \int_0^{x_1} \frac{dx}{v} = \sqrt{\frac{2\,m}{e}}\frac{x_1}{U(x_1)}\left(\sqrt{U(x_1)+U_0} - \sqrt{U_0}\right) \tag{2.135}$$

This depends on U_0 and thus the initial velocity v_0. The drift time t_0 for $v_0 = 0$ follows to

$$t_0 = \sqrt{\frac{2\,m}{e}}\frac{x_1}{\sqrt{U(x_1)}} \tag{2.136}$$

Using pe energies of up to 3 eV and typical values of V_{k1} between 100 and 500 eV, (2.136) gives a deviation of a few percent compared to the full formula (2.135) and can be considered a good approximation. For a fixed distance the transit time t_0 varies with $1/\sqrt{U(x_1)}$.
The full formula can be simplified assuming $U_0 \ll U(x_1)$ for the first term in the bracket, which results in

$$t \approx \sqrt{\frac{2\,m}{e}}\frac{x_1}{\sqrt{U(x_1)}} - \sqrt{\frac{2\,m}{e}}\frac{\sqrt{U_0}\,x_1}{U(x_1)} \equiv t_0 - \delta t \tag{2.137}$$

where δt is the transit time correction resulting from the initial energy boost. For fixed distance and voltage the transit time t_0 is constant; however, δt depends on initial energy, which varies from pe to pe and introduces a jitter. A larger applied voltage $U(x_1)$ or a shorter distance x_1 linearly reduce the variation. The jitter δt can also be expressed as

$$\delta t = \sqrt{\frac{U_0}{U(x_1)}} \cdot t_0 \tag{2.138}$$

Using these formulas, the qualitative wavelength dependence of the transit time t and jitter δt for photoelectrons moving from k to d1 can be examined.
From fig. 2.4 can be deduced that for larger photon energies the mean initial energy increases and the emission energy distribution becomes broader. Eqs. (2.135) and (2.137) show that the transit time t decreases with initial energy, and thus with falling wavelength.
It is more difficult to assess the behavior of the jitter, since a shorter t decreases jitter, but the energy spread increases with wavelength. It is unclear which effect dominates without calculating a transit time distribution from the emission energy distribution and (2.135) and determining the jitter from the histogram. [170] notes that measurements show a decrease of the jitter with falling wavelength.

Quadratic potential

Another frequently encountered field geometry is a square law dependence on distance. For discrete dynode PMTs the field between cathode and d1 approximates a quadratic potential due to the focusing elements.
This potential can be described by

$$U(x) = U(x_1) \cdot \left(\frac{x}{x_1}\right)^2 \tag{2.139}$$

Using (2.132), (2.133) and (2.139) the velocity $v(x)$ results to

$$v(x) = \sqrt{\frac{2\,e}{m}\left(U(x_1)\frac{x^2}{x_1^2} + U_0\right)} \tag{2.140}$$

from which the transit time t to x_1 follows:

$$\begin{aligned} t &= \int_0^{x_1} \frac{dx}{v} = \sqrt{\frac{m}{2\,e}}\frac{x_1}{\sqrt{U(x_1)}}\left[\ln\left(x_1\left(1+\sqrt{1+\frac{U_0}{U(x_1)}}\right)\right) - \ln\left(x_1\sqrt{\frac{U_0}{U(x_1)}}\right)\right] \\ &= \sqrt{\frac{m}{2\,e}}\frac{x_1}{\sqrt{U(x_1)}}\ln\left(\frac{\sqrt{1+\frac{U_0}{U(x_1)}}+1}{\sqrt{\frac{U_0}{U(x_1)}}}\right) \end{aligned} \tag{2.141}$$

Since $\frac{U_0}{U(x_1)} \ll 1$, this simplifies to

$$t \approx \sqrt{\frac{m}{8\,e}}\frac{x_1}{\sqrt{U(x_1)}}\ln\left(\frac{4\,U(x_1)}{U_0}\right) \tag{2.142}$$

In these formulas, the jitter resulting from a varying v_0 is not a separate term, but can be estimated by varying U_0 from 0 to the maximum expected emission energy, e.g. 3 eV. Again t and thus the jitter decrease with higher voltage (the logarithm rises slower than the square root) and shorter distances.

At $U_0 = 0\,\text{eV}$ there is a pole in t, meaning that without initial energy photoelectrons can *not* be collected on d1. So for a square law potential at least a small initial kinetic energy is required for pe collection, whereas this is not the case for the linear potential.
This can also easily be seen from the electron acceleration, which is proportional to the electric field $E = -\frac{dU}{dx}$. For a square law (eq. (2.139))

$$E_q = -2U(x_1)\frac{x}{x_1^2}$$

which is 0 at $x = 0$. For a linear potential (eq. (2.134)), however,

$$E_l = \frac{U(x_1)}{x_1} = \text{const.}$$

The quadratic equations (2.141) and (2.142) yield larger values for t than the linear ones in (2.135) and (2.137) when the same values are used, especially close to $v_0 = 0$. Therefore, the wavelength dependence is expected to be larger for the quadratic potential.

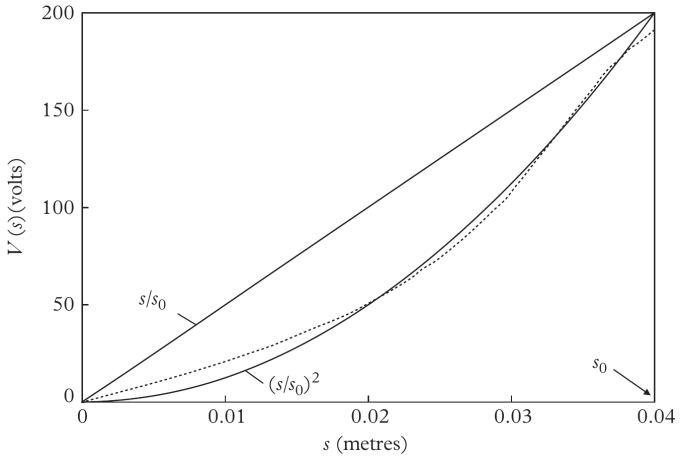


Figure 2.37: Potential distribution between cathode and d1 of an ETE 9814 fast, plano–concave 2" PMT ($T_{\mathrm{FWHM}} = 2.2\,\mathrm{ns}$ [328]) from field simulations (dotted line) compared to linear (upper solid line) and quadratic potential (lower solid line), fig. from [170; p. 325]. x, x_1 and U are termed s, s_0 and V in the figure, $x_1 = 40\,\mathrm{mm}$, $U(x_1) = 200\,\mathrm{V}$. The potential closely follows the quadratic field except near the cathode, where linear fields yield faster acceleration and thus shorter transit time and smaller jitter.

In a linear field, electrons undergo rapid acceleration after emission, whereas in a quadratic field the velocity remains relatively constant at the beginning. As the time required to traverse a distance scales with $1/v$ ($v = x/t$), in the square law potential the main contribution to both transit time and jitter stems from the early fraction of the electron path.

Application to photomultiplier fields

Field between cathode and first dynode

Real field distributions between cathode and d1 can either resemble quadratic (large vacuum photosensors, PMTs) or linear potentials (proximity-focused devices such as mesh PMTs, metal channel PMTs, MCP-PMTs, LAPPD, EBCMOS; see ch. 4).

For PMTs the field between k and d1 typically resembles a quadratic field, but tends towards linear distributions close to the cathode per design (fig. 2.37). The linear contribution is added in order to maximize the extraction potential and the acceleration for the important initial phase, since this reduces transit time and jitter. If an accelerator dynode is used, the field distribution initially is approximately quadratic in shape, but lies *above* a linear potential using V_{k1} during the complete trajectory, shows a *peak* at the accelerator due to its higher potential, and then sharply decreases again (not shown, cf. [170; p. 335]).
For large diameter PMTs, field shapes are hard to come by, but from their geometry and the equipotential line spacing of the ETE D784 [329] it can be concluded that they are closer to a quadratic field than a linear one in the medium and late phases of pe drift. The field shapes are likely similar to the ones for smaller PMTs, with a linear component near the cathode. The larger diameter and more bulbous shape, which would make the field more quadratic, are in part compensated by a larger

field-shaping aperture plate (deck). The ETE 9814, whose field distribution was depicted in fig. 2.37, also has a concave cathode and shows mixed field contributions.

Calculation for a Hamamatsu R5912

Although real potential distributions are likely closer to a quadratic field, the initial part – which has the greatest influence on t and its jitter – deviates from this shape. Thus both theoretical field configurations can only serve as approximations. This shall be analyzed for the 8" Hamamatsu R5912 now.

The distance between cathode and d1 was estimated from the datasheet [315; p. 35] and measuring of a sample PMT to lie between 10.8 and 12.8 cm.
The transit time t_{k1} from cathode to d1 can be estimated from TTD measurements (fig. 2.36) from the effects $\gamma 1$ and L1 and the main peak.
$\gamma 1$ events occur when a photon passes the cathode and creates a photoelectron on d1. The produced pulse thus has a transit time $T_{\gamma 1}$ which is shorter (earlier) than that the main peak (T_0). The difference between T_0 and $T_{\gamma 1}$ originates mostly from skipping the electron transit time t_{k1} between k and d1. This is replaced by the photon flight time between cathode and d1

$$t_{k1,\gamma} = \frac{x_{k1}}{c}$$

which amounts to 0.36–0.43 ns for 10.8–12.8 cm. In addition, the different transit times to d2 of a photoelectron and a secondary electron produced on d1 have to be considered, as will be detailed in 2.3.5 (0.94 ns longer for the pe – the smaller U_0 has a large influence). For $V_t = 1425\,\mathrm{V}$ ($g = 1.3 \cdot 10^7$) $T_{\gamma 1} - T_0$ was measured to -23.75 ns, which results in $t_{k1} \approx 25.1$ ns. $\gamma 1$, however, can only be observed for photons incident on d1, which is only for a small area close to the symmetry axis.
L1 pulses originate from elastic backscattering of a pe on d1, which after flying towards the cathode is refocused by the fields onto d1 where it produces a pulse. For L1, the maximum difference in transit time to T_0 should approximately amount to $2\,t_{k1}$, since elastic scattering losses are small to negligible (p. 220). However, the same statistic and systematic effects which broaden the main peak on its trailing flank also apply to the L1 peak, which makes it challenging to define a maximum lateness. Instead, the $T_{L1} - T_0$ of the L1 peak of 48.81 ns was used, which yields a t_{k1} of 24.4 ns. This value is in good agreement with the one obtained from the $\gamma 1$ peak but slightly smaller, for reasons which can be better explained after the introduction of time shift theory (see p. 298, 316).
All in all, the $\gamma 1$ value seems more reliable, despite of arising from only a fraction of the cathode area, because there are no additional effects influencing it.

The total jitter T_{FWHM} was measured to 2.26 ns. Since this is in large part also due to the CTS, direct comparison to the calculated jitter from pe initial energy alone is not possible. The value can serve as a guideline, however.
The interstage voltage V_{k1} used in the measurements can be reconstructed from the VD layout [315; p. 47] to $514\,V$.

For the linear potential, inserting the distance range and voltage between k and d1 in (2.135) yields t_0 =16.06–19.03 ns for $U_0 = 0$. Using the FWHM values of the initial pe energy distribution for a 405 nm photon in fig. 2.4 (≈0.63–1.45 eV) leads to $t\,(0.63\,\mathrm{eV})$=15.51–18.38 ns and $t\,(1.45\,\mathrm{eV})$=15.23–18.05 ns.

The jitter can be calculated using (2.138) and (2.136). For 1.8 eV pe energy (maximum initial energy) δt vs. 0 eV amounts to 0.95–1.13 ns; in both cases no less than 5.9% of t_0. However, this is the maximum difference in pe energy. If the FWHM is taken instead, the variation in δt shrinks to 0.29–0.34 ns. The maximum difference between the exact formula (2.135) and the approximation (2.137) for $U_0 = 1.8$ eV and $x_1 = 12.8$ cm amounts to only 0.03 ns, which justifies the assumptions made in deriving (2.137).

For the quadratic potential, no value can be calculated for 0 eV. The transit times for the FWHM values are $t\,(0.63\,\text{eV}) = 32.49$–$38.50$ ns and $t\,(1.45\,\text{eV}) = 29.14$–$34.54$ ns; the difference between exact (2.141) and approximated formula (2.142) equates to less than 0.01 ns for 1.8 eV and 12.8 cm. The jitter can be estimated from the difference between FWHM values to 3.34–3.96 ns.

As expected, transit time and jitter obtained with the linear potential are too low, whereas the quadratic field results in too high values. The real field distributions lie somewhere in between both formulas.

In want of 3D electron optics simulations, therefore, in the following the transit time between cathode and d1 is approximated by the averaged value from quadratic and linear potentials. For a mean pe energy of ≈1.1 eV and x_1 =11.8 cm this leads to $t = 24.90$ ns, which is very close to the $\gamma 1$ value of 25.11 ns. The remaining difference is scaled by a correction factor in x_{k1}. Since for both fields the transit time is proportional to the distance, the correction factor can be directly multiplied with the result. If x_{k1} is scaled to match the measured t_{k1}, a value of 11.9 cm results. These simple estimations cannot substitute a simulation and are unfit to measure internal PMT dimensions. In order to estimate the transit time of timing effects and compare their time shift, however, it is necessary to have a formula and adapt it as best one can to the properties of the PMT at hand, and for this reason a scaling factor is acceptable.

Cathode time spread

The cathode time spread (CTS) is the transit time difference resulting from different path lengths between k and d1. Since both (2.135) and (2.141) are linear in x_1, the magnitude of the CTS can be assessed by substituting x_1 by a small additional path length δx in both formulas. In other words, the CTS is proportional to the transit time t_{k1} times $\delta x / x_1$.
Therefore, the path lengths should be as uniform as possible, which can be achieved by using a spherical inner cathode surface centered on d1. Illuminating a reduced diameter of the cathode can also help to improve performance, if this is acceptable.

Spot size on first dynode

Photoelectrons also are subject to lateral drift for finite initial energies and emission angles different from the surface normal. The maximum lateral offset y_{max} accumulated after the transit time t_{k1} from cathode to d1 can be derived using the above formulas by assuming the pe is emitted with 90° to the surface normal. Then

$$y_{\text{max}} = v_0 \cdot t_{\text{k1}}. \tag{2.143}$$

For the linear and quadratic potential follows

$$y_{\max,\,\mathrm{lin}} = 2\sqrt{\frac{U_0}{U(x_1)}}x_1 \tag{2.144}$$

$$y_{\max,\,\mathrm{quad}} = \frac{1}{2}\sqrt{\frac{U_0}{U(x_1)}}\,\ln\left(\frac{4\,U(x_1)}{U_0}\right)\,x_1. \tag{2.145}$$

The offset for photoelectrons emitted at smaller angles can be estimated[98] by

$$y \approx y_{\max}\cdot\sin\theta \tag{2.146}$$

For a linear potential the paths are parabolic.

The lateral drift changes the starting position of the SE cascade on d1 which constitutes another source of jitter. From (2.144) and (2.145) follows that reducing x_1, increasing the voltage and a larger wavelength reduce the spot size on d1 and thus decrease jitter. Using (2.144) Wright calculated a FWHM of the spot size already in the order of 2–5 mm for a 2" PMT [170]. For large PMTs x_1 is determined by cathode diameter, and since $y_{\max}$ scales linearly with x_1, it becomes necessary to either increase the size of the first dynode and/or raise V_{k1}, to make certain that the hit area does not extend beyond d1. For the R5912 the maximum *radius* of the spot size for a maximum pe energy of 1.8 eV is 1.41 cm resp. 2.48 cm from the linear and quadratic formulas – the initial acceleration is smaller for the quadratic field so the lateral drift is far larger. Even taking the mean of both values (1.94 cm) this is larger than the half diameter of (the very large) d1, which accentuates the importance of high extraction fields and good focusing. It also shows why CE values for large PMTs typically are far worse than for small ones – according to Wright all large PMTS suffer in collection efficiency from d1 hit area considerations.

Interdynode transit time

The interdynode transit times and jitters can be calculated from the equations for the linear and quadratic potential, using the smaller interdynode distances and voltages. Looking at fig. 2.20, the equipotential lines are very inhomogeneous. The best approximation with the presented analytic formulas (see 2.3.5.5) was found by taking the mean of the linear and quadratic fields.
For a linear focused PMT the potential difference to the dynode *after next* should be used since the extraction field is mainly determined by dn+2 (fig. 2.20). For a box and line PMT the extraction fields are expected to be affected more by dn+1 (fig. 2.15c), which is why in this work the potential difference to the next dynode is used.
With these approximations transit times can be obtained, which lie in the same order as values from electron-optics simulations.

For the R5912 the interdynode distances x_{ij} were estimated roughly from the dimensions of the PMT outline in the datasheet [315; p. 35] and the generic layout of box-and-line dynodes in fig. 2.15c to

$$\{x_{12},\,x_{23},\,\ldots\,x_{910},\,x_{10\mathrm{a}}\} = \{1.7,\,1.15,\,1.3,\,1,\,1,\,1,\,1,\,1,\,1,\,0.2\}\,\mathrm{cm} \tag{2.147}$$

[98] To be exact, the shorter t_{k1} from $v_0 > 0$ has to be taken into account by using the complete formulas.

The transit times were calculated using the linear potential, a mean value of $U_0 = 5\,\text{eV}$ [170; p. 323] and the potentials of the next stage known from VD layout. The sum of the resulting values is 35.6 ns, which is close to the measured value $T_0 - t_{\text{k1},\gamma 1} = 59.89 - 25.11\,\text{ns} = 34.78\,\text{ns}$ – more cannot be expected from the approximate analytic formulas. To improve the values for calculation of the transit time of TSEs, the dynode distances were adapted with a correction factor (0.98) to match the total transit time, which results in the corrected interdynode transit times

$$\{t_{12},\, t_{23},\, \ldots\, t_{910},\, t_{10\text{a}}\} = \{4.3,\, 2.7,\, 3.5,\, 3.4,\, 4.0,\, 4.0,\, 4.0,\, 4.0,\, 4.0,\, 0.8\}\,\text{ns} \quad (2.148)$$

An even easier and less exact alternative is to estimate the interdynode transit time t_{dd} from the total transit time T_0 for n dynodes through

$$t_{dd} \approx \frac{T_0}{n+1} \quad (2.149)$$

where the total transit time T_0 has to be divided by $n+1$ since the anode has to be taken into account as well.

Anode readout

With the equations for the linear potential, the transit time and jitter for the anode readout can be calculated as well. [170] arrived at a delay of the anode signal to the induced one of 0.5 ns and obtained a jitter of 0.1 ns from anode readout. Since the dimensions and spacings of the later stages do not scale much with PMT diameter, these values give the expected order of magnitude for all PMTs. For the R5912 a value of 0.8 ns was calculated.

Transit time spread [99]

Since the statistic processes in all stages are independent, the overall variance of the transit time can be calculated as the quadratic sum of the weighted contributions of all stages using cascade theory.
Between the cathode and the first dynode only one photoelectron travels (spe pulse), which then initiates the cascade on d1, so the complete transit time variation of this stage is passed to succeeding stages and the variance v_{k1} fully contributes, i.e. has a weight of 1. v_{k1} comprises CTS as well as jitter due to v_0 and emission angle.
The transit time between d1 and d2 is the mean of the individually varying drift times of the $\delta 1$ secondary electrons produced on d1. This reduces the variance v_{12} by the weight $1/\delta 1$. Similarly, the weight of the variance v_{23} is $1/(\delta 1 \cdot \delta 2)$, and between the dynodes n-1 and n it is $1/(\delta 1 \cdot \delta 2 \cdots \delta \text{n})$. During the multiplication process in the dynode system, due to the large number of secondary electrons a time-dispersion mostly results in pulse broadening rather than an increase of the transit time spread [301]. Thus, while usually correlated, a small pulse width not always leads to a small TTS [170; p. 373]. In addition to the reduced weight of the dynode jitter, transit time compensation is active (see p. 171).

[99] Content based on [267; pp. 59–60] [170; pp. 370–373]

The variance of anode readout v_{na} enters with a weight of 1 – there are systematic effects at work here, such as the Shockley-Ramo pulse and the missing transit time compensation.
Putting everything together, this results in the variance v_{TTS} of the transit time

$$v_{\mathrm{TTS}} = v_{\mathrm{k1}} + \frac{v_{12}}{\delta 1} + \frac{v_{23}}{\delta 1 \cdot \delta 2} + \cdots + \frac{v_{\mathrm{n-1,n}}}{\delta 1 \cdots \delta(\mathrm{n}-1)} + v_{\mathrm{na}} \tag{2.150}$$

from which σ_T can be obtained through $\sigma_T = \sqrt{v_{\mathrm{TTS}}}$.
Since v_{k1} fully enters the jitter, and the following terms are reduced by $\delta 1$, V_{k1} should be higher than the other stages not only to reduce the charge resolution, but also to achieve a low TTS.
The equation also shows that, despite the growing denominator, incorporating fewer stages reduces the transit time spread.

If all dynodes except the first are assumed to have the same δ and variance v_{dd}, the dynode terms after v_{12} can be summarized and (2.150) then becomes [170; pp. 370–372]

$$v_{\mathrm{TTS}} \approx v_{\mathrm{k1}} + \frac{v_{12}}{\delta 1} + \frac{v_{dd}}{\delta 1(\delta - 1)} + v_{\mathrm{na}} \tag{2.151}$$

If the fluctuations of the number of produced SE are also taken into account, this *increases* the weighting of the dynodes with approximately the excess noise factor (ENF). However, since F is typically close to 1, the above formulas are good approximations.
Both equations require values for the variances, which have to be obtained from electron-optics simulations, or estimated using the formulas derived above for linear or quadratic fields. While the analytic formulas can only serve as order of magnitude estimates, they are helpful to show the dependence of v_{TTS} on early interstage voltages (mainly V_{k1} and V_{12}) and $\delta 1$.

The TTS shall be calculated exemplarily for the R5912 with (2.150).
Using the average of the FWHM values of t_{k1} for the quadratic field and δt for the linear field yields an FWHM jitter between cathode and d1 of $t_{\mathrm{k1,FWHM}} = 2.00\,\mathrm{ns}$, from which σ_{k1} can be calculated with (2.68), assuming a Gaussian distribution, and $v_{k1} = \sigma_{k1}^2$.
The CTS is unknown, but can be estimated from the resulting total TTS – which will be too small – by increasing the value of v_{k1} until both figures match.
The FWHM jitter of each stage can be assessed by calculating t_0 and δt assuming an SE emission distribution with FWHM of 5 eV centered on the mean energy of 5 eV. The standard deviation and variance can be obtained in the same manner as for k–d1. If the weights are taken into account according to the δn values of (2.41), relevant contributions only come from d1–d2 ($v_{12}/\delta 1 = 5.8 \cdot 10^{-3}\,\mathrm{ns}^2$) and d10–a ($v_{10a} = 4.6 \cdot 10^{-3}\,\mathrm{ns}^2$), which are of similar size. Compared to $v_{\mathrm{k1}} = 0.723\,\mathrm{ns}^2$ these terms constitute minor corrections at best.
The total variance v_{TTS} then amounts to $0.733\,\mathrm{ns}^2$, from which $\sigma_T = 0.86\,\mathrm{ns}$ and $T_{\mathrm{FWHM}} = 2.02\,\mathrm{ns}$ follow. As expected, this is smaller than the measured value of 2.26 ns, and can be used to estimate the CTS: A value of $T_{\mathrm{FWHM,CTS}} = 1.02\,\mathrm{ns}$ reproduces the overall TTS.

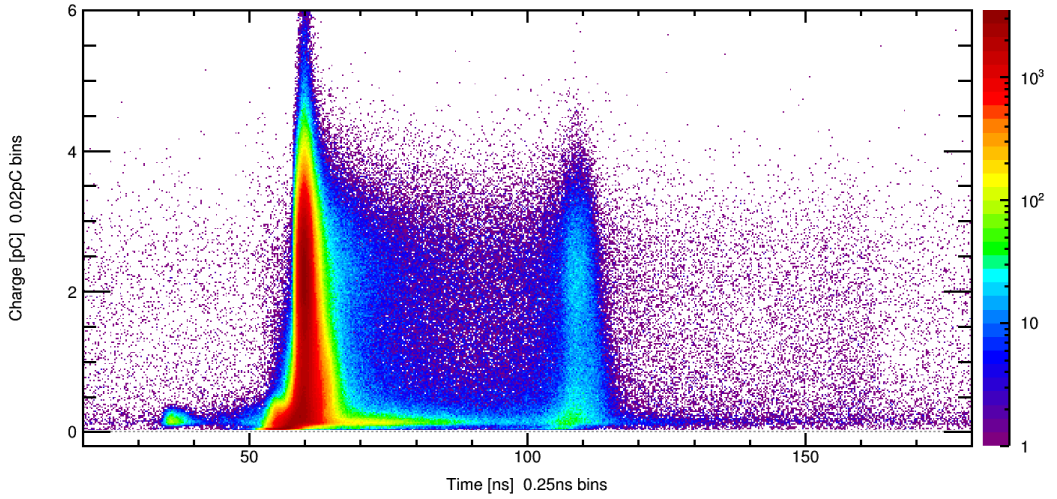


Figure 2.38: Charge–transit time distribution (CTD) of primary pulses (i.e. the charge dependence of the TTD) for a Hamamatsu R5912 at $g = 1.3 \cdot 10^7$, $5 \cdot 10^6$ waveforms. The use of a shielded setup and pulse fitting enabled the exceptionally low threshold of only 1.4% pe. The regular pulses are centered around 2.4 pC at 60 ns. As can be seen, a plethora of other effects populate the distribution.

2.3.5 Time shift theory

When analyzing the primary pulse charge–time distribution (CTD, fig. 2.38), a multitude of effects causing deviations from the regular charge and transit time are observed, especially if measures are taken to lower the detection threshold, such as using a shielded setup or a pulse-search algorithm based on fits [100]. This is even more the case for the early afterpulse charge–delay distribution (CDD; depicts the charge of APs vs. their delay after the primary pulse), described in 2.5.2.2, which in this work for the first time could be analyzed for delays down to 0.8 ns and charges as low as 1.4% pe. As will be seen, early afterpulses are just another side of the same underlying effects which cause the time shifts for the primary pulses – both cannot be understood separately.

Since the vast majority of these effects has not been described in literature so far, it became necessary to develop a theory, which can explain their origins and predict the observed time, charge, height and probability of all primary pulse and afterpulse effects.
In the following section a unified theory of transit time shifting effects (time shift effects, TSEs) in photosensors will be introduced: time shift theory (TST).

The target is to classify underlying effects by physical origin, i.e. particle interactions, alone and not phenomenologically. Present terms for time shift effects such as prepulse, early pulse, late pulse, afterpulse and double pulse are mostly still phenomenological (time shift, number of pulses) albeit the underlying particle interactions are well understood (with some alternative explanation models for some

[100] With pulse-fitting the threshold can be lowered quite substantially below the instrumental threshold – however, *only* if the event contains at least one pulse *above* the instrumental threshold which triggers acquisition of the waveform. This limits the observable small charge effects to AP-generating types.

effects). The term split pulse is an exception; however, it will be seen that there are a few dozens of effects, which could be described by this expression, so it lacks clarity.
This goal requires a thorough theoretical understanding of the processes taking place in all stages of the photosensor and is the main reason for the extensive treatment of photosensor properties in the preceding sections.
Effect properties will be calculated analytically to allow to understand dependences and observations through order of magnitude estimates without having to carry out extensive 3D electron-optics simulations, which is impossible without proprietary knowledge. These calculations are performed mostly ab initio to prove that the effects can be well understood without artificial fine-tuning. Since many values for variables could not be calculated or extracted from literature, the necessary assumptions were often inaccurate. Thus a select few input variables were taken from data to improve the match. By implementing additional values from data or simulations, the predictions could be further improved, albeit the coincidence with observations already is remarkable.
TST is but a rigorous application and combination of all the principles compiled in the previous sections with extensions added where needed. It can, however, explain *all* properties of *all* observed time shift effects at least with the right magnitude.

The consequent combination of effect subtypes leads to the postulation of many new effects. Many of these can be verified promptly from data; others have too small charges or probabilities to be observable. Often it is impossible to predict the impact of an effect without calculating it first, which is why it was necessary to include a large number of effects in the study (see 2.3.6 and 2.5.2.2).

Time shift theory was developed to better understand the workings of PMTs, but can be applied for all vacuum-based photosensors with photocathodes (see 4.3.1, e.g. MCP-PMT, LAPPD, Hamamatsu HPDs [101], ...), since the underlying processes are the same, and the notations introduced in 2.3.5.1 can be generalized to encompass all photodetectors by adding the additional relevant effect types. For example for semiconductor-based sensors such as SiPMs, many of the effects described here are expected to occur as well: photonic feedback resp. crosstalk, pulse lateness through elastic backward scattering of the primary charge carrier, photoeffect occurring on non-standard stages, etc. The missing interactions, like afterpulsing through carrier trapping, or late pulses from diffusion of carriers generated by long wavelengths in the neutral regions beneath the depleted layer [324], can easily be completed.

In the following, first the general effect notations for the possible interactions, as well as the abbreviations used to describe effect properties shall be introduced in 2.3.5.1. Afterwards observable effects are identified from the multitude of possible combinations of particle, interaction and site in 2.3.5.2, which allows to postulate previously unknown effects. All relevant effects are depicted schematically. Following this, the analytic calculation of the probabilities (in 2.3.5.3), charge (2.3.5.4) and time (2.3.5.5) is described together with the estimates and assumptions made. The dependence of TSEs on direction is briefly addressed in 2.3.5.6, before the section closes with a general discussion of the theory in 2.3.5.7.

[101] Some extensions would be required to account for the internal photoeffect used by the avalanche photodiode (APD).

Following this section, the individual TSEs are discussed and their properties are derived. For the resulting primary pulses this is done in 2.3.6, while the resultant early afterpulses are covered in 2.5.2.2. In table 2.2 and 2.3 all relevant TSEs are summarized along with their notations and predicted properties. In 2.5.2.3 TST is also applied to ionic afterpulses, to the extent currently possible.

Since the estimation of parameters and execution of calculations for about 200 time shift effects are rather lengthy and time-consuming and this has to be adapted for each sensor, the calculations presented here were only carried out exemplarily for a well-understood PMT, the Hamamatsu R5912 (8“), for which measurements had been performed with several setups with a front-tapered VD (Hamamatsu E7694-01) at $+1425\,\mathrm{V}$ and $g = 1.3 \cdot 10^7$. The presented numerical values are for this device. While emphasis was put on generality, in places the methodology might therefore be tailored to this PMT.

2.3.5.1 Classification and abbreviations

The notation acts as framework for all calculations. It allows to concisely and unambiguously describe an effect (185 were studied in the context of TST, many others postulated) as well as the construction of higher-order effects consisting of multiple interactions. In a way, this is not unlike perturbation theory in quantum field theory; here the topology form fulfills the function of the Feynman graphs.
This classification is also intended as a clarification of nomenclature, since some terms are used ambiguously in the literature (prepulses, late pulses and afterpulses).

First the notation forms which are used subsequently are explained. These are based on the three-part description particle–interaction–site. The occurring particle types and their interactions with structures in the device are postulated and listed afterwards, based on the processes occurring in PMTs, which were elaborated in the previous sections. The terms for the afterpulse effects are already given in this section in order to have all definitions in one place, the effects are discussed in the AP section 2.5.2, however. Additional abbreviations are introduced as needed in the sections on probability 2.3.5.3, charge 2.3.5.4 and time 2.3.5.5.

In general, the notation is kept as short as possible by defining standard variables, which can be omitted wherever unambiguity allows, in order to make the descriptors clearer, more readable, and better memorizable.

Notations

Three notations will be used, a short form describing the effect and serving as a label, an extended form (topology form) showing the interactions occurring on each stage of the device for visualization and discussion, and a mixed form containing the topological information in a condensed notation. This allows to describe even complex interactions accurately and concisely.
A fourth notation – the impact notation – is used not to describe the event respective pulse, but the influence of a particle's history (kinetic energy, trajectory and transit time) on a single variable, such as the gain this particle will produce on impact. The impact notation is defined later, on p. 218, since the required terms have to be introduced first.

Compact form

- The compact form consists of the concatenated abbreviations for the particle causing the effect, the effect type, and the location in this order [102].
- For clarity, the three notation sections can be separated by colons: (particle:effect:location)
- When discussing multiple effects different variants of particle, interactions or sites can be separated by slashes.
- For each location a standard particle can be defined, which usually causes the interactions there. On d1 this is a photoelectron from the cathode (pk or p), on a later dynode n it is a secondary electron from the previous dynode: s(n − 1). To abbreviate the notation, the standard particle of a location can be omitted.
- Effects can be stacked by concatenating their labels in chronological order. For better readability, a hyphen can be added between the labels of subsequent effects. If the following effect is caused by the same particle, it can be omitted in its notation. This takes precedence over the standard particle wherever possible.
- Regular electron creation processes are omitted in the compact form for brevity. This includes the photoeffect (P) of a photon (γ) on the cathode (γ:P:k) and regular SE generation (R) on a dynode n (x:R:n where x stands for any electron abbreviation (see below for a list), e.g. pk:R:1 and sn:R:n+1). How these processes look in the PMT is visible in fig. 2.39a.
- Examples:
  - In a classic late pulse (fig. 2.40d), a photoelectron from the cathode (p) is elastically scattered backwards (L) towards the previous stage – in this case the cathode – on d1 (1), which results in the compact notation pL1 or p:L:1. The p(k) is the standard particle on d1 and can be omitted, which shortens the notation to L1.
  - If, after returning to d1, this pk is elastically scattered a second time, this can be written as pL1pL1 or pL1-pL1, which can be shortened to L1L1, since the pk is also the particle of the second interaction.
  - When a pk is elastically scattered forward (E) towards the next stage d2 on d1, and the pk is another time elastically scattered forward to d3 on d2, the effect is termed pE1-pE2 (see fig. 2.41a). The label can be shortened to E1E2, since the pk is also the particle interacting on d2. The precedence rule ensures that on d2 a pk and not an SE causes the interaction. The latter would be dubbed pE1-s1E2 (short E1s1E2) and would be impossible, since after pE1 no SEs arrive on d2.

[102] The order is based on English grammar to allow to construct readable sentences of the type subject–verb–place (who–what–where). For example, pL1 can be read as: A photoelectron (p) is elastically scattered backwards (L) on d1 (1).

Event topology form

- The extended form illustrates the interactions taking place in the PMT and visualizes their chronological order and spatial distribution.
- The interactions are listed dynode for dynode. The time axis flows from left to right, and stages are separated by arrows (→).
- Each interaction is written in the compact form, i.e. particle:effect:location, where the colons can be omitted if clarity allows. Effect stacking does not occur, since subsequent effects are described separately with their respective interactions.
- To save space, the notation is stopped once regular amplification (R) on all subsequent stages occurs, or on the step before. Initial steps with regular amplification can also be omitted.
- If the pulse branches, additional pulses are continued in separate lines below the original pulse in the same notation and connected by arrows to the interaction from which they emerge.
- Examples:
  - For L1 first a photon (γ) creates a pe through the photoeffect (P) on the cathode (k), which then is scattered towards the cathode before hitting d1 again and starting a regular cascade (R): γPk → pL1 → pR1 (compare to fig. 2.40d).
  - An L1L1 pulse can be written as γPk → pL1 → pL1 → pR1.
  - If a pk regularly creates secondary electrons on d1, but one of these SEs (s1) is elastically backscattered towards d1 (L) on d2 (2), while the rest of the SE from d1 (r1) is regularly amplified (R) on d2 (2), this results in an s1L2 (short L2):

    $$\gamma\text{Pk} \rightarrow \text{pR1} \begin{array}{l} \nearrow \text{r1R2} \\ \searrow \text{s1L2} \rightarrow \text{s1R2} \end{array}$$

    In fig. 2.42d L2 is depicted schematically; fig. 2.52g shows a typical waveform.

Alternative topological notations

Other types of depicting the interaction in the PMT are conceivable.
For example, instead of showing only the temporal axis in the topological notation presented here, one could plot the interactions in time and space.
In the simplest case this can be done by arranging the dynodes in ascending order on the ordinate and placing the occurring interactions according to the overall transit time on the abscissa. The advantage of this form of representation is that the earliness and lateness introduced by effects as well as their spatial layout become visible at a glance. The drawback is of course that such a display takes up far more space (and this thesis is already long enough as it is).
In a more sophisticated version, one could actually plot the 1D spatial progression $x(t)$ of the average particle of a cascade using the formulas given in 2.3.5.5, where the different stages are only visible as kinks in the paths but can be labeled. In this representation the effects become visible even more clearly.

Mixed form

- The mixed form contains the same information as the extended form in a condensed way.
- Interactions are separated by hyphens, as for the compact form.
- Like for the extended form, all interactions are listed from a suitable starting point up to regular amplification for all following stages.
- Pulse branching, however, is not continued on the next line. Instead, before the stage, where the branching occurs, an equal sign is used in place of a hyphen, followed by the interactions of the first branch. The second branch is described on the same line, starting after a plus sign.
- In the example from above, L2 would be written as: γPk-pR1=r1R2+s1L2-s1R2

Pulse order

- **PP** Primary pulse: the first output pulse
- **AP** Afterpulse: all pulses after the primary pulse

If APs are observed, usually the primary pulse is a regular pulse, possibly with reduced charge, but with regular transit time. Some effects, however, can create a secondary pulse which arrives at the anode earlier than the otherwise regular original pulse and thus becomes the primary pulse. The regular pulse then is an afterpulse.

Particle

There are not many particles types in a PMT which can cause any time shift effects; this must be either the incident photon or one (or more) of the particles produced in the PMT, i.e. a photoelectron, a secondary electron (of whatever sort), a quasiparticle (e.g. through energy loss), an ion, or a photon caused by luminescence (including X-rays). To make the notation more general, radioactivity from passing muons and γ, β or α radiation is also included since it can generate dark counts.

- **Electrons**

  p(n) Photoelectron from structure n. Usually emitted from k, in which case n can be left out. In addition, pk is the standard particle for d1, g, f, m, s and c (see p. 196) and can be omitted there.

  sn Secondary electron from dynode n. Standard particle for d(n+1), can be omitted there.

  shn High-energy secondary electron from dn. An SE with markedly higher emission energy (>25 eV, see p. 224), which reduces the transit time to the next stage and separates the pulse generated by the shn in time from the main cascade.

san Auger electron from dn. These are like SE with higher energies and are emitted very close to the surface. The energy spent in the creation of an san is missing in the production of regular SE, which can reduce the number of total SEs and act as a loss process. Due to the higher energy, Auger electrons are not focused as effectively onto the next dynode and have a higher probability to be lost. Also, since the curve of the secondary emission yield (SEY) levels out, they are amplified a little less effectively at the next stage.

sxn An sn, shn or san

rn The rest of secondary electrons produced on dn

an All secondary electrons produced on dn. Example s2L3 (short L3): This is a regular pulse, where a single SE from d2 is elastically scattered backwards on d3, on d2 all SE from d1 (a1) are still amplified normally; on d3 the rest of the s2 (r2) produce a regular pulse:

$$\gamma\text{Pk} \rightarrow \text{pR1} \rightarrow \text{a1R2} \begin{array}{l} \nearrow \text{r2R3} \\ \searrow \text{s2L3} \rightarrow \text{s2R3} \end{array}$$

j Indicates that separate cascades join at this stage. Example s2BE3 (BE3): For a regular pulse, where an SE from d2 on d3 is inelastically scattered forward towards d4 after creating SEs on d3 which branch off a new pulse (BE), on d4 the regular pulse and the SEs produced by the scattered s2 merge and continue as one pulse:

$$\gamma\text{:P:k} \rightarrow \text{p:R:1} \rightarrow \text{a1:R:2} \begin{array}{ll} \nearrow & \text{r2:R:3} \rightarrow \text{a3j:R:4} \\ \searrow & \text{s2:BE:3} \nearrow \rightarrow \text{s2:R:4} \end{array}$$

One could argue that both cascades at no point in time are separated, but this way the extra electrons switching from the s2 cascade to the regular cascade better become visible. This helps to distinguish a BE3 from a ΛE3, which is also due to inelastic scattering on d3 but caused by loss processes without SE production (Λ):

$$\gamma\text{Pk} \rightarrow \text{pR1} \rightarrow \text{a1R2} \begin{array}{l} \nearrow \text{r2R3 } (\rightarrow \text{a3R4}) \\ \searrow \text{s2ΛE3} \rightarrow \text{s2R4} \end{array}$$

sin Electron from the ionization of a residual gas atom/molecule in the vacuum between stage n and the following stage

- **Ions**

  i(xy) Ion produced in the vacuum between stage x and y, where y can be omitted and the specification of the origin is optional. i can be replaced by the specific ion type, e.g. H_2^+. Generic negative ion denoted by i^-.

- **Photon**

  γ Photon

  γ_{ij} Photon from structure i hitting j

γ_c Cathodoluminescence

γ_b Bremsstrahlung

γ_l Photoluminescence

γ_f Fluorescence

γ_p Phosphorescence

γ_d Electric discharge

- **Radioactive radiation**

μ Muon

γ_g Gamma radiation

β Beta radiation

α Alpha radiation

Effect types

- **Electron interactions**
  The electron-dynode interactions were already discussed in 2.1.3 and starting on p. 106 in detail and are summarized and labeled here. Fig. 2.12 is particularly helpful as it shows the actually occurring interactions, which helps to substantiate the effects discussed here.

  R Regular amplification. The electron loses its complete energy through various processes while creating a regular amount of SEs (δ corresponding [103] to fig. 2.10). All emitted electrons are true secondaries with typical emission energies ($\mathcal{O}(5\,\mathrm{eV})$), confer fig. 2.12.
  Since this is the standard interaction, the abbreviation normally is omitted in the compact form. At the anode R denotes regular collection.

  Λ Loss processes: A reduced number of SEs is produced. This can be due to various reasons:

  Λ_s Statistical losses. In extreme cases this can lead to zero SE emission from a dynode, which adds to the multiplication efficiency (ME). Strictly speaking, this is a subset of R; Λ_s can be viewed as separate effect due to the phenomenological difference in form of reduced charge. There is a fluid transition to R and also Λ_p (see below), since the losses are due to particle interactions in the end (e.g. with phonons).

  Λ_i Inelastic backscattering losses (term identical to BT): After using a fraction of its energy to create SEs, the primary electron is backscattered out of the dynode and lost (lands on an inactive part). Therefore,

[103] To be exact, for regular amplification δ likely will be *higher* than the mean measured stage gain, since the average value is reduced by various following loss processes.

not all of its kinetic energy is converted into SEs. These events are part of the inelastic plateau in fig. 2.12.

Λ_p Particle losses: In this case a substantial fraction of energy is lost in a) the generation of bremsstrahlung (Λ_b), characteristic X-rays (Λ_x), or Auger electrons (Λ_a), which do *not* subsequently start another cascade (if they do this is treated as separate effects), or b) by scattering on a (quasi-)particle such as a plasmon (Λ_{pl}). From the remaining energy a lower amount of SEs is generated, leading to an underamplified pulse. This is visible in the SE energy distribution in the inelastic region as peaks (Auger electrons as additional peak, plasmons as escape peak) or continuously (bremsstrahlung losses span the whole spectrum).
Characteristic X-rays do *not* appear as loss peaks, since they are emitted due to relaxation processes after a bound electron was knocked out from an inner shell. The energy carried away by the X-ray is lost for SE generation but does not affect the energy of any SE directly. The ionization, which leads to X-ray production but should cause plateaus in the SE energy spectrum, which start at the respective binding energies and are smeared and diminished by scattering of the knocked out electrons on their way to the surface. Likely, this plays no large role.

ζ Collection losses: The electron fails to hit the next stage. This is connected to the CE from k to d1, but can be defined for any two successive stages. For d2+ it contributes to the ME, if all SEs fail to propagate.

Φ Ionization of a residual gas molecule/atom. Through the scattering the electron loses a fraction of its energy and changes its direction, and one or more sin-electrons are produced.

γ Photon production: Through any of the interactions listed in 2.6.1. Photons can be produced as a side-effect of a cascade without reducing its energy and charge. This is denoted by combining both symbols to Rγ.

Backscattering

Type

ε Elastic backscattering: The electron is elastically backscattered from the stage. This corresponds to the elastic peak in fig. 2.12. Elastic backscattering produces the most prominent features in the TTD, which are best known and studied. This is therefore defined as standard time-shifting electron-dynode interaction and the abbreviation is usually omitted, since the direction classifier (see below) is always given, so no possibility of confusion with R exists.

C Elastic cascade branching: For d2+, whenever elastic scattering of an SE occurs, this *automatically* produces a new cascade. This is defined as standard production method for cascade branching and the descriptor is normally omitted.

Type (continued)

I Inelastic backscattering: The electron is backscattered out of the dynode either *elastically* but loses a fraction of its energy on the way in or out, or inelastically (e.g. plasmon scattering). Due to the reduced kinetic energy it produces less SEs in the next step. This generates the events in the inelastic plateau as well as the plasmon loss peaks in fig. 2.12. Two sub-effects are distinguished depending on SE production before backscattering:

(I)Λ Loss processes without SE production or starting a separate cascade (through Λ_p or Λ_s). The electron is (in)elastically backscattered out of the dynode after incurring the losses, or loses energy on the way out. Λ_p can at the same time cause both energy loss and backscattering. The I in the descriptor can be left out.

(I)B Inelastic cascade branching: The electron is elastically backscattered out of the dynode with a fraction of its original kinetic energy, but regularly loses energy while producing SEs before or after scattering. If the backscattered electron hits an active area of the PMT, it can create a second, independent cascade. The energy normally used for SE generation then is split up on two cascades, each with reduced total charge compared to regular amplification.
The I in the abbreviation is usually omitted, since it is not needed for unambiguity.
The abovementioned loss processes – with the exception of Λ_i, which would prevent the second cascade – can affect both the SE production and the backscattered primary.
Elastic cascade branching (C) is not B, since in C the primary (an SE) does not use part of its energy on SE production. For example, inelastic cascade branching of a d1 SE on d2 is described by s1:IBL:2, which can be shortened to BL2 without losing unambiguousness. The same effect for elastic cascade branching is s1:CL:2, abbreviated to L2. Both effects are shown in fig. 2.42d.

The inelastic plateau in the SE energy spectrum consists of mainly B and in small part Λ_p in form of bremsstrahlung. Plasmon peaks could also be present.

X_s To summarize all possible scattering effects (ε, C, Λ, B), the variable X_s can be used.

X_sx
The emission of a particle x (e.g. an shn or a γ) before scattering of type X_s occurs (compare Rγ, mentioned under γ). This *does* reduce the backscattered energy.

Direction (backscattering)

The scattering direction affects the time shift of the ensuing pulse if the electron hits an active structure, or whether the particle is lost and no pulse is produced. Fig. 2.39a and 2.16 help to visualize the possible paths and the solid angle occupied by other structures.

E Forward scattering toward the next stage, resulting in the pulse being *earlier* than regularly (cf. figs. 2.40c and 2.42c).

L Backward scattering toward the previous stage, resulting in a *later* pulse (see figs. 2.40d and 2.42d).

Enn, Lnn

Intradynode scattering onto the same stage n. Enn denotes a hit closer to the next stage, Lnn nearer to the previous stage (compare figs. 2.40e and 2.40f).
Both Lnn and Enn can result in a *later* pulse due to the intradynode drift time and a longer path to the next dynode from the impact site. If the interdynode drift time is reduced by more than the intradynode drift time, however, the pulse leaves the PMT a little *earlier* than normal [104] – this is only possible for Enn.
For Lnn the scattered particle hits n directly, whereas for L it is scattered away from the surface but accelerated back toward n through the fields; the transition between both is fluent.

Eij, Lij

Generalized formulation for scattering off i onto j.
E = Ei(i+1), Enn = Eii, L = $L_r$ii (r for refocused), Lnn = $L_d$ii (d for direct). It is possible for particles to skip stages and be focused onto the dynode after next or later. E.g. if d1 is hit on the part close to k, E13 should be possible.

T Backscattering losses (BSL): The electron hits a non-active part of the PMT and *terminates* there without producing secondaries. This affects the largest part of backscattered pk – the solid angle acceptance regions for the other target areas are smaller – and contributes to the ME being < 1.

X_d To denote any of the above directions, the term X_d can be used.

E(nn), L(nn) and T can occur for both elastic and inelastic scattering. The original, phenomenological terms “early” and “late” pulse were kept for forward and backward scattering due to their clarity and familiarity. Alternatively, forward scattering can be denoted by F and backward scattering by B, which requires to rename inelastic cascade branching to ζ, and collection losses to K.

Ξ Emission: For secondary electrons (sn, shn, san) emitted from a stage, the emission direction plays a similar role as for backscattering. Therefore, the same direction labels are used: E denotes emission toward the next stage, L toward the previous; for Enn/Lnn the same stage is hit closer to the next/previous.

[104] Assuming that SE production takes a negligible amount of time compared to drift times

Since the interaction (the emission) occurs for the same stage n as the origin, the complete emission process is written as sxn:Ξ(E/L):(n/nn), where Ξ can be omitted to shorten the syntax. To preserve unambiguity, an exception to the general notation is introduced: For a particle whose interaction location is the same as its origin, the direction label refers to emission instead of scattering, if Ξ or X_s is omitted.
sn:E:n then denotes the normal emission process during regular amplification (small green arrows in fig. 2.42a): An SE from dn is emitted toward d(n+1). Such secondary electrons form the regular cascade and are indistinguishable from it. If the same effect occurs for SEs with higher emission energies (shn or san), however, the transit time to the next stage will be shorter and a discrete pulse might result (red in fig. 2.42a).
For sn:L:n the SE is emitted in direction of the previous stage (due to emission angle variations) but then is refocused onto its origin stage by the fields (blue in fig. 2.42a).
In the case of sn:E/L:nn (fig. 2.42b) the SE strikes its origin stage with its small emission energy and either is backscattered, knocks out 1 SE or is absorbed.

- **Photon interactions**

  P Photoeffect: If a photon hits a photosensitive part of the PMT, it can produce a photoelectron (figs. 2.39a and 2.39b). This is the standard effect of a photon on a photosensitive surface and can also be written as R. Besides the cathode , sensitive structures include the dynodes (a material which is a good SE emitter is also a good photoemitter), aluminized glass (metals are photoemitters, albeit poor ones due to high surface reflectance and high electron scattering rates), the focusing grid (if installed) and the deck of the dynode mount, where the latter two are coated with cathode material by the evaporation technique employed. Looking at the deck through the cathode one can clearly discern concentric rings on deck and grid which are focused on the beads of raw materials on the wires above where the ring color reflects the thickness of the deposited layer as known from semiconductor science.

  M Reflection (M for mirror): When a photon is reflected off a surface, e.g. the (aluminized) glass or deck (see figs. 2.39c and 2.39d). Mij denotes the reflection off stage i onto j.

  (τn) Transmission through structure n (glass or cathode): This is relevant for effects, which pass the cathode without absorption, such as γ1 (fig. 2.39b). Since this would lengthen the effect description excessively without adding information, it is generally omitted.

- **Regular pulse (RP)**
  A pulse is labeled regular pulse if it is produced by regular interactions on all normally involved stages, i.e. photoelectron production on the cathode and subsequent regular amplification on all dynodes. This pulse then has the same transit time T_0 as γk (fig. 2.39a), which is the classic example of an RP. However, an RP needs *not* have regular charge, as e.g. cascade branching may take place

and then the impact energy of an electron at a stage is not fully converted into SEs (undersized RP), but also creates an additional pulse; for example for BE1, BL1 or BL2 (schematics in figs. 2.40c, 2.40d and 2.42d; exemplary waveforms of BE1 and BL1 in fig. 2.52). In this case, depending on the transit time of the additional pulse, the RP either becomes the PP or an AP.

- **Split pulse (SP)**
  When an effect splits the pulse charge on two (or more) pulses. This is a phenomenological effect group, as several effects can cause split pulses, such as elastic cascade branching (C), inelastic cascade branching (B), intradynode cascade branching where the backscattered electron hits the same dynode (E/Lnn, can be elastic or inelastic), afterpulses created by γ_b emission where the bremsstrahlung photon carries a notable fraction of primary energy, and in principle even iAPs, since the ionization of a residual gas molecule/atom reduces the kinetic energy of the electron.
  In the literature, the term split pulse is used for a single B-type effect, namely BL1 (regular pulse + late pulse after inelastic pk cascade branching on d1, which divides the charge among them) [190, 276].

- **Double pulse (DP)**
  Effects, which create two pulses. Again, this is a phenomenological group. Double pulsing can occur due to a split pulse or other effects, e.g. all photonic afterpulses (γAP, see below). Some groups use the term only for BL1 and synonymous to split pulses.

- **Afterpulse**

  iAP Ionic afterpulses (see 2.5.2.3): The classic afterpulses caused by ionization of residual gas molecules/atoms between k and d1, where the ions drift to the cathode and produce one or more SEs on impact. The short notation for iAP production is $\mathrm{i}\Phi\mathrm{v}_{\mathrm{k1}}$, where i can be substituted by the respective ion species, and v_{k1} is the (imperfect) vacuum between k and d1. After drifting to the cathode, the interaction there is iRk, where R refers to the typical amplification for this ion type.
  If the ion hitting k originates from the vacuum $\mathrm{v}_{\mathrm{n(n+1)}}$ between n and n+1, this is denoted by inRk.

  iiAP Interdynode ionic afterpulses: Ions are also created in the vacuum between dynodes from ionization through SEs. In fact the electron density is far higher in the cascade between dynodes than for k–d1, so the probability is increased dramatically. The ions typically drift to the next dynode, where they create underamplified afterpulses, but they can also hit the cathode (see 2.5.2.3). The notation for production is $\mathrm{i}\Phi\mathrm{v}_{n(n+1)}$ for ions originating between dn and d(n+1). inRn denotes an ion from $\mathrm{v}_{n(n+1)}$ which hits dn and creates an AP on dn, inRk hits the cathode.

  BAP Branching (or electronic) afterpulses originate from any effect with cascade branching. If the second pulse (the AP) of the branching is meant, an A can be added as prefix to the interaction type, e.g. for

s1:L:2 (L2) the AP would be referred to as s1:AL:2, or short AL2. If only afterpulses are compared, the A can be omitted, since then only the origin process is relevant.

γAP Photonic afterpulses: This group contains all effects creating photons, which subsequently hit a photosensitive part of the PMT and create a photoelectron which starts a new cascade (figs. 2.43a–2.43c). γmn denotes a photon from m hitting n. For example a γAP created on d1 which then hits the cathode is written as γ:R:1k, or short γ1k. Typical waveforms are shown for γ5k and γ51 in figs. 2.52d and 2.52e.
Photons can be produced by a cascade due to bremsstrahlung, cathodoluminescence, excitation and ionization of residual gas, or photoluminescence (fluorescence or phosphorescence), see 2.6.1.

EAP Early afterpulses (see 2.5.2.2): Again, this is a phenomenological group and contains multiple effects, such as all effects with cascade branching (BAP) and photonic afterpulses (γAPs).

LAP Late afterpulses (see 2.5.2.4): Afterpulses which occur at delays beyond what is explainable by ion drift times. The phenomenological term is out of necessity, since presently the origin of LAPs is uncertain.

Location

In fig. 2.39a the most important structures are shown; please also confer figs. 2.1 and 2.16.

g Aluminized glass

gg Glass

k Cathode

c Accelerator between cathode and d1

f Focusing grid between cathode and d1

fi Inside of focusing grid, facing d1

m Dynode mount or deck

mi Inside of dynode mount

s Light shield

n n-th dynode

ij From stage i to j, e.g. photon emission from i hitting j. For clarification a comma can be placed between i and j.

nn Intra-dynode effect on dn: When a second effect occurs on the same dynode. For example p:BL:11 (BL11), where a pk after SE creation on d1 is elastically backscattered onto a position on d1, which is closer to the cathode (fig. 2.40f). This introduces a lateness for the pulse created by the scattered pk due to the drift time d1–d1 and the increased drift time d1–d2 compared to the other pulse (exemplary waveform shown in fig. 2.52f).

v_{ij} Vacuum between stage i and j, e.g. v_{k1} for the origin of ionic afterpulses and $v_{n(n+1)}$ for iiAPs from between dn and d(n+1). If iiAPs are quoted without specifying their origin, v_{dd} is used.

no Outside of dn

a Anode

b Voltage divider or base

e Epoxy or glue cast around the VD or PMT

o Output connector

x Any location

Resources for reading the notation

The compact notation can be a bit taxing to decipher without practice, since the abbreviations of the constituents have to be looked up in the lists on pp. 188–197 and mentally converted into physical interactions.
Therefore, four aids were included for the reader's convenience:

- The most important interactions are depicted *spatially* in figs. 2.39–2.43, which should help to visualize an effect. To aid understanding of the notation and abbreviations, the figure captions show the mixed notations for all depicted effects.
- Fig. 2.52 shows measured waveforms for some selected effects, which can assist in picturing the resulting voltage pulses.
- In addition, the interactions which occur in a specific effect are described in detail on pp. 273–294 (EPs) and 315–327 (LPs) for all relevant effects.
- Finally, in tables 2.2 (EPs) and 2.3 (LPs) all calculated effects are listed with compact and topological notation, all probability factors (2.3.5.3), and the relative charge and transit time formulas (2.3.5.4 and 2.3.5.5) for PPs and all occurring APs. In addition, the tables show calculated values for the R5912 of the probabilities, charge, time, delay, charge sum and charge fraction for all pulses (terms explained later).

Effect distributions

To display the timing of measured pulses and calculated effects, various distributions are used in this work. These either depict the dependence between two effect properties (1D histograms; e.g. probability over time – the TTD) or three (2D histograms; e.g. charge vs. time with probability indicated by the bin color – the CTD).
To provide an overview, the names and contents of all used timing distributions are summarized here, and the differences between them are clarified.

Independent of whether a pulse is a primary pulse (PP, first pulse in waveform) or an afterpulse (AP, all subsequent pulses), it is an early pulse (EP), if its transit time T is smaller than T_0 (the mean transit time of γk, the standard effect). If $T > T_0$ it is a late pulse (LP). An AP can only be an EP, if the PP is even earlier and the delay of the AP (transit time difference between an AP and the PP) is short enough.

An RP is a pulse which results from regular electron production on all stages and thus occurs with a mean transit time equal to T_0. RPs can have undersized charge, however, since other interactions can withdraw a fraction of the impact energy or remove electrons from the cascade.

A 2D distribution is the same as the corresponding 1D histogram with an additional pulse parameter resolved by adding an extra axis. This greatly helps to distinguish, identify, and characterize effects; especially, when multiple 2D histograms depicting different pulse properties are viewed. 1D distributions can be considered projections of a 2D histogram onto the x-axis, i.e. integrating over the extra axis. For this reason, the total probabilities of effects are easier to read from the 1D distributions, which is a very important means to characterize the time response of a sensor and justifies the use of 1D distributions. In addition, 1D histograms are far easier to measure.

- **1D distributions**

  TTD Transit time distribution: x = transit time of pulses, y = number of pulses / probability. When no prefix is specified, the PP-TTD is meant.

  PP-TTD The TTD for the primary pulses. Commonly used to characterize the time response of a sensor.

  EPP-TTD Fraction of the PP-TTD for early primary pulses (PPs which are EPs), see fig. 2.44. If an effect produces an EP and an (undersized) RP, the EP enters the TTD, while the RP becomes an AP.

  LPP-TTD Fraction for late primary pulses (PPs which are LPs), see fig. 2.46. If an effect produces an RP and an LP, the RP is present in the TTD at T_0 (usually indistinguishable from γk), while the LP is only visible in the AP distributions. Merely LP effects, which create only one pulse or which shift all created pulses behind T_0 can show up at times $> T_0$ in the LPP-TTD (discussed on p. 294). Thus, the LPP-TTD contains far less resolvable effects than the EPP-TTD, where effects producing RPs *are* visible.

  AP-TTD The TTD can also be plotted for the APs, but since the PPs are smeared in time by TSEs (as indicated by the PP-TTD), so are their APs. This makes it hard to disentangle effects, because it is ambiguous whether APs at a certain time originate from earlier PPs (larger delays) or later PPs (shorter delays). Thus, the AP-TTD is only of use for very specific questions (e.g. the fraction of RPs which become APs) and not shown here.

  DD Delay distribution of afterpulses (see figs. 2.54 and 2.65): x = delay of APs after the PP, y = number of pulses / probability. Commonly used to characterize AP behavior. The DD resolves the ambiguity found for the AP-TTD, since each AP is related to its PP by calculating the intermediary delay.

- **2D distributions**

  CTD Charge–transit time distribution: x = transit time, y = pulse charge, z = number of pulses / probability. The charge-resolved TTD. This helps to disentangle effects overlapping in time, since often their charge differs. If no prefix is used, the PP-CTD is meant.

  PP-CTD The CTD of primary pulses.

  EPP-CTD For early PPs (see fig. 2.45). The charge-resolved EPP-TTD; all EP effects producing RPs are visible.

  LPP-CTD For late PPs (see fig. 2.47). The charge-resolved LPP-TTD; LP effects producing RPs are not resolvable.

  AP-CTD In the charge-resolved counterpart of the AP-TTD, the ambiguity between PP time and AP delay is largely removed, since the charge information helps to distinguish effects (fig. 2.60). Therefore, this plot provides complementary information to the CDD (the AP time) and was studied here (see p. 373).

  CDD Charge delay distribution of afterpulses (see figs. 2.56 and 2.66): x = delay of APs after PP, y = afterpulse charge, z = number of pulses / probability. The charge-resolved DD. For the identification of AP effects this has a similar impact as the shift from TTD to CTD for PP interactions.

  FDD Charge fraction–delay distribution (see figs. 2.61 and 2.67): x = delay of APs after PP, y = afterpulse charge fraction (charge of an AP divided by charge sum of this AP and the PP), z = number of pulses / probability. The DD with resolved charge-fraction. This helps to distinguish effects with similar charges but different correlations between PP and AP charge with delay (most notable for BL1 and γnk: compare fig. 2.56 with fig. 2.61).

- **AP distributions**

  Subscript a This denotes that the distribution contains all occurring afterpulses (e.g. DD_a, fig. 2.55a), which corresponds to plotting the AP rate (number of APs per PP, can be >1; see 2.5.2.1 for definition of rate and probability).

  Subscript 1 In this case only the *first* AP on a waveform enters the distribution (e.g. DD_1, fig. 2.55a). This is helpful to remove higher-order afterpulses (APs created by APs) and is equivalent to plotting the AP probability (whether one or more APs occurred, ≤ 1). However, the histogram is subject to shadowing (see p. 164).

  Subscript 1u The distribution of the first APs after unshadowing (e.g. DD_{1u}, fig. 2.55b). This removes the statistic effect of shadowing, but can not account for correlations between AP types (see p. 357) and thus in general does not reproduce the distribution for all APs. The unshadowing turns the AP probability into the rate of first APs, which can be >1.

Prefix To avoid ambiguity, the AP type can be specified by a prefix (EAP-, iAP-, LAP-; see 2.5.2 for an overview of types), which for the AP-TTD and AP-CTD replaces the existing prefix "AP-". Examples: EAP-DD and EAP-CTD.

2.3.5.2 Relevant effects

From the plethora of possible combinations of particle, effect and location, only a few classes of effect types are observable by producing transit time shifts or afterpulses:

R Regular amplification on all stages produces a regular pulse; if no additional interactions take place this results in γPk-pR1 = γk, which is the benchmark for all other effects.
Regular particle interactions on all involved stages can, however, also cause *irregular* pulses if stages are skipped or non-standard stages partake. Examples are cascades started by photoeffect on d1 (γP1 = γR1 = γ1), the aluminized glass (γg), the deck (γm) or the focusing grid (γf), and photoelectrons hitting the aluminized glass (Rg), the deck (Rm), the grid (Rf) or d3 (pR3 = p3).
In addition, all effects occurring on later stages, such as E2 or L3, have regular amplification in the first stages before a deviation from γk occurs.
For some effects it is necessary to explicitly state when regular amplification occurs on an unusual stage, for example for E1-L2-R1. Here first a pk is forward scattered onto d2, where it is backward scattered towards d1 and, since its kinetic energy exceeds the potential difference between d1 and d2, hits d1 and initiates regular SE production with basically the normal amount of energy deposited in d1 (the backscattering losses are negligible).

En, Ln Elastic forward/backward scattering of the standard particle on a stage leads to the most significant effects. Tn only causes losses and does not affect timing.

Enn, Lnn
Scattering can not only occur between stages, but also *stage-internal*, with far smaller time shifts.

Eij (j > i + 1), Lij (j > i)
Stage skipping after scattering can occur with a small probability.

snEn, snLn
Instead of the primary, one (or more) of the produced SEs can be emitted towards the previous stage, causing L-type behavior. Emission in direction of the next stage (E-type) leads to regular behavior, unless the secondary has higher than usual energy and arrives earlier than the rest of the SEs. Secondary electrons can have up to about 50 eV [268; p. 5], which is more than sufficient to create notable drift time differences

compared to the mean emission energy of ca. 5 eV. High energetic secondary electrons emitted in E- and L-direction are denoted by **shnEn** and **shnLn**.

sanEn, sanLn
If one of the secondary electrons is an Auger electron, this automatically has above-average energy compared to normal SEs, and both L- and E- type effects are possible.

snEnn, snLnn
An SE can only produce secondary electrons and a cascade on the dynode it originates from, if it has sufficient emission energy, since the potential is identical at exit and entry. For a typical emission energy of 5 eV a δ of 0.25 is expected for Cs_3Sb, which can result from either elastic backscattering (counted as SE due to the low energy) or actual SE production (likely only a single SE). This would create a 1 SE cascade once every four times. For small emission energies δ rises about linearly with energy. High energy SEs and Auger electrons thus can have a $\delta > 1$; the respective effects are labeled **shnEnn**, **sanEnn**, **shnLnn** and **sanLnn**.

snEij, snLij (j > i + 1)
Depending on emission position, energy and vector, stage skipping can occur for secondaries with a certain probability.

C, B All E- and L-effects can be subject to elastic/inelastic cascade branching.

Λ, T, ζ Loss effects can either completely eliminate the pulse or reduce the charge. Λ in particular can also affect the particle energy and thus its transit time. These interactions can always happen as an additional effect, albeit with reduced probability since then another interaction with likelihood $p < 1$ is stacked. The additional event classes created by T and ζ are not covered here, Λ is not treated separately, but discussed together with the respective E/L effects, since it can apply a secondary time shift.

γn As mentioned for type R, *external* (incident) photons hitting a photosensitive part n apart from the cathode produce separate event classes.

γij *Internal* (secondary) photons created inside the PMT on structure i can hit another stage j and produce an independent cascade there.

M All photons can be reflected or diffusely scattered from one structure onto another.

Φ Ionization of residual gas particles gives rise to the class of ionic afterpulses and, if caused by the pk, should also introduce a short lateness in the transit time of the primary pulse.

In addition, many combinations of the above effects are possible. Dual interactions were studied for the following type groups, including X_s subtypes (see 2.3.6):

(sxn)En-(sxm)E/L:m with m > n

(sxn)Ln-(sxm)E/L:m with m ≥ n (special case: L1Ef)

(sxn)E/L:nn-(sxn)E/L:n

(sxn)En-(sxm)E/L:mm with m > n

(sxn)Ln-(sxm)E/L:mm with m ≥ n

Any interaction plus photon emission

Based on these interaction type groups, also triple or even quadruple interactions can be formulated.
In principle, all possible combinations are expected to occur, but the probability drops with each additional interaction and is small to begin with for many effects. In addition, some effects are indiscernible from the regular pulse, e.g. s1E1 (these *are* the regular SEs emitted from d1).

Effects starting amplification beyond d4 were not covered. For one thing the produced charge is too small to be observed (as it turns out, no effect starting amplification on d4 produces observable charges either). For another thing, due to the high number of electrons (i.e. number of trials) present in the cascade at these stages, effects occurring there are likely to constitute an always present statistical contribution to the pulse shape. With rising occurrence rates the relative fluctuation of the number of observed events shrinks, making it less likely to observe differences to the average pulse shape.

At the other extreme, the relevance of multiple interaction effects is limited by probabilities dropping below the statistical detection limit, in which case they also play no practical role for sensor performance. Which effects were included in the study was therefore also based on expectable probability and capped at a certain number of interactions. For triple interactions or more, only the most dominant effects were combined as only then the possibility for observation is given. Nevertheless, it is difficult to predict which effects are observable before actual calculation, so many effects had to be included in the calculations, which turned out to be of no practical relevance.

Focusing grid

In some PMTs a focusing grid is installed between cathode and d1. This is the case for e.g. all current Hamamatsu PMTs with 5" diameter or more [213, 330] and the ETE PMTs 9354 (8") and D784 (11") [262, 263, 306].

The presence of a focusing grid introduces several additional time shift and afterpulse effects. Because the grid is covered with photocathode material during cathode evaporation as well, it is expected to show external photoeffect and emit a photoelectron when hit by an incident photon (γf) or a secondary photon produced in the PMT (γgf, γ1f or from later dynodes). This also should make it an adequate SE emitter. Thus when a pk hits the grid, it either can produce secondary electrons (Rf), be backscattered (Ef, ΛEf, BEf, Lf, ΛLf, BLf, and even LfLf) or create a photon through e.g. bremsstrahlung, which then can trigger a pulse on another sensitive

structure (γfk, γf1). Finally, instead of a photoelectron an L1 can hit the grid, where it can be subject to forward scattering towards d1 (L1EfR1) or one of the other interactions listed above.

Whether or not these effects actually lead to observable pulses depends on the field distribution set by the voltage divider, however.
If the grid potential lies between that of d1 and d2, as is the case for the Hamamatsu R5912 studied here, the grid acts as a moderate accelerator electrode. This means that the pk first is accelerated by a potential gradient even stronger than from d1 alone, which is beneficial for CE, transit time and TTS. However, after passing the grid, the pe is actually decelerated slightly towards d1. This has the added benefit that it suppresses effects based on photoelectrons or SEs originating from f (γf, γgf, γ1f, Rf, afterpulse of BEf, primary pulse of BLf), since they cannot overcome the potential barrier to d1 (28 V for the R5912) with their initial energy. Only the most energetic SEs can reach d1, but their generation probability is strongly suppressed, which is why this was not studied here. An accelerator-type grid is found for the Hamamatsu R12860 (20"), R11780 (12"), R7081 (10"), R5912 (8") and R6594 (5") [213, 330] [105] and the ETE D784 (11") and 9354 (8") [262, 263, 306].
For a grid potential at d1 level, SEs from f can reach d1, although with very slow drift velocities and low gain $\delta 1 < 1$. This means that despite the short distances the transit times will be rather long and that only for a fraction of d1 hits a pulse is produced. This layout is the case for the R11833-03 (5"), which has a low gain and bad TTS, however, and thus is not considered in this work.
If the potential lies between k and d1, the full scope of grid effects become observable without limitations. This configuration is found for the R1250 (5" linear-focused, very fast) and R877 (5" box-and-grid, extremely large TTS). It should be noted that since the SEY has a diminished return for increasing voltages, such a configuration can lead to gains even *higher* than for γk for effects producing SEs on f, as the grid can act as an additional amplification stage.

The grid is assumed to be of accelerator-type in the following, as is the case for all large-area PMTs considered here.

Effect schematics

In order to give an overview over the TSE types discussed in this work, in figs. 2.39–2.43 the processes relevant for understanding are depicted schematically. The individual effects shown in the graphs will be discussed in detail in section 2.3.6 and 2.5.2.2. The PMT layout is based on the Hamamatsu R5912 and roughly to scale, to the best of available information ([213; p. 38] [169; p. 45] and measuring of a specimen). Λ-based effects for E-type scattering have similar trajectories to the base effect (e.g. ΛE1 resembles E1, ΛE2 is akin to E2). The reduced initial energy modifies the paths somewhat but principally leads to increased transit times. For L-scattering, Λ-effects resemble the B-variants, since these also have reduced energies (for example the path of ΛL1 is similar to BL1).
Interactions on later dynodes are just like the effects shown for d1 and d2 (e.g. E2, L2) except for other SEs from previous stages also being present.

[105] In the VD design of the R5912 (8") and R7081 (10"), three grid potentials g1–g3 are given, where g2 is at d1 potential, and g1 and g3 lie between d1 and d2 levels. For the R6594 (5") also three grids with applied voltages between that of d1 and d2 were listed with g1 at d1 potential. For lack of labeling it was assumed that g1 corresponds to the actual grid.

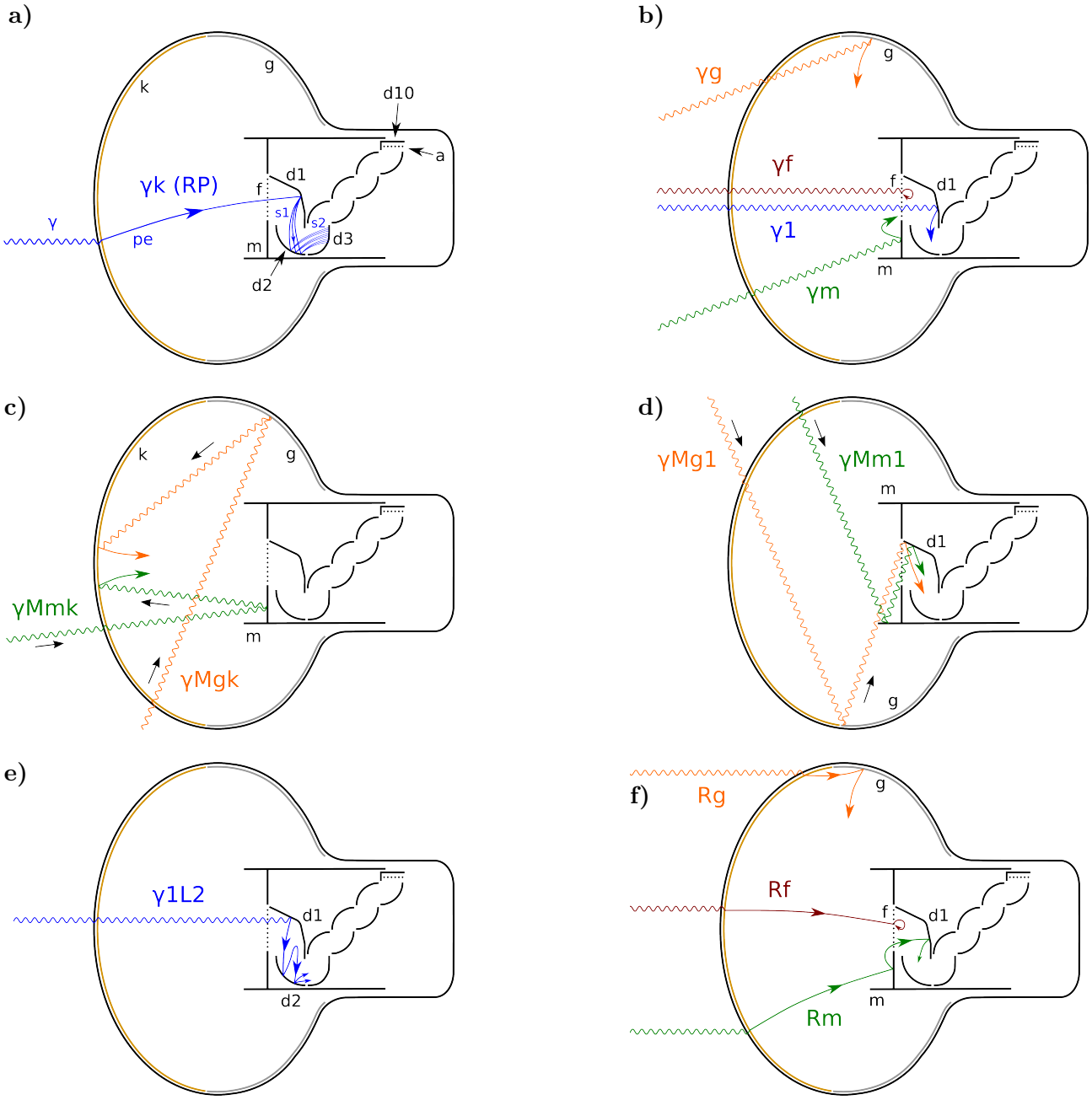


Figure 2.39: Transit time shifting effects (TSEs): a) - e) from incident photon interactions, f) from regular pk amplification on structures different from d1. PMT dimensions roughly to scale for the Hamamatsu R5912, a large area box-and-line PMT; figures based on [213; p. 38] and [169; p. 45]. Structures: k = photocathode, g = aluminized glass, f = focusing grid, m = dynode mount / deck, dn = n-th dynode, a = anode. Particles: γ / curly lines = photons; solid lines = electrons: pe / large arrowhead = photoelectron, sx / small arrowhead = secondary electrons from structure x (for dynodes x = n). Effects interactions listed in mixed notation. Discussion of effects and topographic notations see text.
a) **γk** = γPk-pkR1: Regular pulse (RP) with normal transit time ($T = T_0$) and mean charge (1 pe). Early pulses (EPs) have a transit time $T < T_0$ and an earliness of $|T - T_0|$, late pulses (LPs) have $T > T_0$ and a lateness of $T - T_0$.
b) **γ1** = γP1-p1R2: EP, large earliness due to omission of the electron drift time from k to d1, reduced charge of $\approx {}^1/_{\delta 1}$ since the gain from d1 is skipped. **γm** = γPm-pmR1: EP or LP depending on pk drift time to d1, charge ${}^1/_{\delta 1}$ – on d1 the pm produces merely 1 SE or is backscattered, as illustrated for Rm in f). **γg** = γPg-pgR1: EP/LP depending on pg drift time, charge reduced by smaller traversed potential difference and non-ideal d1 hit position. **γf** = γPf-pfTf: No pulse for accelerator-type grid, since pf is trapped at f (pf drawn with small arrow).
c) **γMmk** = γMmk-γPk-pkR1, **γMgk** = γMgk-γPk-pkR1: Both LPs with short lateness due to extra photon flight times, and regular charge.
d) **γMm1** = γMm1-γP1-p1R2, **γMg1** = γMg1-γP1-p1R2: Both EPs with earliness slightly lower than γ1 and charge ${}^1/_{\delta 1}$.
e) **γ1L2** = γP1-p1L2-p1R2: EP, later than γ1 but earlier than all EPs caused by electron interactions only; charge ${}^1/_{\delta 1}$.
f) **Rg** = γPk-pkRg-sgR1: LP, medium lateness; depending on the SEY of Al and the incident pk energy, the charge can be larger than for RPs. **Rm** = γPk-pkRm-smR1: LP, medium lateness due to low fields on the path of sm; charge ${}^1/_{\delta 1}$. **Rf** = γPk-pkRf-sfTf: No pulse, SEs trapped at f.

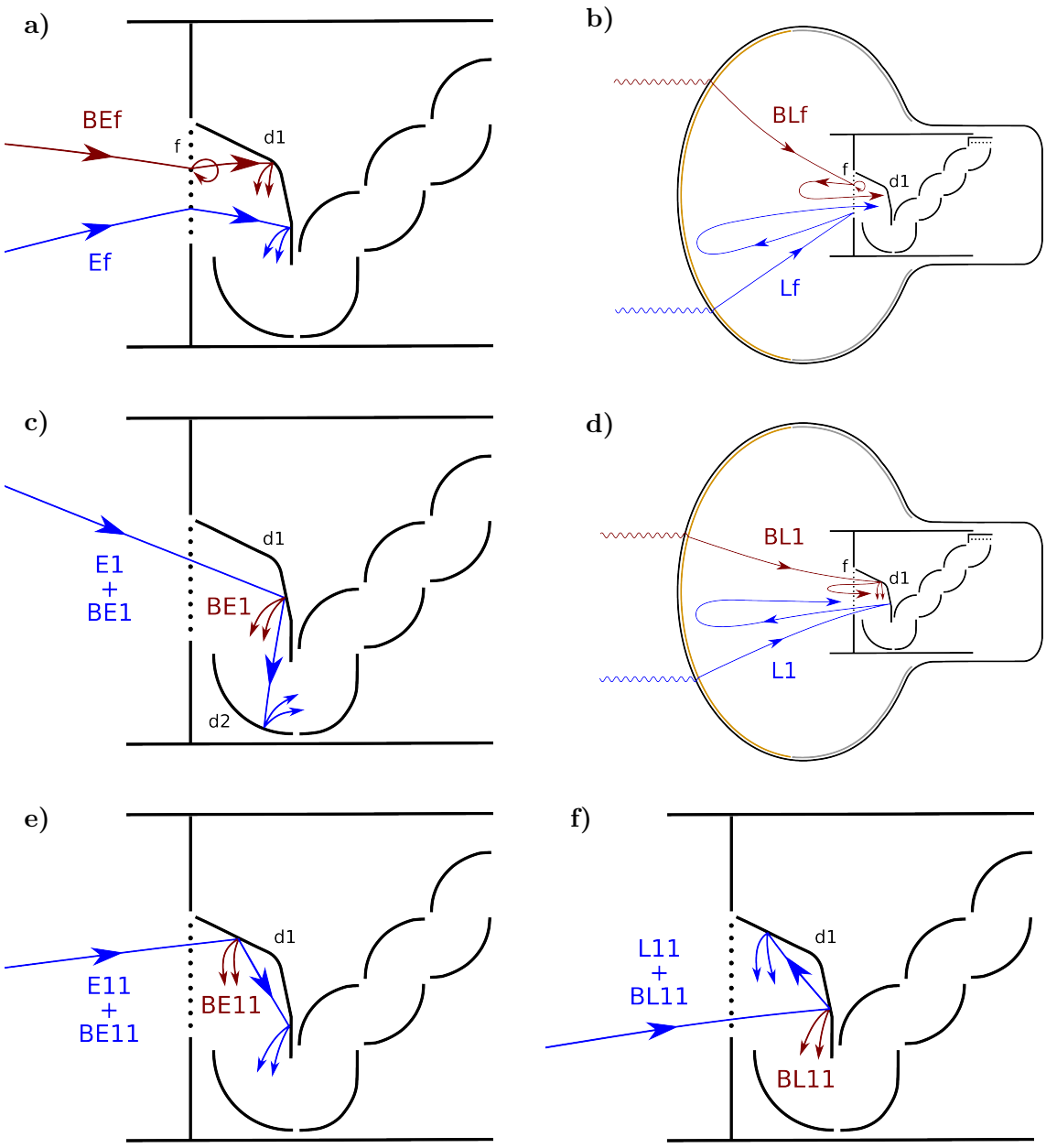


Figure 2.40: Time shift effects from photoelectron backscattering. Description see fig. 2.39, discussion of effects and topographic notations see text.
a) Ef = γPk-pkEf-pkR1: EP/LP depending on d1 hit position, charge can be reduced by d1 impact location. **BEf** = γPk-pkBEf=pkR1+afTf: EP/LP, time as Ef but can have larger lateness due to energy reduction from branching; charge as Ef or lower, down to $^1/_{\delta 1}$; no afterpulse (AP) since sf are trapped.
b) Lf = γPk-pkLf-pkR1: LP with large lateness, but shorter than L1; regular charge.
BLf = γPk-pkBLf=pkR1+afTf: LP with lateness as Lf or shorter for smaller energies, charge from $^1/_{\delta 1}$ to regular; no AP due to trapping.
c) E1 = γPk-pkE1-pkR2: EP, charge slightly larger than $^1/_{\delta 1}$ due to increased impact energy on d2.
BE1 = γPk-pkBE1=pkR2+a1R2: EP, earliness as E1 or shorter for smaller energy, charge from $^1/_{\delta 1}$ up to E1; AP formed from a1 with regular time and charge ranging from $^1/_{\delta 1}$ to regular.
d) L1 = γPk-pkL1-pkR1: LP, large lateness due to pk drift time towards k and back to d1, normal charge.
BL1 = γPk-pkBL1=pkR1+a1R2: Primary pulse (PP) is RP formed from a1 with reduced charge from $^1/_{\delta 1}$ to 1 pe; AP with lateness as L1 or shorter, charge $^1/_{\delta 1}$ to 1 pe.
e) E11 = γPk-pkE11-pkR1: EP, earliness depends on d1 hit position, charge reduced by non-ideal hit position. **BE11** = γPk-pkBE11=pkR1+a1R2: reduced pk energy after branching can increase the drift time to d1 that much that the produced pulse is later than the RP; a1 form RP with normal time and reduced charge. Depending on pe energy after branching either the pk- or the a1-pulse is the PP, the other the AP. Both charges range from $^1/_{\delta 1}$ to 1 pe.
f) L11 = γPk-pkL11-pkR1: LP with short lateness depending on hit position, charge can be reduced by impact location on d1. **BL11** = γPk-pkBL11=pkR1+a1R2: PP formed by a1 with regular time, AP is from scattered pe, lateness can be larger than for L11 due to energy reduction. Both charges from $^1/_{\delta 1}$ to 1 pe.

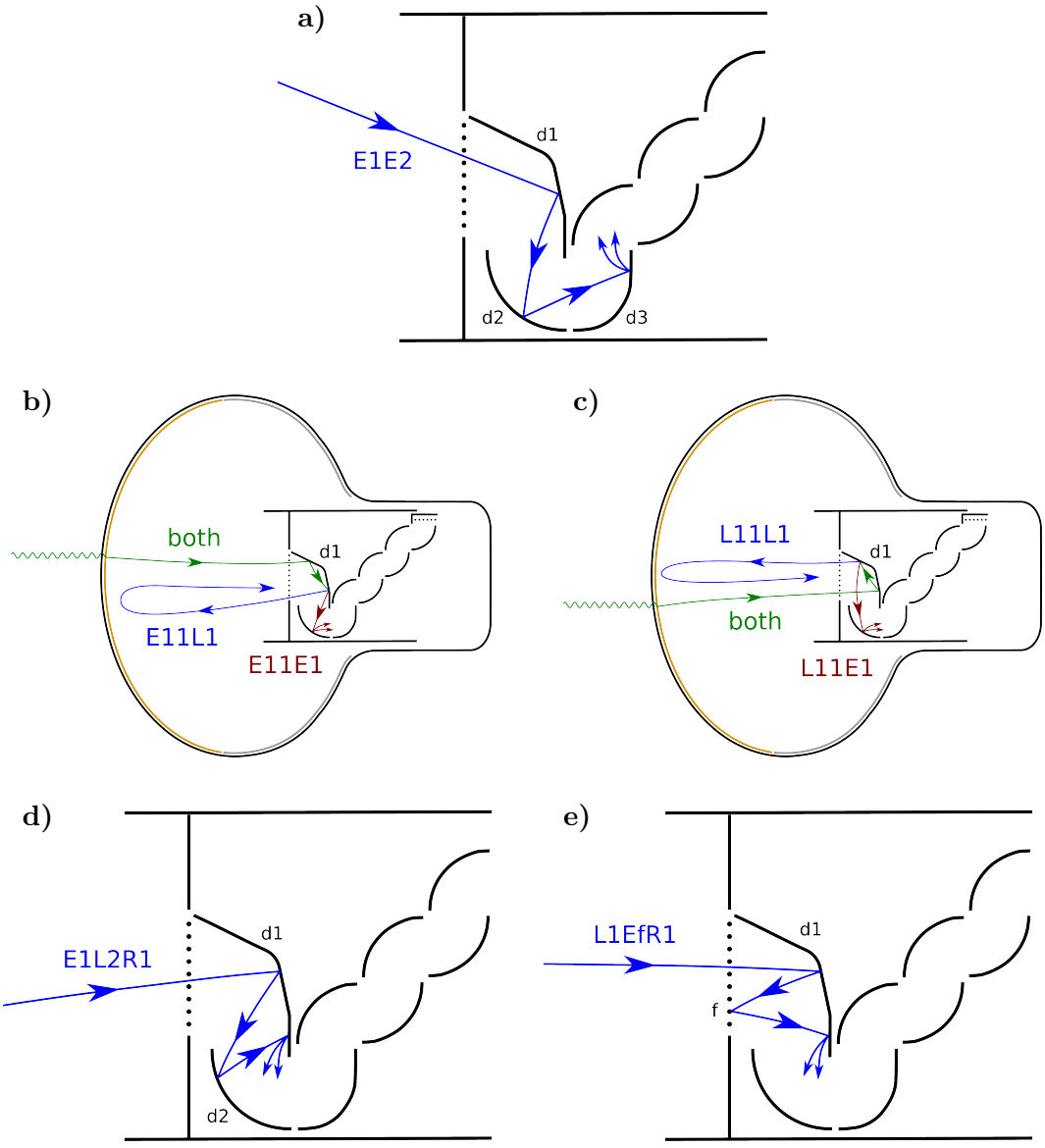


Figure 2.41: Time shift effects from dual photoelectron backscattering. Description see fig. 2.39, discussion of effects and topographic notations see text.
a) E1E2 = γPk-pkE1-pkE2-pkR3: EP, earlier than E1, charge reduced by two gain steps, charge too far below instrumental threshold for observation. For E1E2E3 the pe is forward scattered on d3 as well, causing a still earlier EP with one more gain step missed, making observability even less likely.
b) E11E1 = γPk-pkE11-pkE1-pkR2: EP, earliness as E1 or slightly less, charge as E1.
E11L1 = γPk-pkE11-pkL1-pkR1: LP, lateness as L1 or slightly shorter, regular charge.
c) L11E1 = γPk-pkL11-pkE1-pkR2: EP, earliness as E1 or less, charge as E1.
L11L1 = γPk-pkL11-pkL1-pkR1: LP, lateness as L1 or later, regular charge.
d) E1L2R1 = γPk-pkE1-pkL2-pkR1: LP with short lateness, regular charge.
e) L1EfR1 = γPk-pkL1-pkEf-pkR1: LP with short lateness, regular charge.

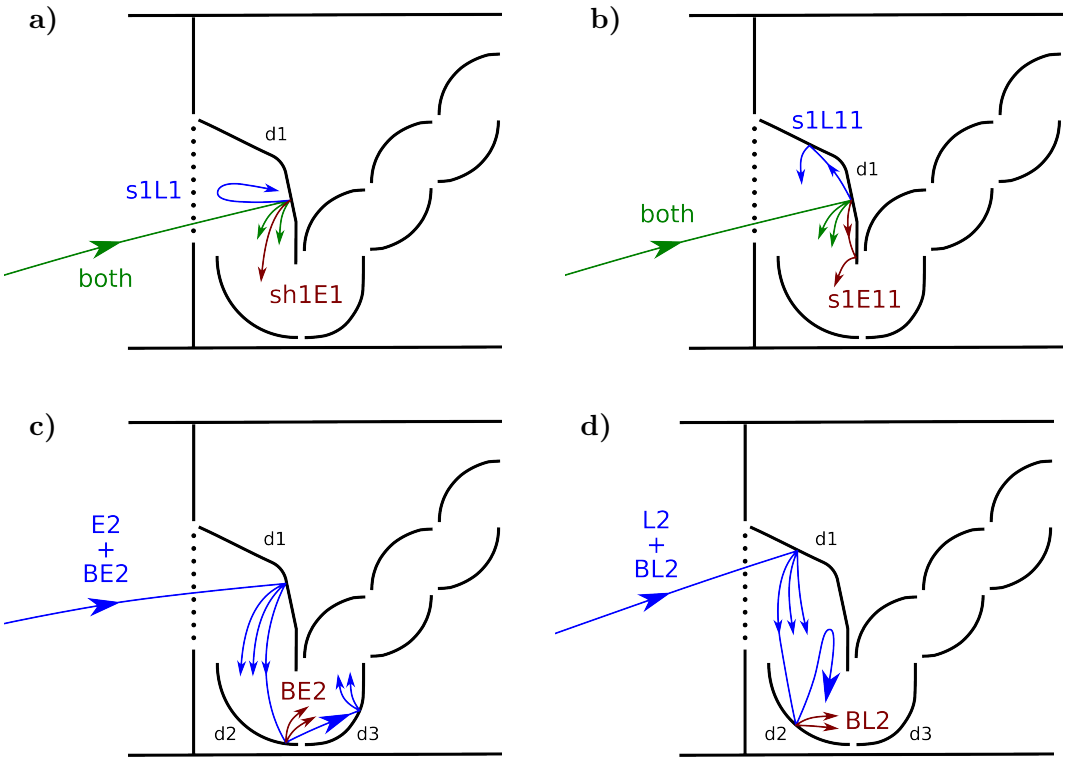


Figure 2.42: Time shift effects from interactions of secondary electrons originating from d1. Description see fig. 2.39, discussion of effects and topographic notations see text.
a) sh1E1 = γPk-pkR1=r1R2+sh1E1-sh1R2: EP from sh1, very short earliness, charge slightly above $1/\delta 1$; RP forms AP. **s1L1** = γPk-pkR1=r1R2+s1L1-s1R1: RP forms PP, s1 creates AP with short lateness and charge $1/\delta 1$ or higher for sh1.
b) s1E11 = γPk-pkR1=r1R2+s1E11-s1R1: s1 pulse has lateness due to small s1 energy and slow drift velocities, charge $1/\delta 1$ or more for sh1; the RP is thus the PP, the s1 pulse the AP. The s1E11 path curvature was drawn towards d1 for better visibility, the trajectory is actually curved away from d1 as for s1L11. **s1L11** = γPk-pkR1=r1R2+s1L11-s1R1: RP is PP; s1 pulse is AP with lateness, charge as s1L1.
c) E2 = γPk-pkR1=r1R2+s1E2-s1R3: EP with charge slightly larger than $1/\delta 1\delta 2$ due to higher impact energy on d2; RP forms AP. **BE2** = γPk-pkR1=r1R2-a2jR3+s1BE2-s1R3: EP with earliness from 0 to E2, charge as E2 or down to $1/\delta 1\delta 2$; RP is AP. Both E2 and BE2 occur with high probability due to high $\delta 1$.
d) L2 = γPk-pkR1=r1R2+s1L2-s1R2: RP is PP, s1 forms AP with lateness and charge $1/\delta 1$. **BL2** = γPk-pkR1=r1R2-a2jR3+s1BL2-s1R2: RP forms PP, AP from s1 has charge from $1/\delta 1\delta 2$ to L2, lateness as L2 or less. High probability for both effects.

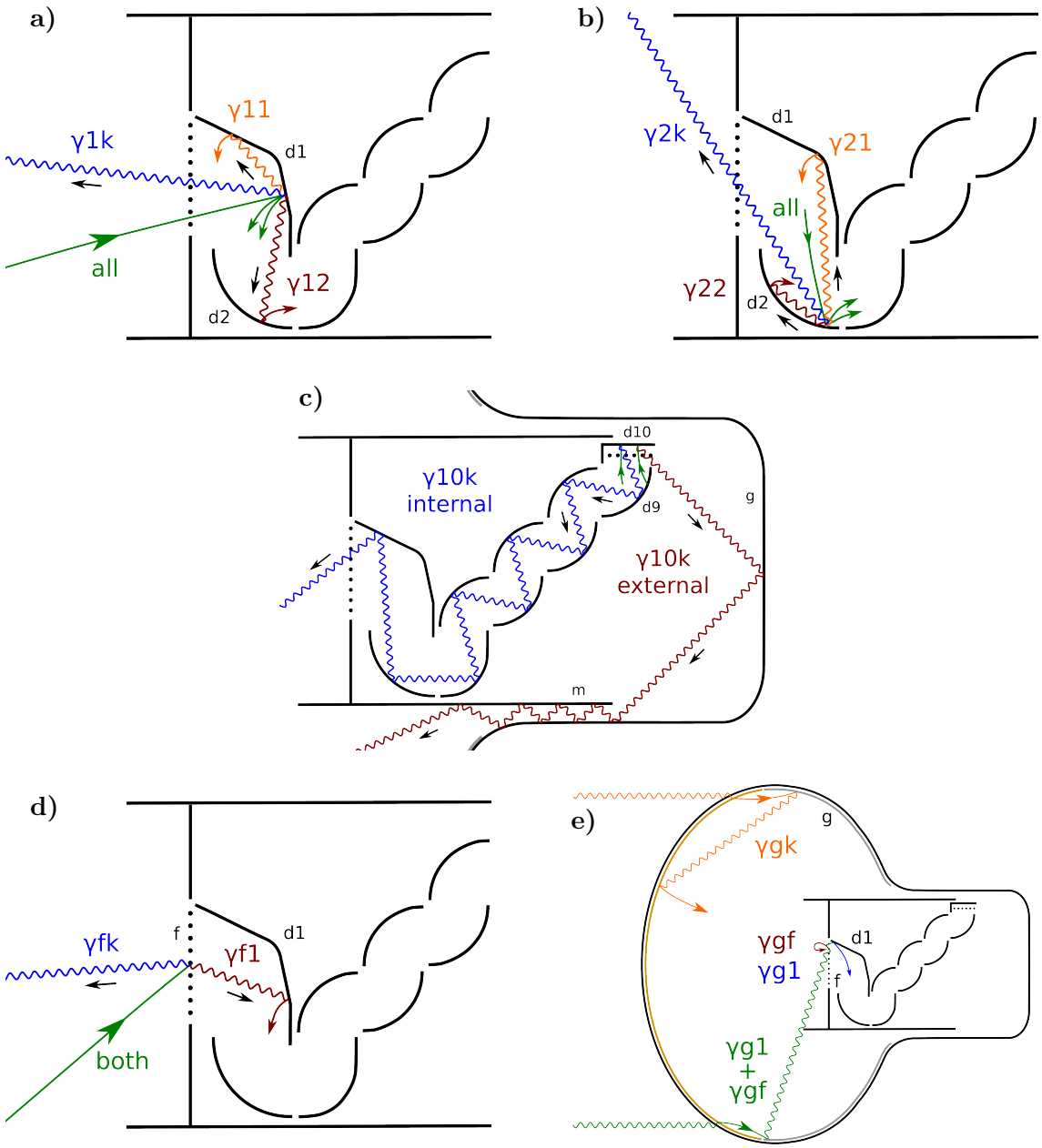


Figure 2.43: Time shift effects from secondary photon interactions. In this work, photon production following an RP is discussed; however, any effect depositing energy on the respective dynode can produce secondary photons as a byproduct. Description see fig. 2.39, discussion of effects and topographic notations see text. The charge of pulses from photon hits on k is regular, for hits on d1 reduced by $1/\delta 1$ and for hits on d2 reduced by $1/\delta 1 \delta 2$.
a) **γ1k** = γPk-pkRγ1=a1R2+γ1k-γPk-pkR1: AP, medium delay compared to RP due to pk drift from k to d1. **γ11** = pkRγ1=a1R2+γ11-γP1-p1R2: AP, very short delay **γ12** = pkRγ1=a1R2+γ12-γP2-p2R3: EP, RP becomes AP.
b) **γ2k** = a1Rγ2=a2R3+γ2k-γPk-pkR1: AP, higher delay than γ1k. **γ21** = a1Rγ2=a2R3+γ21-γP1-p1R2: AP, short delay. **γ22** = a1Rγ2=a2R3+γ22-γP2-p2R3: AP with very short delay.
c) **γ10k** = a9Rγ10=a10Ra+γ10k-γPk-pkR1: AP with large delay and lateness even above L1. Photons can arrive at k through internal or external paths, the former through reflections inside the dynode chain, the latter through gaps between the dynodes (one exemplary path shown). Note that d2 can be hit *directly* from d10. γ10,1 has the same path as γ10k–internal but is already absorbed on d1, γ10,2 produces a photoelectron on d2.
d) **γfk** = pkRγf=afTf+γfk-γPk-pkR1: LP with medium lateness, no pulse from f unless pk is forward scattered towards d1 after photon production (pE$_\gamma$fk). **γf1** = pkRγf=afTf+γf1-γP1-p1R2: EP, no pulse from f unless pk is forward scattered (pE$_\gamma$f1).
e) **γgk** = pkRγg=agR1+γgk-γPk-pkR1: LP, low probability. **γg1** = pkRγg=agR1+γg1-γP1-p1R2: EP with large earliness compared to electron-interaction EP (however, later than γ1L2R1) but very low probability. **γgf** = pkRγg=agR1+γgf-γPf-pfTf: No pulse due to trapping.

Effects known before this work

Many time shift effects have been postulated or observed throughout the past decades. Unfortunately, no review listing all of them exists, making it difficult to collect the published information. By thorough literature research the following list of all effects that have been postulated or identified previously – to the best of the author's knowledge – was compiled. The effects are listed in chronological order of discovery, as far as is reconstructible:

γ1 "Prepulses": First discovered by [331](1964) according to [170; p. 435]

γAP "Anode/dynode glow" (see pp. 379, 418): The consensus in literature is that light emission from PMTs, associated with γAPs and most prominent for the later stages, was first analyzed by [332](1967).

L1 "Late pulses": First mention by [333](1972) according to [170; p. 435]

γgk [173; p. 56](1980) lists this effect as one source of light emission from PMTs visible by eye.

BE1 [232; p. 189](1985) deduced the presence of BE1 due to deviations of the pulse shape observed for some pulses, similar to the methods employed here. [184; p. 77](2007) mentions them in passing, noting that they usually pose no problem due to the short earliness and small charge. [301](2012) assumed one observed effect to originate from the merged sum pulse of BE1 and its RP, but likely observed γg instead.

E1 "Early pulses": While the first publication mentioning them could not be established, it is clear that this effect has been known for a long time, given that is it easily observable. [334](1993) knew about them, although they were still labeled "prepulses" at the time, together with γ1. [335](2002) also labeled them prepulses but gave a correct description of the effect.

BL1 "Split pulses", "double pulses": [335](2002) mentions the effect, but attributed it to s1L1. [184; p. 77](2007) describes the physical process correctly.

s1L1 Postulated by [335](2002), but falsely attributed to observed effects: They observe EAPs within 8–60 ns delay; however, the calculations performed here for the same PMT type indicate that s1L1 are expected only up to delays of about 6 ns and that it is not certain whether the effect occurs at all due to the attracting fields of f and d2, which strongly draw s1 away from d1, making a return to it unlikely (see p. 321).

E1E2, E1E2E3 First described by [275](2004) for an ET [106] 9351 (8"). The observed early effects, however, can not originate from E1E2 and E1E2E3 (discussion on p. 291), since the instrumental threshold was at 16% pe, which strongly suppresses dual and triple site interactions – for the

[106] Electron Tubes Ltd.

similar R5912 studied here, already E1E2 lies below a detection threshold of only 4.5% pe. γm, sx1E13 and γ1L2 are possible alternative candidates for the observed behavior.

ΛL1 "Inelastic late pulses": Postulated in [275] as well in order to explain the events between main peak and L1, where it was claimed that it could explain the trailing right flank of the main peak. However, since the observed match is a result from the employed phenomenological fit function, most likely a combination of other effects such as γMgk, γMmk, ΛL11, L11, BEf, ΛEf, etc. (confer 2.3.6.2) causes the trailing behavior.

Lf Again this was first found by [275].

γfk Also found by the Borexino collaboration and first identified in [336] (2004) – in [275] the effect was still interpreted as an artifact from reflections in the laser optics.

γMmk Listed by [170; p. 74](2017). Wright observed a 10% drop in PMT output current when a light spot moving over the deck of the dynode stack disappears in the aperture to d1, which shows the magnitude of the contribution of the reflected photons to the effective quantum efficiency (QE).

γM1k Also [170; p. 74]. This effect has a smaller probability than γMmk and similar lateness, and thus is not studied here.

γg Again [170; p. 74].

All in all, the existence of 15 effects plus the various γAPs was predicted or observed before this work.

TST allows to greatly increase the number of expected effects from logical combination of interactions, limited only by the reduction in charge and probability. For this work a *selection* of 185 relevant effects was studied, which contains the abovementioned effects. Not all of these occur in every PMT, however.

Newly postulated effect types include intradynode scattering, the extension of effects onto later dynodes (d2+), interactions for sh and sa, circular effects (e.g. E1L2R1), interactions of internal photons with structures other than k, photoelectrons producing SEs not on d1, and multiple interactions (with the exception of E1E2 and E1E2E3). Nevertheless, the majority of the studied effects was obtained not from the inclusion of new interactions, but merely the consequent combination of all effect parameters.

2.3.5.3 Probability

Absolute and relative probability

The probability of occurrence of an effect is crucial to determine, whether it is observable and impairs the time resolution of the sensor.
However, only *relative* probabilities can be retrieved from the TTD or the charge–time distribution (CTD) directly, that is, the fraction of events belonging to an effect to

all recorded events. If the exact probability of one effect or the total probability of all effects is known or can be estimated, it is possible to obtain *absolute* probabilities in a second step. Relative likelihoods are completely sufficient to analyze the TTD/CTD though, since they determine the amount of trouble each effect poses.
It is most natural to compare the probability of a group of events to that of the regular pulses γk. First, because this is by far the most frequent occurring effect in the TTD, second, since it is far more resistant to threshold variations than most other effects (as will be seen), and third and foremost, because the terms required to calculate its absolute probability are routinely measured and usually well-known, since they determine sensor performance. The absolute probability to detect a photon via γk per second amounts to

$$\begin{aligned} p_{\gamma\mathrm{k},a} &= E_\gamma \cdot \mathrm{A_k} \cdot \mathrm{QE_k} \cdot \mathrm{CE_{k1}} \cdot \mathrm{u1} \cdot (1 - p_{0\mathrm{SE1}}) \cdot (1 - p_{\mathrm{T12}}) \cdot (1 - p_{0\mathrm{SE2}}) \\ &\quad \cdot (1 - \mathrm{p_{T23}}) \cdot \ldots \qquad (2.152) \\ &\approx \mathrm{E}_\gamma \cdot \mathrm{A_k} \cdot \mathrm{QE_k} \cdot \mathrm{CE_{k1}} \cdot \mathrm{u1} \qquad (2.153) \end{aligned}$$

where E_γ is the incident photon irradiance (photons per area per second), $\mathrm{A_k}$ the cross-section area of the cathode [107], $\mathrm{QE_k}$ the quantum efficiency of the cathode, $\mathrm{CE_{k1}}$ the collection efficiency of photoelectrons from k on d1, u1 the probability that on d1 the pe is *not* backscattered (unscattered) but produces SEs regularly, $p_{0\mathrm{SEn}}$ the likelihood of zero SE production on dn, and p_{Tij} the probability of total SE loss (termination) between dynode i and j. Due to the large δ1 of modern high-gain PMTs, the last four terms are already virtually indistinguishable from 1, all following SE-production and loss terms from d3+ can be neglected.
The relative probability p_r of an effect to γk is then defined as

$$p_r = \frac{p_a}{p_{\gamma\mathrm{k},a}} \qquad (2.154)$$

with p_a the absolute probability of the effect.
For most interactions, the first two terms of $p_{\gamma\mathrm{k},a}$, irradiance times cathode area – the photon flux through the cathode – are also contained in p_a and cancel out. Since all effects discussed in the following are based on incident photons [108], E_γ always cancels and is not mentioned again. Omitting the first two factors from eq. (2.152) corresponds to calculating the absolute probability of γk per photon. Using values for the R5912 ($\mathrm{QE_k} = 25\%$, $\mathrm{CE_{k1}} = 80\%$, u1= 69.3%; discussed below), this is merely 13.86% of all photons hitting k.
Taking into account that not all pulses lie above the threshold further reduces $p_{\gamma\mathrm{k},a}$: Using (2.157) (derived shortly) an instrumental threshold h_s of 1.1 mV yields 13.72%, using the fit routine developed here (threshold 0.35 mV) 13.82%. For an event to be detected it has to be recorded by the setup as well as found in the analysis (see (2.158)), so in this case the probability is determined by the instrumental threshold. This threshold-corrected $p_{\gamma\mathrm{k},a}$ – dubbed $p_{\gamma\mathrm{k},d}$ (detected probability) – is the base probability with which a p_r has to be multiplied to obtain p_a.

While the absolute likelihood of γk can be estimated with sufficient accuracy, the fraction of γk pulses to all observed primary pulses $\eta_{\gamma k}$ in the TTD, however, is

[107] This can be the whole of the cathode area for a light source surrounding the PMT.
[108] This is different for dark counts of all manner, including iAPs produced by them and radioactive decays.

not known – and not easily estimated, since it depends on all effects discussed in the following, many of which cannot unambiguously be matched with features in the TTD. Furthermore, the detected fraction of many small effects strongly depends on threshold.
Therefore it is necessary to establish $\eta_{\gamma k}$ from data. This can be done by comparing the probability calculated for the central bin of the γk peak using (2.159) with the observed value from the TTD; the required scaling factor is $\eta_{\gamma k}$. This formula assumes a Gaussian distribution. For the R5912 at g=$1.3 \cdot 10^7$ and h_s =1.1 mV a value of 78.2% is found. The remaining 21.8% of detected pulses in the TTD are caused by other effects and (by definition) produce EPs and LPs. This shows the importance of measuring and understanding the TTD.

Total probability

There are three different probabilities for each effect which are relevant here.
The total probability p_t shall denote the summed likelihood to observe a given effect once or more often per event (incident photon). For pe-based effects this just equals p_r, since only one photoelectron can be produced per photon [109]. For effects involving SEs though, one must account for the number of trials by exponentiating the probability p_1 for the effect to occur for one SE by the number of particles n. This is simply one minus the probability that the effect occurs for none of the particles:

$$p_t = 1 - (1 - p_1)^n \tag{2.155}$$

The value of n depends on the local gain factors δ, which can differ from the regular δn due to backscattering, skipping of amplification steps and loss processes.
Leading probability terms which involve single particles or are identical to γk are independent of p_1 and not exponentiated.
The total probability p_t, and not p_1, determines the probability to observe an effect, since multiple instances can occur for the same PP if more than one particle can cause it. Therefore p_t enters the calculation of p_r. In the TTD multiple occurrences for the same PP cannot be separated if the resulting pulses are close enough to be inseparable and contribute to the same distribution, in the CTD this can show in form of multiple charges. But in most cases the higher number of particles which are able to cause the effect merely increases its likelihood of observation and pileup does not occur.

To give an example, for E2 p_1 contains just the probability factors which affect the pulse production by an s2, the factors leading to the production of the s2 are not contained in p_1:

$$p_1 = \mathrm{b2} \cdot \mathrm{e2} \cdot \mathrm{E2} \cdot \mathrm{u3_{E2}} \tag{2.156}$$

with the probability factors b2 that an s1 is subject to scattering on d2, e2 that the scattering is elastic, E2 to be scattered forward such that d3 is hit, and $\mathrm{u3_{E2}}$ that on d3 no backscattering occurs - which would lead to other, secondary effects.
Substituting (2.156) in (2.155), and plugging its result, the probability factors which do not enter p_1, and (2.153) into (2.154) results in the relative probability p_r for E2:

$$p_r = \frac{\mathrm{QE_k}}{\mathrm{QE_k}} \frac{\mathrm{CE_{k1}}}{\mathrm{CE_{k1}}} \frac{\mathrm{u1}}{\mathrm{u1}} \left(1 - (1 - \mathrm{b2} \cdot \mathrm{e2} \cdot \mathrm{E2} \cdot \mathrm{u3_{E2}})^{\delta 1}\right) = 1 - (1 - \mathrm{b2} \cdot \mathrm{e2} \cdot \mathrm{E2} \cdot \mathrm{u3_{E2}})^{\delta 1}$$

[109] Double and triple photoelectron emission can occur in vacuum ultraviolet (VUV) PMTs [192,193], but this is irrelevant here.

In the following, p_r is shortened to p, since it is the basis for the subsequent corrections.

The expected number N of occurrences per event of an effect is calculated by simply multiplying all probability factors involved in p with each other – i.e. all p_1 and multiplicities δ, since at least for the first dynodes no dependences and covariances are expected. The probability factors, which are not contained in p_1, are a prerequisite that one or more pulses of this effect can occur at all (i.e. that there is such an event) and thus do not enter N. For E2, $N = \mathrm{b2} \cdot \mathrm{e2} \cdot \mathrm{E2} \cdot \mathrm{u3}_{\mathrm{E2}} \cdot \delta 1$.

Detectable probability

For a real measurement also the influence of the threshold on event detectability has to be considered. p corresponds to an ideal, unreachable threshold of zero; it states how many pulses of this type are produced in the device and could be observed for a perfect setup without noise. To compare theoretical predictions with data, it is necessary to take into account which fraction of p is observable due to pulse height. Two steps are required (in the method used here) to observe a pulse. First, data acquisition has to be triggered by a PMT pulse which exceeds the instrumental threshold of a discriminator h_s and lies within an acquisition window after a triggered light pulse. In a second step, the recorded waveforms are analyzed offline by a pulse-search algorithm employing a fit routine. This also has a detection threshold in form of a minimum detectable pulse height h_f due to the difficulty of distinguishing noise spikes from events for too small pulses. The fit threshold h_f is smaller than h_s of the setup. This has the consequence that a small pulse which is larger than h_f not necessarily has to be larger than h_s to be observable, if another, larger pulse exceeding h_s is present in the waveform – otherwise the acquisition is not triggered by the setup. This is not the case for all effect types, so this threshold extension through data analysis is only partially possible and will not reveal all effects which lie below h_s but above h_f.

To calculate the fraction η_p of p lying above a generic threshold h_t, a Gaussian probability density function (PDF) is assumed for the height and charge distribution of the pulses of an effect. For most effects this is a valid approximation, for the few exceptions it at least allows an order of magnitude estimate; the estimation could be refined by using a more appropriate PDF for these cases.
The fraction η_g of a Gaussian function with mean μ and standard deviation σ which lies above a threshold h_t is

$$\eta_g = \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{h_t - \mu}{\sqrt{2}\,\sigma}\right)$$

μ can be calculated from the expected gain of a pulse and the measured values of the RP gain ($1.296 \cdot 10^7$ for the R5912) and mean height (24.28 mV). σ is not known [110], but can be estimated from the measured standard deviation of the RP by assuming that the relative standard deviation σ/μ is the same for RPs and the effect:

$$\sigma \approx \mu \cdot \frac{\sigma_{\mathrm{RP}}}{\mu_{\mathrm{RP}}}$$

[110] It could be estimated by numeric calculation of the respective compound Poisson charge distributions (see 2.2.2) and finding parameters for Gaussian functions approximating these distribution well enough, but this involves tremendous effort.

σ/μ of the RP height can be approximated with the same ratio for the RP charge (here $1.083/1.764$).
In addition, the cutoff of the Gaussian PDF at zero has to be considered by redistributing the probabilities lying below zero to the remainder of the distribution. This can be done with the correction factor

$$f_0 = \frac{2}{1 + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\,\sigma}\right)}$$

Finally, for effects with large p, μ should be replaced with the mean height of a single pulse times the expected number of pulses N (see (2.45)), and σ should be substituted with the standard deviation for N pulses (see (2.46)). Since at least one pulse has to be present for observation, N should be set to 1 for expected values $N < 1$. This leads to:

$$\mu \to \max(N, 1) \cdot \mu$$
$$\sigma \to \sqrt{\max(N, 1)} \cdot \sigma$$

Altogether, the detectable probability p_d of an effect for a threshold h_t and a relative probability p amounts to

$$\begin{aligned} p_d &= p \cdot \eta_p = p \cdot \eta_g \cdot f_0 \\ &= p \cdot \frac{1}{1 + \operatorname{erf}\left(\frac{\mu_{\mathrm{RP}}\sqrt{\max(N,1)}}{\sqrt{2}\,\sigma_{\mathrm{RP}}}\right)} \left(1 - \operatorname{erf}\left(\frac{h_t - \max(N,1)\cdot\mu}{\sqrt{2\cdot\max(N,1)}\;\mu\,\frac{\sigma_{\mathrm{RP}}}{\mu_{\mathrm{RP}}}}\right)\right) \end{aligned} \tag{2.157}$$

This leads to the other two probabilities relevant in TST: The likelihood p_s for data acquisition and the probability of detecting all pulses pertaining to an event in data analysis p_f. A time shift effect can produce one or more afterpulses and is defined as detectable only, if all its pulses are recognized.
Both p_s and p_f can be obtained with eq. (2.157) using the respective thresholds h_s or h_f. There is an important distinction, however. For p_s the mean/maximum height of the *largest* pulse an effect produces is used for μ to find out whether data acquisition is triggered. For p_f on the other hand the mean/maximum height of the *smallest* pulse is employed to find out whether all pulses of an effect can be found and the event can be reconstructed completely.
For effects without a charge range from systematic causes, the mean height is calculated with TST and used to estimate the detectable probability. For effects with a charge range due to e.g. a dependence on transit time, the maximum height is used, which gives an upper bound for the expected detectable probability. For greater accuracy, a more appropriate PDF could be used, e.g. a convolution of a boxcar function with a Gaussian (representing statistical broadening through multiplication) for probabilities which are distributed evenly in height.
In the calculation for the R5912, an effective h_s of 1.1 mV (4.5% pe) was reconstructed from data, whereas it was possible to lower h_f to 0.35 mV (1.44% pe) in the evaluation.

Since an effect has to pass both p_s (the largest pulse in the waveform) and p_f (all pulses pertaining to the effect), the detected probability of an effect cannot be

larger than any of both. For this reason, the combined probability p_c is introduced, defined as

$$p_c = \min(p_s, p_f) \tag{2.158}$$

and used whenever data is compared with calculated probabilities, e.g. in the TTD and CTD plots.

Plotting probabilities

In order to be able to plot the calculated detectable probabilities in the TTD such that realistic comparison with data is possible, the smearing of probability over transit time has to be considered.
Effects with point-like distributions in time can be approximated by a Gaussian, for effects spanning a range in time the easiest approach is to assume a flat distribution in time smeared with a Gaussian spread, equivalent to the convolution of a boxcar function with a Gaussian.

Effects can best be compared to data by plotting their maximum probability. For the point-like distribution this corresponds to plotting the probability at the center, which can easily be matched with the observed peak; for effects with time ranges the simple flat function can be directly compared to the actual distribution. Since binned data, i.e. histograms, are used to represent data, this means that to find the maximum probability p_m, the PDF has to be integrated over the width of the central bin located at the predicted time. When the detectable probability p_f is distributed like a point-like Gaussian PDF with standard deviation σ_t and a bin width w is assumed, the Gaussian cumulative distribution function (CDF) can be used and p_m is simply

$$p_m = p_c \left(\mathrm{CDF}\left(\frac{w}{2}\right) - \mathrm{CDF}\left(-\frac{w}{2}\right) \right) = p_c \cdot \mathrm{erf}\left(\frac{w}{\sqrt{8}\,\sigma_t}\right) \tag{2.159}$$

Unless σ_t could be obtained from data, the standard distribution of the RPs (σ_T) was used (see 2.3.3).
For the extended distribution in form of a convolution of a boxcar with a Gaussian, calculating p_m is more complicated: For ranges which are large compared to σ_t, the resulting function reaches a plateau in the middle, but for short spans the central region does not saturate, resulting in complicated integrals. In these cases the integral over the central bin was calculated numerically.

Looking at the TTD in fig. 2.44 and 2.46, even for the low-DCR measurement shown here, it is difficult to observe effects below the $p_m = 10^{-5}$ level, which is why this was used as cut criterion. TST effects were only included in the plots of TTD (fig. 2.44, 2.46) and CTD (fig. 2.45, 2.47) if both $p_s > 1 \cdot 10^{-5}$ and $p_f > 1 \cdot 10^{-5}$, i.e. $p_c > 1 \cdot 10^{-5}$ was fulfilled. The same applies to all TST plots shown for early afterpulses in 2.5.2.2.
In these plots the combined detectable probability p_c is drawn in respectively used for coloring the effect markers. Colored arrows or lines indicate the calculated time and charge range (horizontal and vertical, respectively) or the qualitative charge–time dependence within these ranges (diagonal) of TSEs.

Abbreviations

In the following the abbreviations concerning probability calculations introduced so far and used subsequently are summarized.

Probability types

- p_a Absolute probability of an effect
- $p_r = p$ Relative probability to γk
- $p_{\gamma\mathrm{k},d}$ Detectable probability of RPs, used to obtain p_a from p_r for an effect
- p_1 Probability that a single initial particle which can start an effect is created during a cascade; enters p_r
- p_t Total probability for the creation of one or more initial particles, contributes to p_r
- p_d Detectable relative probability, reduced compared to p by a detection threshold h_t
- $h_t \rightarrow p_d$
  - $h_s \rightarrow p_s$ From instrumental (setup) thresholds; the probability to trigger data acquisition of the waveform. h_s is the lowest instrumental threshold.
  - $h_f \rightarrow p_f$ From data analysis algorithms (fit); the likelihood to detect all pulses constituting the event offline. h_f is the smallest reliably detectable pulse height.
  - μ Expected pulse height
  - μ_{RP} Mean charge or height of RP
  - σ_{RP} RP charge/height standard deviation
- $p_c = \min(p_s, p_f)$ Combined detectable probability; the detected probability can exceed neither p_s nor p_f.
- p_m Maximum binned probability in a histogram, occurring at the center of the time/delay range covered by the effect
  - w Bin width of the central bin
  - σ_t Standard distribution in time/delay of a point in time for an effect, smeared by fluctuations in transit time; set to σ_T (standard deviation of the main peak) where not available from data
- N Expected number of occurrences per event, equal to the product of all p_1 and multiplicities
- $\eta_{\gamma\mathrm{k}}$ Fraction of RPs of all events in the TTD, influenced by instrumental and analysis thresholds

E_γ Photon irradiance incident on the device

γcΛn Fraction of Λ-scattering of a single electron on stage n which creates photons

γcn Probability of photon creation on n through an incident electron, including all factors: $\gamma\mathrm{cn} = \mathrm{bn} \cdot \mathrm{in}\Lambda\mathrm{n} \cdot \gamma\mathrm{c}\Lambda\mathrm{n}$ (see below)

γen Probability of photon emission from n following photon creation

γtij Total transport probability of a photon from i to j, including solid angles and different paths, possibly involving reflections

γtMij Transport probability through reflection off i onto j for an incident photon

Area

A_n Area (cross section) of a stage; n = k, g, f, m, (d)1

A_{Mij} Effective area for reflections off a surface i onto another j, which occurs for a fraction of the incident photons: A_{Mg1}, A_{Mm1}, A_{M12}

FF_f The cross section fraction covered by wires of the focusing grid f (fill factor)

$\mathcal{T}_k$ Transparency of the cathode.

R_n Reflectance of a structure; n = k, g, gg, m, dn

QE_k Quantum efficiency of a stage; n = k, g, f, m, dn

Collection efficiency

CE_{ij} Collection efficiency on j of an electron emitted from i

p_{Tij} Probability of total SE loss from i to j through termination on inactive structures

Multiplicity

δn_x Gain on stage n given conditions x, corresponds to the number of SEs produced there and able to cause effects subsequently

p_{0SEn} Probability of zero SE emission from n

c Creation factor, likelihood to produce 1 or more SE. $c = 1 - p_{0SEn}$. Usually written out instead of using the abbreviation.

sn_h, sn_a Fraction of SEs from dn which are sh/sa.

Backscattering

b Backscattered fraction of incident electrons from a structure

e Elastically scattered fraction of b

i Inelastically scattered fraction of b, $i = 1 - e$

Λ Fraction of i incurring energy reduction through loss processes

B Fraction of i losing energy through branching, $B = 1 - \Lambda$

u Unscattered fraction, $u = 1 - b$

Direction fractions

E(Λ/B) Forward scattered fraction, requires that an active stage is hit (next or original stage). Energy reduction through Λ or B influences the hit fractions and is thus added to the abbreviation.

L(Λ/B) Backward scattered fraction hitting *any* active stage (original or previous stage)

An Arrival factor, fraction of backward scattered electrons (L) landing on the *specific* stage n after the processes described in the subscript

T(Λ/B)(o/i) Fraction terminating on inactive structures; if originating from backward scattering: o = on the way out, i = on the way back in after returning

Enn, Lnn Intradynode scattering can be abbreviated as Enn or Lnn

Ejk, Ljk(o/i) Scattering off j onto k; for L effects: o = on the way out, e.g. from d1 to k; i = on the way back in after reaching the turning point. This notation is used when a different stage than usual is hit.

Impact notation

In order to determine the number of SEs an incident electron will knock out of a surface on average, the impact energy, impact stage, and exact location on the stage are important; the latter due to the edge effect. This depends on previous interactions which influence particle properties such as energy and trajectory. In the equations used to calculate probabilities, charge, and transit times it is therefore necessary to make the properties of an impact particle visible in form of its history. For this purpose the impact notation is used:

- For a variable x (e.g. probability or transit time), previous interactions are listed by appending the impact stage i and a subscript y to x, in the form $x\mathrm{i}_y$.
- The subscript consists, in this order, of the impact particle type and its origin, its emission direction, followed by all its previous interactions.
- Interactions are written as interaction type followed by site and direction (where applicable), separated by colons where necessary. The same abbreviations as in the other notations are used.
  - Λ and B denote the value resulting from minimal particle energy / deposited energy in the loss process; the maximum energy is described based on the lossless interaction.
- Multiple interactions are concatenated, separated by commas where needed. The order of interactions matters and has to be retained. n identical subsequent interactions can be summarized by using an exponent n.
- Standard particles can be omitted. On d1, f, g, m, s this is pk. On d2 to a the standard particle is sn, with n the previous stage. For intradynode scattering

the default particle is the one of the involved stage, elastically scattered on it. For transit times between stages it is the standard particle of the target stage.

- Direction fractions are usually implied implicitly by the stage subsequently hit. Apart from gain reduction due to non-ideal hit positions – which is taken into account through an extra factor – the traversed path is irrelevant for the quantities described by the impact notation, since the energy at impact is the same.
  For forward scattering, one or more additional potential steps can be traversed if a subsequent stage is hit. This is usually clear from context – if e.g. a pk scatters off d1 and then interacts with d2 it has to traverse the potential difference somehow – just as deceleration is obvious for backward scattering, and thus the labels E or L are usually not used.
- Regular amplification as the last interaction can be omitted.
- Examples:
  - The backscattered fraction (b) on d2 (2) for an s1 is $\mathrm{b2_{s1}} = \mathrm{b2}$, since s1 is the standard particle on d2.
  - The gain on d1 (δ1) of a pk (standard particle), which on f with loss processes (Λ) was forward scattered (E, can be omitted) is $\delta\mathrm{1_{pk\Lambda fE}} = \delta\mathrm{1_{\Lambda f}}$.
  - An sh1 emitted in forward direction (E, omissible), which is elastically scattered (ε) on d2 (2) backwards (L, omissible), and hits d1 again (R1, regular amplification) produces a number of SEs there described by $\delta\mathrm{1_{sh1E,\varepsilon 2L,R1}}{=}\delta\mathrm{1_{sh1\varepsilon 2}}$.
  - The forward scattered fraction (E) on d1 of a pk which was previously elastically scattered on d1 (ε1) backwards (L) is $\mathrm{E1_{pk\varepsilon 1L}} = \mathrm{E1_{\varepsilon 1}}$.
  - The same after elastically backward scattering twice on d1 is $\mathrm{E1_{pk\varepsilon 1L,\varepsilon 1L}} = \mathrm{E1_{\varepsilon 1^2}}$.

Impact energy calculation

The impact energy is gained from the initial emission energy U_0 and the subsequently traversed potentials, possibly in several steps (see 2.3.5.4). Energy gain through acceleration is simply added to the starting energy of a particle from the last stage, decelerations are subtracted in the same manner for backward scattering. Energy loss through elastic backscattering losses as well as Λ or B, however, are relative to the total energy at the time of the loss. This complicates the calculations considerably, since the impact energy at d1 is not the same anymore for δ1, $\delta\mathrm{1_{\varepsilon 1}}$ and $\delta\mathrm{1_{\varepsilon 1\varepsilon 2}}$.

Initial energy

U_0 depends on whether the particle is a pn, sn, shn or san, all of which can be produced on various stages. Photoelectrons can be from k, g, m, f, s, and all dynodes. In the calculations it is assumed that the mean pe emission energy from other structures is similar to the cathode, but there could be differences due to other band structure and surface topology. Secondary electrons of the types s, sh, and sa can be produced on g, m, f, s, and (naturally) all dynodes. Here, the assumption is made that the average SE emission energy from aluminum (g) is comparable to $\mathrm{Cs_3Sb}$.

The transit times from dn to d(n+1) are already significantly different for sn and pn due to the differing initial energies.

Loss processes

Λ and B can reduce the energy of a particle. The maximum reduction follows from the requirement to still be able to produce a pulse on the subsequent stage, since otherwise the effect is not observable with the same topology. For interdynode forward scattering, a scattered energy corresponding to that of a regular SE is assumed. For intradynode scattering the required energy was set to the value of δn = 1 which follows from the SEY(E), that is, an energy high enough to on average create 1 SE on impact. This is an arbitrary choice; incident electrons with lower energy have a finite probability of creating 1 SE (or being backscattered themselves; this is indistinguishable), which, however, drops with falling impact energy due to the falling SEY-curve. Demanding $\delta = 1$ is the most natural choice and the calculated values based thereon can serve as guide values. It should be kept in mind, however, that sometimes lower energies will produce pulses, too.
For Λf and Bf to produce 1 SE on d1, the minimal backscattering energy of a pk is *higher* than for intradynode scattering on d1. This is due to the additional potential difference V_{f1} for accelerator-type grids, which has to be overcome first.

Energy reduction through particle production

Auger electrons and high energetic SEs can only be produced near the surface, since these particles would otherwise quickly lose their kinetic energy on the way to the surface and could not be emitted – the sh or sa must be created in one of the first interactions of the primary. Thus as an approximation, the sh / sa energy is subtracted from the impact energy and the quantity of produced SEs is determined from the SEY(E) with the reduced energy.
This simplification was not applied to the emission of single regular SEs (s), which likely are from deeper inside the material judging from their smaller energy. An sn creating a separate pulse (e.g. snLnn) thus is subtracted from the number of SEs available to the main pulse.
Photons can be from various processes; bremsstrahlung, characteristic X-rays and X-ray fluorescence can be created deeper than SEs and still be emitted; for cathodoluminescence the optical properties of the material become important and emission depths are likely smaller. Photons thus need not be produced near the surface, and should not be subtracted from impact energy before SE creation. Whether photon production is energetically possible at all can be estimated by requiring the impact energy to be larger than a cutoff energy set by the QE of the structure hit by the photon; e.g. around 3 eV (413 nm). This is relevant for γg1, γgk and γgf – competing processes for Rg – which can not occur since the impact energy of the pk at g is too low.

Elastic scattering losses

Elastic backscattering is not completely loss free. While the mass difference between an atom and the incident electron is tremendous, nevertheless a momentum and energy transfer takes place. The energy fraction which the scattered particle retains is called kinematic factor, and abbreviated here as εn for elastic scattering on stage n.

For Rutherford scattering assuming an elastic hard-sphere collision [337] the kinematic factor in the most extreme case, for 180° scattering on Cs, amounts to 0.999983, which at the energies relevant here is indistinguishable from 1, so elastic backscattering losses (εn factors) can safely be ignored.
ε1 can be estimated from data by comparing the L1 and γk mean charges. L1 differs from γk by one elastic scattering on d1, and possibly by an reduced δ1 at the d1 re-hit due to edge effect. The challenge lies in the separation of both effects from other close event classes, which introduces uncertainties. For the R5912 (+1425 V, $g = 1.3 \cdot 10^7$) the L1 charge is actually higher than that of γk ($q_{\mathrm{L1}}/q_{\gamma\mathrm{k}} = 1.035$), likely due to contributions of undersized pulses to the used selection of γk events, so the only possible statement is that ε1 is compatible with one within the supposed errors.

Probability factors

All factors influencing the detectable probability revolve around two basic processes: particle creation and transport. The former is relevant to produce the initial particle(s) but also at the last and all intermediary stages in order to create secondaries which continue the pulse or start another. The latter determines the direction fractions and transport losses through various processes. In most effects several stages of transport and creation are involved. In addition, for effects which can be created by more than one particle (e.g. on d2), this has to be considered through exponentiation with the particle number.
The respective probability factors which enter the calculations are listed in the following. The list is similar to that of the abbreviations but in this case ordered by physical processes to provide a better overview.

- **Particle creation**

  Photon impact

  QE

  Electron impact

  SE production

  s

  $\boldsymbol{p_{\mathbf{0SE}}} = \exp(-\delta)$
  Zero SE emissions can in principle occur statistically on all stages for every effect and thus is relevant for all effects, but especially for c:

  $\mathbf{c} = 1 - p_{0\mathrm{SE}} = 1 - \exp(-\delta))$
  One or more SEs must be generated on impact to create a pulse. For effects based on sn, sg, sm or sf also the particle causing the effect must be created first, which results in an additional c term at the start of the equation.

  sa, sh

  $\mathbf{sn_{h/a}}$ For sh/sa creation the sn fraction which are sh/sa is important in addition to c.

Backscattering

b For effects caused by backscattering. This depends on energy and material.

e

i $= 1 - \mathrm{e}$

Λ

B $= 1 - \Lambda$

u At the final stage of an effect no backscattering must occur, since this would lead to independent higher-order effects, which should be covered separately. E.g. for γ1 the effect γ1L2 is a higher-order process, but is initiated by a γP1 as well. Therefore for *all* effects, an u term is added for the first stage of regular multiplication; afterwards the terms are identical to γk and cancel out.

γcΛn The fraction of Λ creating photons on stage n. Photon generation is closely related to inelastic scattering, since Λ creates bremsstrahlung photons, which can carry up to the complete impact energy, and characteristic photons. However, low energetic bremsstrahlung and shell transitions can also be produced in B or R interactions. In addition, photons created in other, low-loss processes such as cathodoluminescence can originate from all dynode interactions.

γcn $= \mathrm{bn} \cdot \mathrm{in} \cdot \Lambda\mathrm{n} \cdot \gamma\mathrm{c}\Lambda\mathrm{n}$ The probability of an incident electron to create a photon on n, considering all factors required for Λ scattering leading to photon creation.

- **Transport**

Photons The probability of photon propagation onto a structure depends on the solid angle it occupies seen from the emission/incidence point, as well as absorption and reflections occurring in between.

γen Photons produced inside a structure have to leave the material without being absorbed first to be able to trigger an effect somewhere else. Emission towards the interior of the material eliminates a large fraction of created photons.

$\mathcal{T}$ The transmission through glass and cathode

R The reflectance is relevant for basically all structures.

A For active stages, which are able to produce photoelectrons, the relative area compared to the cathode enters the probability.

γtij Total probability of transport onto a target stage; this includes the required solid angles, and multiple possible paths such as direct hits or after other processes such as reflections (M).

Electrons

Single particle

CE When photoelectrons or single SEs are involved

E, L, T Fraction of backscattered / emitted electrons hitting a structure; depends on impact energy, stage, and scattering position.

An For L scattering (which requires that the electron hits an active stage) the fraction of electrons hitting a previous stage instead of the scattering stage is important.

All SE from a stage

p_{Tij}

Calculation and estimation of values

The procedure to calculate the effect probabilities observed in the TTD and CTD was derived (p. 210), the components required to carry it out were identified (p. 221), and their abbreviations were introduced together with the impact notation, which summarizes the previous processes of an event (pp. 216 and 218). The final requirement to carry out actual calculations are the numerical values of the probability factors.
As mentioned, the goal is to identify upper estimates for the detectable probability to allow to assess whether an effect can in principle explain the observed features, so for values lying within a range the upper bound is used. Wherever possible, literature values were used, or quantities were calculated based on other variables. Nevertheless, due to lack of available information, many quantities had to be estimated based on the data at hand.
Due to the great effort required for the calculations presented here, these were carried out exemplarily for the R5912 only. Therefore the values given subsequently are for this PMT type, more specifically for the specimen examined in this work, since some variables were extracted from measured data.
The values and assumptions used here are summarized shortly in the following. All calculated probabilities for the R5912 resulting from this are listed in table 2.2 and 2.3, together with their probability formulas.

Quantum efficiency

For the quantum efficiency of the cathode $\mathrm{QE_k}$, the value at the QE peak at 420 nm of 25% was taken [212], but the QE is quite flat between 350 and 430 nm. Of course the mean QE of individual PMTs varies compared to the series average (for the used R5912 follows a value of 25.3% from comparing the cathode luminous sensitivity S in the manufacturer's test sheet to the mean value of S and QE in the catalogue). Also the QE depends on cathode position and since some effects show asymmetric occurrence probabilities, a fixed value can only be an estimate.
For the aluminized glass, $\mathrm{QE_g}$ was set to the upper estimate of 0.1% inferred on p. 56; the value is likely lower.
The focusing grid ($\mathrm{QE_f}$) and the deck ($\mathrm{QE_m}$) are covered with cathode material during evaporation, as is clearly visible from the concentric deposition rings around

the material droplets positioned on wires between deck and cathode. As a worst case assumption which maximizes the probability of (undesired) TSEs resulting out of this, the QE was assumed to have the same value as k. The value could be lower due to suboptimal chemical composition and growth conditions; on the other hand the "cathode" on m and f is reflective, which could increase its QE.
The dynodes of the R5912 consist of Cs_3Sb, which according to Hamamatsu when used as a reflective cathode material has a QE of 14.6% at 400 nm [169; p. 34]. For a dynode the QE likely is not as high, but the value was used as upper bound for d2 and later dynodes. d1 is a high gain dynode with likely better crystallinity and higher QE. The same increase in QE as from Hamamatsu standard QE (SQE) to high quantum efficiency (HQE) PMTs (25%→35%) was assumed, resulting in a value of 20.4%. Even if these estimates are too high, the QE of the dynodes is *far* from zero, so effects based on P on d1+ *will* play a role and have to be studied.
All QE values already contain reflection losses for incident photons, which thus do not need to be included as an extra factor.

Secondary electron production

The probability of zero SE emission p_{0SE} is negligibly small for regular δn values of the first four dynodes of the R5912 (see (2.41)); the accumulated probability is $3.1 \cdot 10^{-7}$ for δ1 and already virtually 0 beyond; the zero SE emission probability for individual stages d1 to d5 is $3.1 \cdot 10^{-7}$%, 0.02%, 0.01%, 0.03%, 0.75%, for d6 to d10 it is 4.71%.
The creation factor c is thus only used when the impact energy is very small (e.g. an sn impact on dn). Already an shn/san impact on the same dynode dn results in a $\delta > 1$. For regular amplification of s(h/a)n on dn+1, the probability of SE generation is assumed to be 1; on later dynodes δn is far lower, but no effects are relevant where c would be of importance. The individual δn values and their calculation are discussed in 2.3.5.4.

Fraction of high-energy secondary electrons and Auger electrons

The sa fraction of SEs was estimated at 1% for all dynodes judging from the small size of the Auger peaks in the SE energy distribution (fig. 2.12b). Taking into account c leads to a total sa creation probability at d1, d2 and d3 of 14.0, 73.0 and 100%, the mean expected number of Auger electrons is 0.150, 1.30 and 12.33. This means that at d2 an Auger electron is present most times but not always. At d3 it is already a regular occurrence and a statistic part of the mean pulse shape. In addition, whether 12 or 13 Auger electrons are created at d3 cannot be resolved reliably due to the small charge difference.

The choice of the sh fraction of SEs is somewhat arbitrary, since there is a tail towards large energy in the SE distribution (fig. 2.12a). An SE was defined to have high energy from transit time considerations and the SE energy distribution. Calculations show that at least 25 eV are required for a notable time difference to the RP, which allows to discern the effect. The RP *includes* SEs with above average though lower energy, since e.g. a fraction of 10% of s1 has an energy of more than 15 eV, so from δ1 on average 1.5 such particles are expected, so s1 with such energies can be considered a statistical part of the RP. Only at 25 eV is the difference in transit time for E scattering marked enough compared to regularly occurring s1 that

sh1E effects are expected to be observable. From fig. 2.12a a value for $s1_h$ of 1.1% was obtained through integration of the values extracted from the plot, which is adopted for the other dynodes. This results in total sh creation probabilities for d1, d2 and d3 of 15.3, 76.4 and 100%, and expected sh numbers of 0.165, 1.43 and 13.57, similar to sa. Thus sh2 will occur most of the time and sh3 will likely be statistically inseparable from the regular pulse shape.

The production of shn and san is a function of impact energy, which is unknown, however. It was approximated by a constant energy dependence, fixed to the factors derived above. The expected number of sh/sa thus is obtained simply by exponentiating $sn_{h/a}$ with the local number of produced SE.

Since the emergence of shn and san assumes that a regular SE production takes place on dn, the first probability factors are identical to γk and disappear.

Backscattered and unscattered fraction

The backscattered fraction b and the unscattered fraction $u = 1 - b$ depend on the material and impact energy (see fig. 2.14).
Two materials are relevant here [111], Cs_3Sb for the dynodes, and Al for the aluminized glass. For the former, values were taken from [170; p. 218] (see fig. 2.14). To give a few numbers, at an energy $E = 0$, 100, 250, 500, 1000 eV the backscatter coefficient b is 4.5, 12.0, 23.1, 30.5, 35.6%. For Al [268; p. 144] was used, where the curve was extrapolated to 0 V by eye, resulting in values at $E = 0$, 100, 300, 500, 700, 1000 eV of 37, 33, 27, 23.0, 20.4 and 19.5%. Note the different behavior of both materials with rising impact energy.
Values at the required impact energies were obtained by linear interpolation between available data points. This results in e.g. b1, b2, b3, b4, bg = 30.7, 18.4, 21.8, 16.2, 37.0%. g was assumed to lie at the same potential as k.
As can be seen from the quoted values, for the dynodes b grows with impact energy (and thus usually with δ), whereas u decreases with energy. For effects with impact energy ranges, the energy boundary maximizing the total probability was taken, since an upper estimate for the detected probability is desired.

Elastic and inelastic backscattering fraction

The elastic fraction e of backscattered electrons (BSEs) for Cs_3Sb was estimated from available SE energy distributions of comparable materials, chiefly [170], to 25% at 100 eV. Due to the scarcity of available data only crude estimates are possible.
The energy dependence of e given by [257, 269, 270] does not match well with data at low energies, as it likely only applies for the higher energies common in scanning electron microscopy (SEM) and Auger electron spectroscopy. Thus e is assumed to be roughly independent of energy for the energies relevant here and set to 25%.
According to [268; pp. 151–152], the total fraction of BSE contained inside the elastic peak is $\propto Z/E$. The graph on p. 152 in the source confirms a linear dependence on Z beyond the value corresponding to Cs_3Sb, so the value of e_{Cs_3Sb} was scaled with the effective atomic number of both materials to obtain $e_{Al} = 6.0\%$, assuming constant energy dependence.

[111] Ignoring Lm, which should be similar to L1.

Branching and loss fraction of inelastic scattering

For inelastic scattering the energy loss can occur either by branching, i.e. that the primary electron is scattered out of the material or alternatively a high energetic SE is ejected, or through loss processes producing other (quasi)particles such as bremsstrahlung photons, plasmons or phonons. In both cases a fraction of the impact energy creates SEs, since at least 1 SE is necessary to observe a secondary pulse for B respective a Λ pulse at all, and an electron with less energy than the elastic peak (down to 0) leaves the material.

The allocation of inelastic BSE (cf. fig. 2.12) into B and Λ is unknown to the author, but it was assumed that it is more likely that the incident primary electron creates at least one secondary electron than merely losing energy through creation of other, unobserved particles. This can be substantiated by studying the possible loss processes, which can fill the inelastic region in the SE energy distribution.
A total energy loss is equivalent to zero SE emission, which was calculated to be negligible for d1. Since the BSE distribution for *small* impact energies is more or less flat and not falling towards smaller energies, this suggests that statistic energy loss merely through frequent phonon scattering is not the main contribution to BSE energy reduction.
Plasmon creation can be estimated to be subdominant from the small size of the plasmon peaks, which in addition are located in the vicinity of the elastic peak and cannot explain the spread down to zero energy.
The emission of Auger electrons on the other hand can be interpreted as a subtype of B, but is small in relative probability as well – the Auger peaks are not large and often only discernible when magnifying the respective region.
This leaves two processes to explain the majority of BSEs and the energy spread from zero to the full impact energy: The emission of a high-energetic photon (bremsstrahlung, characteristic X-ray) or the creation of SEs.
Bremsstrahlung can span the complete energy range, but is less likely to occur for higher photon energies according to (2.8) and thus cannot constitute the majority of events, which must stem from SE creation.

Therefore the fraction B was set to a tentative value of 0.8 and assumed to be independent of energy for lack of knowledge. From $\Lambda = 1 - \mathrm{B}$ follows $\Lambda = 0.2$. The probabilities calculated based on this assumption agree with data (see 2.3.6), but it should be kept in mind that many other estimated factors are involved. The true value will be energy-dependent, which introduces an energy-dependence for pulses created by B and Λ effects, since these always cover an energy range between zero and the total impact energy.

Another conclusion is that a Λ fraction above a few percent cannot be explained by plasmons alone and must be caused in large part by photons or phonons. Due to the low energy loss in electron–phonon scattering (meV range), a large number of such interactions would be required to cause notable energy losses, while at the same time no SEs may be created, which should be very unlikely at energies somewhat removed from the elastic peak. Λ is thus expected to stem mostly from photon creation, which again should be mainly from bremsstrahlung (see 2.6.1). For simplicity, the fraction $\gamma c \Lambda n$ was thus set to 100% (discussed below).

From these considerations, Λ can also be derived for Al: The number of photons generated through bremsstrahlung scales with Z, so Λ_{Al} should be reduced compared to $\Lambda_{\mathrm{Cs_3Sb}}$ by the factor $Z_{\mathrm{Al}}/Z_{\mathrm{eff,\,Cs_3Sb}} = 0.24$, resulting in $\Lambda_{\mathrm{Al}} = 4.8\%$ and $\mathrm{B_{Al}} = 95.2\%$.

Photon creation

Photons are created in a fraction of Λ interactions through bremsstrahlung and characteristic X-rays in the complete energy range between zero and the impact energy. High energetic photons can be created only through these processes.
However, photons can also be generated with less energy from R and B, either from low energetic bremsstrahlung / characteristic X-rays or other processes such as cathodoluminescence. The energy loss then can reduce the amount of SE generated [112].
As a simplification, B and R were assumed to not create any photons, and the fraction of Λ interactions generating photons was set to 100% of Λ to compensate for the missing creation channels. The resulting probabilities match data quite well, as can be seen from fig. 2.54, but photon emission and propagation factors also are included on the probability formulas, which could compensate deficiencies in these assumptions.
Therefore for an incident electron the total probability γc to create a photon equals $\mathrm{b} \cdot \mathrm{i} \cdot \Lambda \cdot \gamma\mathrm{c}\Lambda$. For incident standard particles on d1 to d5 this amounts to 4.6%, 2.8%, 3.3%, 2.4% and 1.6%, on later dynodes and the anode γc is 1.2%, on the grid 4.7%. For the aluminized glass a value of 1.7% would result; however, the pk impact energy of 1.1 eV is too low to create photons, so the value is reduced to 0%.
For m and s, Λ and $\gamma\mathrm{c}\Lambda\mathrm{m/s}$ can be estimated as for Al above. The effective atomic number of stainless steel is $Z_{\mathrm{eff,\,SS}} = 25.5$ [338; pp. 25–26]. Assuming that m and s are at d1 potential results in a total photon emission probability per incident pk of 2.2% for both.

Photon emission

The fraction of generated photons which are actually emitted from the material was estimated at 30% for all structures. 50% of photons are emitted in the wrong direction, heading further inside the material, and not all photons are created close to the surface, which causes losses on the way out. For the wires of f and a the same probability as for the dynodes and m was assumed, because the curvature of the surface is always large compared to the interaction volume [113], which corresponds to the emission solid angle of a flat surface in the first approximation.

Areas

To determine the likelihood of a photon hit on a sensitive surface, knowledge of its relative area to the cathode suffices.

[112] Paradoxically, it can also *increase* the number of SEs produced: If e.g. a bremsstrahlung photon is created near the surface and carries away a moderate fraction of a high energetic primary electron (e.g. after E1 or E1E2), the penetration depth of the primary is reduced, which can allow more SEs to escape the material. This is equivalent to reducing an impact energy lying beyond the SEY peak to a value closer to the peak.

[113] 10 nm–10 µm for SEM electron energies [268; p. 3] – here far lower voltages are used and the dimensions should be even smaller.

The aluminized area covers the complete backside of the bulb opposing the cathode, so the area of g is about as large as the cathode, minus the average cross section of the dynode stack which is visible from the cathode. A_g/A_k was estimated at 0.94.
The relative area of the first dynode which is visible from the cathode was gauged from photos of the PMT from the front, where all internal structures are clearly visible, to amount to $A_1/A_k = 1.2\%$.
The area taken up by the grid was approximated from its fill factor FF_f of the cross-section leading to d1. All photons have to pass f to reach d1; however, there is a fraction of photons which pass f and land on the dynode mount (or d3), so the solid angle supplying photons to f is larger than for d1. Estimating the fill factor from pictures through the number of pixels covered by the grid and taking into account the solid angles leads to $A_f \approx 10\%\, A_1 \rightarrow A_f/A_k \approx 0.12\%$.
The cross section of the dynode mount also was estimated from photos from the front to $A_m/A_k = 7.5\%$

The effective area (including solid angles) of the deck, for which photons incident on m from k are reflected onto d1 was assessed to $A_{Mm1}/A_k = 0.375\%$, for g to $A_{Mg1}/A_g = 1\%$ (compare fig. 2.39d). Photons reflected off d1 are expected to hit d2 with ca. 100% probability from geometry, so $A_{M12} = A_1$.

Transmission

The transparency of the cathode $\mathcal{T}_k = 0.23$ was taken from [200], using the value at 443 nm.

Reflectance

The reflectance R_k of the cathode for rays coming from the inside can be calculated with the Fresnel equations using the refractive indices of vacuum and the cathode (2.7), which results in a large value of 21.1% for vertical incidence.
For Al, the reflectance R_g averaged over the relevant wavelengths was estimated to 90%.
Reflections on the glass gg also follow the Fresnel law. Using the refractive index of 1.5 for borosilicate glass given by Wright in [170; p. 38] results in a vertical reflectance R_{gg} of 8.0%.
For the reflectance of m, the values quoted for steel by [339] were taken as a basis and a value of $R_m = 50\%$ was used. The dynodes were estimated by eye to have a similar reflectance and thus were also put at $R_d = 50\%$.

Photon transport

From the box-and-line schematic [213; p. 38] and the R5912 layout [169; p. 45], combined in e.g. fig. 2.39, the fraction of photons hitting adjacent structures was estimated. Possible targets include active stages (k, g, m, d1 to d10) before and after the emitting one which are visible from some part of the emitting structure, as well as the dynode mount m which is inactive and leads to losses. In fact, the fraction hitting m is the largest term for most emitting stages, since the side walls (not visible in 2D representations) occupy a large solid angle.
Possible internal paths through reflections between stages were included through the reflectance of the reflecting surface and transfer probabilities between stages accounting for solid angles. From late stages many previous dynodes – up to d2 – are visible *directly*, albeit with decreasing solid angle. In addition, the geometry of

the dynodes unfortunately facilitates reflections onto the previous stage. Optical coupling between stages likely is the price which has to be paid for optimal electron optics.
For dn, the solid angle occupied by the dynodes dm with m < n − 2 was estimated through a geometric reduction factor of 50 compared to m+2. This reduction is applied multiplicatively until arriving within the dynodes, for which values were estimated manually. For the rear dynodes dn with n ≥ 5 the transport factor γtn(n−1) to the previous stage was estimated individually based on the schematics, and the transfer probability γt_{dd} between earlier adjacent stages was put at 40%. For γ4k and γ3k, the γtij were set to 70% for ij = 32, 21 and 1k. For γ2k γt1k was estimated at 90%.
The listed values for the reduction factor and the transfer probabilities give rather good agreement with data, although the detectable probabilities for γnk and γn1 are slightly underestimated. However, these calculations involve many parameters, some of which are exponentiated multiple times, so even slight deviations from the true value accumulate to notable differences. The results could likely be adapted to match data by judicious fine-tuning.
Photons which are emitted from a stage m can be reflected back onto m by surrounding structures, which is called self-reflectance here. This yields non-negligible contributions. For d1 (γ11) self-reflectance can occur via d2 or k, for all following dynodes n (γnn) via d(n−1) or d(n+1).

The resulting p_c lie in the order of 0.05 to 1.1% for γnk, 0.3 to 4.1% for γn1, and for self-reflectance amount to $1.6 \cdot 10^{-4}$ for γ11, and $3.1 \cdot 10^{-6}$ for γ22.
To give an example of the calculation, for γ10k (fig. 2.43c):

$$\begin{aligned}\gamma\text{t10k} = {} & \gamma\text{t10,9} \cdot (\text{R}_\text{d} \cdot \gamma\text{t}_\text{dd})^9 + \gamma\text{t10,7} \cdot (\text{R}_\text{d} \cdot \gamma\text{t}_\text{dd})^7 + \gamma\text{t10,5} \cdot (\text{R}_\text{d} \cdot \gamma\text{t}_\text{dd})^5 \\ & + \gamma\text{t10,3} \cdot (\text{R}_\text{d} \cdot \gamma\text{t}_\text{dd})^3 + \gamma\text{t10,2} \cdot (\text{R}_\text{d} \cdot \gamma\text{t}_\text{dd})^2 + \gamma\text{tagk}\end{aligned}$$

where γtagk denotes paths through slits between dynodes which are reflected from the glass onto the cathode (external path in fig. 2.43c). Gap paths were included only in the emission calculations for d10 and the anode, which explains the higher probability calculated for γ10,k/1 and γa,k/1 in fig. 2.54 compared to γ9,k/1. If external paths are included in the other effects as well, this could raise the probability above the data, which allows reduction by other means to achieve a match.

The probability that a photon, which passes the cathode and hits g, m or d1, is reflected on another active stage were estimated as follows: From g with 50% directly onto k, with 45% onto g again which can be reflected onto k in a next step with higher order, onto d1 with 1%. From m with 91% onto k and 5% onto d1. From d1 with 100% onto d2 – the geometry favors this.

Collection efficiency

The term collection efficiency is extended, as now not only the CE between consecutive stages is relevant, but the likelihood to hit any stage which can forward the pulse.

The structures which a pk can hit (with CE values assessed from field layout and geometric consideration) are g (0.5%), f (8.9%), m (10%), s (0.1%), d1 (80%), d2 (0% – disfavored from geometry but possibly slightly above 0%), and d3 (2.5%).

For CE_{k1} the only values which could be found in the literature for the R5912 were crude estimates from the IHEP [114] MCP-PMT group which are very low, fluctuate strongly between publications and are subject to a conflict of interests due to the development of the NNVT [115] MCP-PMT for JUNO [116], and therefore cannot be used as reliable input. In the end a value of 80% was estimated which is typical for all but the newest fast, large PMTs; the true value could be lower or perhaps also a little higher. The pk hit fractions estimated for all other structures given above – derived ab initio – sum up to about 20%, which reinforces this estimate.

The grid is hit by a fraction of pk which normally would hit d1, which corresponds to the grid fill factor FF_f. Since the value of CE_{k1} already accounts for these losses, CE_{kf} can be calculated through

$$CE_{kf} = CE_{k1} \frac{FF_f}{1 - FF_f}$$

The standard CE for later stages $CE_{n,n+1}$ is assumed to be higher than CE_{k1} due to the close proximity and high fields, nevertheless losses can still occur. For example between d1 and d2, SEs which are emitted close to the grid f are at risk of being drawn to f instead of d2. Alternatively the dynode mount or d(n+1) can be hit instead of dn. In the latter case the missed gain step effectively eliminates the pulse from detection. For these reasons CE_{12}, CE_{23} and CE_{34} were estimated by subtracting all other direction fractions Enn, Lnn, Ln, Tn (discussed below) from 100%. This results in $CE_{12} = 92.3\%$, $CE_{23} = 87.7\%$ and $CE_{34} = 87.0\%$.

The first dynode can also be hit by pg, pm and pf. For photoelectrons from the aluminum, a CE reduction by another factor of 0.8 compared to CE_{k1} was expected since the electron optics is not designed and optimized for these paths.
For particles from the deck, a similar CE reduction factor of 0.8 was assumed, since the potential differences are very small in this region: pm are drawn to f and then decelerate towards d1 where they land with very small impact energies, making deflections onto the inside of m likely.
From f no photoelectron can escape for an accelerator-type grid, so $CE_{f1} = 0$ here.

For Rm not a photoelectron but SEs are accelerated from m onto d1. In most cases only a single SE hits d1, as evidenced by the TTD in fig. 2.47. Since then the pulse propagation depends on only a single particle, a collection efficiency term is present in the p calculation. In fig. 2.46 Rm is well discernible, which can be used to determine δm by matching p_c with the observed probability at the lower lateness boundary. From this a value of about 0.5 results, which is in agreement with the absence of 2-SE pulses for Rm in the TTD.

Direction fractions

A direction fraction denotes the part of backscattered or emitted electrons which hit a certain structure (see pp. 218 and 223).
These values are difficult to quantify without field simulations. Even with simulations the values are likely to differ from reality due to e.g. surface properties, charge buildup,

[114] Institute of High Energy Physics
[115] North Night Vision Technology Co., Ltd.
[116] Jiangmen Underground Neutrino Observatory

small variations in dynode geometry from assembly and the like. Therefore these quantities were extracted from data where possible, and where not, the available figures were used as base for estimates.

Only E1 and L1 are separated well enough from other effects with similar or higher probability to allow to obtain their probabilities with sufficient accuracy (see figs. 2.44 and 2.46 and discussions in accompanying text). Individual cuts on PP charge and excluding APs were used to better isolate the effect peaks in the TTD. For L1 a single Gaussian was then fitted to the left flank of the peak (red. $\chi^2 = 0.61$) since other effects add to its right flank, resulting in $p_{c,\mathrm{L1}} = 1.26\%$. For E1 a fit with three Gaussian functions best described data: One for the γk peak, another for E1 and minor underlying effects, and the third to phenomenologically describe a conglomerate of effects in between the other peaks (mainly sh1E1, sa1E1, BE1, BE2, sh2E2, sa2E2 and E2, see fig. 2.44). The γk peak is difficult to fit after the peak ($T > T_0$), likely due to the CTS and γMgk, and the multitude of effects present between γk and E1 in the TTD can only be described approximately with the fit function, so the red. χ^2 amounts to 1.77. This could be improved by excluding large charges in addition to small ones (here only the latter was used). Through this procedure for $p_{c,\mathrm{E1}}$ a value of 2.54% was found. Using

$$p_{\mathrm{E1}} = \frac{\mathrm{b1}}{\mathrm{u1}} \cdot \mathrm{e1} \cdot \mathrm{E1} \cdot \mathrm{u2}_{\mathrm{p}\varepsilon 1}$$

$$\text{and} \quad p_{\mathrm{L1}} = \frac{\mathrm{b1}}{\mathrm{u1}} \cdot \mathrm{e1} \cdot \mathrm{L1} \cdot \mathrm{u1}_{\varepsilon 1}$$

with values of the other variables as described above, the direction fractions could be calculated to E1 = 34.2% and L1 = 16.4%.
In order to not further complicate the calculations, L1 was defined as the direction fraction for a re-hit on d1, that is, f hits are excluded. L1fo denotes the *additional* direction fraction for pk which are backward scattered (L) off d1 (1) and hit f on the way *out* (o) to k, which is relevant for L1EfR1. An f hit on the *return* path from k to d1 – described by L1fi – leads to different effects, e.g. L1Lf.

Values for the other effects were estimated based on E1 and L1, the assumption of Lambertian emission, the PMT layout and the expected field shapes.
E2 was set to 30%, E3 to 40%.
Forward scattering off the grid (Ef) was assumed to only be able when a pk glances off the outer diameter of a wire, so a reduced direction fraction of 5% was estimated. On the other hand forward scattering on f for a pk which was backscattered from d1 ($\mathrm{Ef}_{\varepsilon 1}$) uses the full cross-section of each wire. Since a subsequent impact on d1 is very likely, $\mathrm{Ef}_{\varepsilon 1}$ was set to an increased value of 65%.
Based on the dynode chain geometry (for example fig. 2.42d) and the expected interstage fields (compare fig. 2.20), L2 was estimated to be far more likely than L1 and set to 40%. L3 was put at 30%, L4 at 35%.
Lf1 describes the fraction hitting d1 after backward scattering off f, and with 45% was deemed far more likely than L1, since for the latter the pk has to leave the dynode stack first and can hit its walls. In contrast, Lff denotes the fraction which is scattered backward off f and hits f again and was approximated with the same value as $\mathrm{CE}_{\mathrm{kf}} = 8.9\%$.

L1fo can be derived using L1 (which excludes f hits), FF_f and T1o. The latter is the fraction of pk backward scattered on d1 which terminate on the inside of the dynode mount, and was estimated at 50% of T1 (see below). From

$$L1fo = L1 \cdot \left(\frac{1}{1 - FF_f} - 1 \right) + T1o \cdot FF_f$$

follows a value of 3.5% for L1fo.
$L2_{sh1}$ is put at 40%, since it is expected to be similar to $L2_{p\varepsilon 1}$ despite the smaller energy, because the losses caused by the attracting field of d4 should be comparable. $L3_{sh2}$, however, should be smaller than $L3_{s1\varepsilon 2}$ (which was assumed to resemble L3), since the d4 fields here have a far more notable influence on particles with lower energy, and was set to 25%.
Intradynode direction fractions were estimated based on the typical hit point, solid angles, expected fields, and assuming Lambertian emission. The estimated values are: E11 5%, E22 1%, E33 3%, L11 10%, L22 5%, L33 10%.

Now the direction fractions for SEs which hit their origin dynode shall be addressed.
For regular SEs, $L1_{s1}$ is expected to be very low in view of the field distributions shown in fig. 2.20, which strongly draw s1 away from d1, and was put at 0.5%. Judging by the same field schematics, $L2_{s3}$ and $L3_{s3}$ should be notably higher, since the field lines are more or less parallel to the surface close to it, and were estimated at 2% and 3%.
For $L11_{s1}$, $L22_{s3}$ and $L33_{s3}$ slightly higher values – 1% , 3% and 4% – than for interdynode backward scattering are assumed, since the paths are close to the dynode surface and the field gradients are smaller than for Ln paths.
Intradynode forward scattering should be even stronger suppressed due to the field gradients and the small solid angle for which this is possible: $E11_{s1} = 0.2\%$, $E22_{s3} = 0.3\%$, and $E33_{s3} = 1\%$.
The terms En_{sn} describe the SEs, which are part of a RP, and are equivalent to $CE_{n,n+1}$. In order to determine the $CE_{n,n+1}$ values (see p. 230) the loss fractions Tn_{sn} had to be estimated. $T1_{s1}$ was put at 6% from SEs which are drawn to f or hit the dynode mount walls. $T2_{s2}$ should be somewhat higher to SEs escaping to d4 and later dynodes (effectively lost for cascade generation due to missed gain steps) besides m hits, and thus estimated at 7%. $T3_{s3}$ is likely a bit smaller than $T1_{s1}$, with losses occurring due to m and d5+ hits, and was set to 5%.
For Auger electrons, the fraction which terminates on the dynode mount is expected to be small, since the electron optics should still mostly focus these events onto the next dynode. Due to the larger energy interdynode E-effects become distinguishable from the RP. However, other directions are more likely to occur than for sn now, which reduces En_{san} compared to $En_{sn} = CE_{n,n+1}$. This was accounted in the quantities as follows: $E11_{sa1}$ 1.5%, $E22_{sa2}$ 0.5%, $E33_{sa3}$ 2%; $L1_{sa1}$ 4%, $L2_{sa2}$ 4%, $L3_{sa3}$ 10%; $L11_{sa1}$ 3%, $L22_{sa2}$ 4%, $L33_{sa3}$ 8%; $T1_{sa1}$ 8%, $T2_{sa2}$ 10%, $T3_{sa3}$ 7%. The forward emitted fraction collected on the next dynode (En_{san}) equals the remaining direction probability, which results in $E1_{sa1}$ 83.5%, $E2_{sa2}$ 81.5%, $E3_{sa3}$ 73%.
The energy of shn (25 eV) is higher than sn (5 eV) and lower than san (36 eV), see 2.3.5.4. Their direction fractions thus should lie in between the both and were set to: $E11_{sh1}$ 1%, $E22_{sh2}$ 0.4%, $E33_{sh3}$ 1.7%; $L1_{sh1}$ 3%, $L2_{sh2}$ 3.5%, $L3_{sh3}$ 8%; $L11_{sh1}$ 2.5%, $L22_{sh2}$ 3.7%, $L33_{sh3}$ 7%; $T1_{sh1}$ 7.5%, $T2_{sh2}$ 9.5%, $T3_{sh3}$ 6.5%. The forward emitted fractions then follow to $E1_{sh1}$ 86%, $E2_{sh2}$ 82.9% and $E3_{sh3}$ 76.8%.

For inelastically scattered electrons (B and Λ), the E fractions are expected to be higher than for elastic scattering, since the lower energy facilitates collection on the next stage through the electron optics. L scattering could be either increased or decreased through the smaller mean energy; this depends on the exact field distribution. For now, all EB, EΛ, LB and LΛ values were set to the same numbers as E or L.

Regarding impact of non-standard particles on a dynode, $E1_{s1\varepsilon1}$ should be similar to $E1_{s1}$ (92.3%). $L1_{s1\varepsilon1}$ is comparable to $L1_{s1} = 0.5\%$. $E1_{sh1\varepsilon1}$ likely is similar to $E1_{sh1}$ (86%), which is lower than $E1_{s1\varepsilon1}$ since the higher energy leads to a broader angular distribution.

Other effects on the same dynodes are estimated to have similar direction fractions as comparable effects with similar energies. An exception are $E3_{p\varepsilon1\varepsilon2}$ and $L3_{p\varepsilon1\varepsilon2}$, which were set to 80% of the E3 / L3 value due to the tremendous energy of the pk between d2 and d3, which renders focusing onto d3 through the electron optics ineffective.

The termination fractions Tn after elastic scattering of the standard particle on n can be calculated from the difference between 100% and the sum of all other direction fractions from a stage and effect. Their values were required to be ≥ 0 for obvious reasons. The Tn values were T1 34.4%, T2 24%, T3 17%, Tf 50%.

The total backscattering losses, which are relevant for the determination of the multiplication efficiency (MEn = 1 − Tn), however, also include B and Λ scattering, which have far higher contributions than en. Summing up all weighted terms results in the total backscattering losses $T_t n$, which here are identical to Tn, however, due to the assumptions which were made.
For d1 a ME1 of 65.6% results from E1 34.2%, L1 16.4%, E11 5%, L11 10% (the first two from data), and identical values for B and Λ of the same direction fractions. Moderate errors are expected for the measured values; E11 and L11 have large uncertainties but the order of magnitude should be correct and their contribution is small compared to the sum of E1 and L1. All in all, the tentative calculated ME1 value for the R5912 is in agreement with the general, series-independent value of 75% derived on p. 84. For d2 a value of ME2 = 76% results and for d3 ME3 = 83%.

Arrival factor

For sn scattered on d(n+1) the probability of return to dn can be assumed to be very small. This is due to the small remaining energy at dn, equal to the SE emission energy, which makes focusing away from dn very likely. The arrival factors $An_{sn\varepsilon(n+1)}$ are thus set to 0.

The case is different for EnL(n+1) effects or sh/sa. In the first case it is actually rather likely that dn is hit – due to the large remaining energy the path is not influenced much by the fields – and extremely unlikely that a path is taken which *passes* dn, turns around between d(n−1) and dn, and then is focused onto d(n+1). The arrival factors of a d(n+1) hit $A2_{pk\varepsilon1\varepsilon2}$ and $A3_{s1\varepsilon2\varepsilon3}$ are therefore assumed to be 0, while $A1_{pk\varepsilon1\varepsilon2}$ and $A2_{s1\varepsilon2\varepsilon3}$ are set to 1.
If shn are scattered backward on dn+1, the probability to hit dn again is smaller than for EnL(n+1), since the lower energy makes refocusing on dn+1 more likely.

$A1_{sh1\varepsilon 2}$ was assumed to be 40% ($\rightarrow A2_{sh1\varepsilon 2} = 60\%$). $A2_{sh2\varepsilon 3}$ is more likely (70% $\rightarrow$ $A3_{sh2\varepsilon 3} = 30\%$), because the electrons would have to run up the potential towards d1 to avoid hitting d2 and thus most hit d2.

Total secondary electron loss between stages

The probability that all regularly produced electrons from a stage n get lost on the way to the next equals

$$p_{Tn,n+1} = (1 - CE_{n,n+1})^{\delta n} \tag{2.160}$$

$p_{Tn,n+1}$ is the loss probability between n and n+1 for a *single* particle incident on n. From the cathode only 1 pk is emitted, which is why $p_{Tk1} = 1 - CE_{k1} = 20\%$.
Because δn is large for the first four dynodes (> 8, see (2.41)), $p_{Tn,n+1}$ is negligibly small there, however: $p_{T12} = 2.0 \cdot 10^{-17}$, $p_{T23} = 1.2 \cdot 10^{-8}$, and $p_{T34} = 4.2 \cdot 10^{-9}$. The probabilities for a continued cascade to disappear can be obtained by multiplying the individual loss probabilities and are even smaller.
The total SE-loss probability can therefore be neglected if a regular amount of SEs is produced on a regular position [117].

Photonic effects

For all γij effects regular amplification at i is assumed. For many effects studied here an RP is assumed as basis, but photons can be generated as a side product in all dynode impacts. Some effects do not require production of a second pulse through regular amplification, e.g. on f, since no SEs can escape. In this case the total probability of the effect can be a little higher, since inelastic scattering contributes as well.

Determining the effect probabilities from data

The ultimate goal would be to assess the probabilities of all effects discussed here from data, in order to identify discrepancies of the assumptions and improve the model. This would allow more accurate predictions when TST is used, e.g. in photosensor development.
However, this is extremely difficult due to the multitude of effects, which overlap and for whose probabilities only theoretical predictions and no data exist. The only possibility would be a simultaneous fit of all effects on both PPs and APs in charge, time and probability with a fit model incorporating *all* factors discussed previously with their interdependence. It would be even better to use calculated fit functions $q(T)$ (such as the one derived for BL1, see p. 243) for all interactions instead of the simple linear correlations used for most effects.
Since for many interactions knowledge of the exact geometry and field layout is required for correct calculation, such a fit unfortunately is not possible at the time. The calculated probability and time values are too uncertain to clearly attribute all effects to features in the TTD and CTD in the present state of the model. This was only feasible for a select few effects, which are well separated from features with ambiguous association, e.g. L1, E1, and γ1.
A possible intermediate step to such a complete fit would be to calculate the summed probability function for the TTD and DD from all effects using the current models for

[117] For the edge effect both $CE_{n,n+1}$ and δn could be smaller at irregular hit positions and, therefore, total SE loss might play a non-negligible role.

probability and time calculation. By comparing this to the measured distributions the probability factors and timing calculations could be refined.

As a next step, it would be desirable to be able to reconstruct the CE (meaning CE_{k1}) from measured effect probabilities. This is difficult though, since it requires to determine all other CE_{kx} with x = f, m, g, s, d2, and d3, for which it is necessary to determine or conclude all other probability factors involved. For example, the d3 effects will likely be impossible to measure, as the produced pulses are too small for acquisition: For p3 $p = 2.9\%$ – this is in fact a large effect – and p_f is still 0.45%; however, $p_s = 1.8 \cdot 10^{-13}$, since no other, larger pulse is produced, which prevents detection.

2.3.5.4 Charge and related quantities

In order to calculate the charge of an effect, first the occurring individual amplification steps have to be identified, before the gain at each stage can be calculated from the respective impact energy. In order to achieve this, first the abbreviations which are relevant for the charge calculations are introduced, before the calculation of impact energy is discussed, which directly leads to the stage gain. From this follows the total charge and its properties, and for effects with APs the charge sum and charge fraction. Finally, the pulse height, which is closely related to the charge, is discussed shortly.

Abbreviations

- **Impact energy**
  This complements the terms introduced with the impact notation on p. 218:

  $\boldsymbol{E}$ Energy

  $\boldsymbol{E_0}$ Initial energy of a particle leaving a stage; either emission energy or energy after backscattering

  - $\boldsymbol{E_0}$**xn** Emission energy of particle type x = p, s, sh, sa from stage n
  - $\boldsymbol{E_0}$**n** Backscattered energy of particle in impact notation subscript scattered on n.
  - $\boldsymbol{E_{0,c}}$**n** Crossover energy: The initial energy where the change in transit time causes the PP to become the AP and vice versa

  $\boldsymbol{E_i}$**n** Impact energy on n of particle in impact notation subscript

  $\boldsymbol{E_{i,x}}$**n** Impact energy after emission of particle x (s, sh, sa, γ) near surface

  $\boldsymbol{U}$ Potential, $E = e \cdot U$

  $\boldsymbol{U_{ij}}$**,** $\boldsymbol{U}$**ij** Potential difference between stages i and j

  $\boldsymbol{\varepsilon}$**n** Elastic backscattering loss factor

  $\boldsymbol{\Lambda}$**n** Inelastic backscattering loss factor

- **Gain**

  δn The gain on stage n of particle in subscript

  δ_cn Gain for crossover energy

  δ_xn Gain after emission of particle x (s, sh, sa, γ) near surface

  δnn For BEnn / BLnn, the notation would be ambiguous, so δn_{Xnn} denotes the gain from the energy deposited by the incident particle *before* intradynode scattering (the first pulse), and δnn_{Xnn} the gain of the re-hit on the same dynode *after* intradynode scattering.

  rn Gain reduction factor on n from non-ideal impact location (material and surface properties, edge effect)

  r$_c$n Gain reduction factor at crossover charge, where PP and AP change places

- **Charge**

  q Charge

  Σ_q Charge sum for PP and *one* AP, $\Sigma_q = q_{PP} + q_{AP}$

  η_q Charge fraction of an AP to its charge sum, $\eta_q = \frac{q_{AP}}{q_{PP}+q_{AP}}$

- **Height**

  h Pulse height

Gain

Potential differences

The traversed potential between two stages can be calculated based on the resistance distribution of the voltage divider. Ideally, this also lists the potential of focusing structures like f and m, and of the light shield, if implemented. Otherwise, these values have to be estimated, as is also necessary for g if no field simulations are performed.

Here, a R5912 with front-tapered voltage divider Hamamatsu E7694-01 (values from [315; p. 47] and datasheet provided with VD) at +1425 V was taken as basis for the computations.
The potential of f was inferred to lie between d1 and d2 (g1) as described on p. 203. Th deck m and the light shield s (if present – this could not be verified from layout schematics or visual inspection) were assumed to be at d1 potential (g2), as is often the case [277]. The aluminized glass is expected to have the same potential as the cathode, since the aluminum provides the voltage (HV or ground) to the low-conductivity cathode.

From these considerations the following values were derived:

For regular trajectories: (2.161)

$$U\{\text{k1, 12, 23, 34, 45, 56, 67, 78, 89, 910, 10a}\} = \{514, 182, 227, 151, 75.6, 45.4, 45.4, 45.4, 45.4, 45.4, 45.4\}\ \text{V}$$

For non-standard paths: (2.162)

$$U\{\text{kg, km, ks, kf, k3, g1, 1f}\} = \{0, 514, 514, 542, 923, 514, 27.5\}\ \text{V}$$

Emission energies

For photoelectrons emitted from a K_2CsSb cathode, a mean energy of about 1.1 eV can be extracted from fig. A1.8 in [172; p. 1-32] (fig. 2.4) for 3.06 eV (≈405 nm). This was used for pk, and for lack of alternative values also for photoelectrons from the dynodes (Cs_3Sb), g (Al), m (stainless steel coated with cathode material), f (wires coated with k material) and anode a (wires coated with dynode material).
The mean dynode emission energy is stated to lie in the region of 5 eV by [170; p. 323]. This value was used for regular SEs (s) from the dynodes and all other structures. For high-energetic SEs (sh) U_0 was estimated at 25 eV (see p. 224). In the case of Auger electrons (sa) a tentative value of 36 eV was assumed, based on Auger emission line measurements of atomic Cs, which show prominent lines around this value [271].

Calculation of impact energy

Impact energy is based on emission energy, traversed potentials and energy losses, and calculated using impact notation, which was already introduced on p. 218.

For effects based on Λ and B, the initial energy of the scattered particle has a range, which is why its impact energy and the resulting gain and charge also are not centered on one value but show a spread.
This also applies for effects started by sn, since sn always include a small fraction of shn with substantially different initial energy, which changes gain and transit time. In principle all effects have a gain range through gain statistics, especially from the first multiplication stage (cf. 2.2.2). Here only the mean gain is used except for effect types for which a systematic gain or charge range is listed, since this allows to assess the match with the peak in the CTD.

In a few cases, an EP effect (an effect capable of producing shorter transit times than γk, earliness) can also produce pulses with lateness (longer transit times than γk) in a part of its parameter space. If the effect also produces a second pulse at RP time, the PP and AP change places when the transit time of the PP becomes larger than the one of the AP. This requires to treat the two types of APs (RP-pulse and effect-pulse) separately. Such a behavior is the case for BE11, ΛE33, BE33, E2L33 and s1L11E1. The earliness or lateness of the pulse with varying time depends either on the initial energy of its generating particle (after a loss process) or its trajectory (intradynode scattering). By equating the transit time to the turning point where the pulse-order changes, the initial energy corresponding to the exchange (crossover energy) can be determined. From this the impact energy and then the gain can be calculated.

Secondary emission yield

The SEY(E) curve for Cs_3Sb was extracted from fig. 2-7 in [169; p. 19] (fig. 2.10) using DataThief III [340].
For aluminum the values presented by [341] for Al 99.5% were taken; comparison with the final TST results points to lower values, likely due to poorer material properties such as crystallinity and surface structure.

Stage gain calculation

The SEY at the exact impact energy was obtained by means of linear interpolation between available SEY datapoints.
For several effects, impact energies lower or higher than the data range were required. For higher energies the curve was linearly extrapolated based on the largest available energies, which due to the small slope of the curve in this region is a good approximation.
The matter is more complicated for low energies. Here, an extrapolation from the lowest value to $E = 0$, SEY $= 0$ was used, which underestimates the SEY slightly owing to the concave shape of the SEY curve. In addition, it would be more realistic to assume that at zero energy, no SEs are produced and only the primaries are backscattered with a probability according to b for $E = 0$. Thus as the end point for the linear extrapolation $E = 0$, SEY $=$ b(0) should be taken. This, however, raises the question of how to separate the unscattered fraction u and the backscattered fraction b then (see p. 217 for definitions), since b would be incorporated into the SEY, which is used for SE and photon production following u. No clear solution presents itself, and so the simpler extrapolation to the origin was used.

When the original SEY values as taken from data are used, the total gain g is lower from calculation ($1.424 \cdot 10^6$) than measurement ($1.296 \cdot 10^7$). To adapt the SEY to the PMT at hand, the δn of all ten dynodes was scaled by the same correction factor η_c so that $^{1.424 \cdot 10^6}/_{1.296 \cdot 10^7} = \eta_c^{10}$. Since d1 is a high-gain dynode, subsequently $\delta 1$ was set to the value obtained in the single electron response (SER) fit with the compound Poisson distribution (see p. 100) and the δ of the rest of the dynodes was scaled down using a uniform factor to match the measured total gain again. The final correction factor of $\delta 1$ vs. the original SEY curve was 1.69, for the dynodes 1.21. Instead of from the shape of the SER, $\delta 1$ could alternatively be obtained from the relative mean charge of $\gamma 1$ to γk. Both methods produce comparable results, as can be seen from fig. 2.45 where the value calculated using the SER fit is plotted in for $\gamma 1$, which perfectly matches the peak from data.

The SE production on the deck m was best described with $\delta \mathrm{m} = 0.5$, as outlined on p. 230. Since m is at d1 potential and f is at a potential between d1 and d2, When a pm hits d1, it is first accelerated towards f, passes the grid, and then is decelerated towards d1, which it hits with its initial energy of about 5 eV. On d1 it thus can either be elastically backscattered towards d2 or create SE, but likely at most only 1 SE. To reflect this, $\delta 1_{\mathrm{pm}}$ was set to 1.

For emitted particles from f (pf and ps), $\delta 1_{\mathrm{pf}}$ and δf were set to 0, since the emission energies are too low to escape (1.1 resp. 5 eV compared to $U_{1\mathrm{f}} = 27.5$ eV).

Stage gain values

From the previous deliberations follow the values for δn (confer (2.41)):

On the RP path: (2.163)

$$\delta\{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10\} = \{15.00,\ 8.69,\ 9.46,\ 8.06,\ 4.89,\ 3.06,\ 3.06,\ 3.06,\ 3.06,\ 3.06\}$$

From other interactions: (2.164)

$$\delta\{\mathrm{g},\ \mathrm{m},\ \mathrm{f},\ 1_{\mathrm{pg}},\ 1_{\mathrm{pm}},\ 1_{\mathrm{pf}},\ 3_{\mathrm{pk}}\} = \{0.03,\ 0.5,\ 0,\ 15.00,\ 1,\ 0,\ 9.87\}$$

From this, the mean number of SE present in the cascade after dn for a RP can be calculated by multiplying the δn:

$$g \text{ after } \{\mathrm{d1}\ldots\mathrm{d10}\} = \{15,\ 130,\ 1.23\cdot10^3,\ 9.93\cdot10^3,\ 4.86\cdot10^4,\ 1.49\cdot10^5,\ 4.54\cdot10^5,\ 1.39\cdot10^6,\ 4.24\cdot10^6,\ 1.30\cdot10^7\} \quad (2.165)$$

In addition to reduced dynode gain from too low or too high impact energy, a suboptimal hit position on a dynode can lower δ through edge effect, or reduced SE emission due to material properties (activation, surface). This can occur for effects which involve Ef, Enn or Lnn and electron emission from g. Since the reduction factor rn is mostly used for Enn and Lnn, its standard particle on dn is sn, not s(n−1).
For pg and sg which hit d1 a reduction factors $\mathrm{r1_g} = 0.8$ was estimated.
For Ef, Enn and Lnn effects a maximum gain reduction $\mathrm{r1_f}$, r1, r2, r3 on the respective dynode of 0.5 was assumed.

Where low impact energies lead to a small δ, in the gain calculation a term $\max(1, \delta)$ is introduced, since at least one SE needs to be created for the pulse to continue. A smaller δ than 1 cannot further reduce the pulse charge due to the quantization of charge carriers.

E-type scattering can increase the impact energy greatly compared to the standard particle on a stage. For E1E2 this is 924 eV on d3, for E1E2E3 even 1075 eV on d4. For such extreme energies, the SEY decreases again due to the larger penetration depth and reduced SE escape probabilities (see p. 70). Nevertheless, for d2 and beyond the regular impact energy usually is far enough below the SEY peak that an overall increase is observed through forward scattering. For an E-effect the factor of the δ on the last stage n to the regular δn amounts to: E1 1.22, E2 1.12, E3 1.30, E1E2 1.04, E2E3 1.33, E1E2E3 1.14.

Charge

Relative charge

As for the detectable probability of an effect, it is expedient to calculate the predicted charge of its pulses relative to the RP (γk), with the unit % pe. For the depiction and comparison in the CTD this plays no role – the charge axis can be plotted either in absolute values (pC) or relative to the RP (% pe). However, a relative charge calculation

instead of an absolute one is far simpler and has smaller uncertainties, since the majority of terms in numerator and denominator (confer (2.102)) cancel each other. The absolute charge of a pulse can then be obtained in a second step by multiplication with the measured value of q_{RP}, which here is 1.764 pC.

Relative charge notation and calculation

The charge of a pulse of an effect therefore is noted down in the relative charge notation. In this the product of all gain factors (see above) relevant for pulse production are listed: for the effect pulse in the numerator, for the RP in the denominator. These factors range from the first amplification step to the last multiplication on d10 before readout at the anode. Loss effects and gain modification on a stage through backscattering or altered kinematics enter by substituting the regular δn of a stage by the modified one, using impact notation. In the shortened relative charge formulas (after canceling out all terms), the only remaining factors are for skipped gain steps, modified stage gains through changed impact energies, or reduced stage gains through suboptimal hit positions or angles. It is only necessary to note down the stages with differences to the RP. As an added benefit, this also greatly reduces the length of the notation.
For γk the total charge amounts to

$$q_{\gamma\mathrm{k}} = \delta1 \cdot \delta2 \cdot \delta3 \cdot \delta4 \cdot \delta5 \cdot \delta6 \cdot \delta7 \cdot \delta8 \cdot \delta9 \cdot \delta10 \cdot e \tag{2.166}$$

with e the elementary charge.
Thus, for example for $\gamma1$ the relative charge formula reads

$$\begin{aligned} q_{\gamma1} &= \frac{1 \cdot \delta2_{\mathrm{p1}} \cdot \delta3 \cdot \delta4 \cdot \delta5 \cdot \delta6 \cdot \delta7 \cdot \delta8 \cdot \delta9 \cdot \delta10 \cdot e}{\delta1 \cdot \delta2 \cdot \delta3 \cdot \delta4 \cdot \delta5 \cdot \delta6 \cdot \delta7 \cdot \delta8 \cdot \delta9 \cdot \delta10 \cdot e} \\ &= \frac{1}{\delta1} \cdot \frac{\delta2_{\mathrm{p1}}}{\delta2} = \frac{1}{15.00} \cdot \frac{8.56}{8.69} = 6.6\%\,\mathrm{pe} \end{aligned}$$

Here at d1 a p1 is produced, so $\delta1$ is set to one since no multiplication takes place at d1 and only one particle is transported to d2. The gain at d2 is slightly different for a p1 with U_0 =1.1 eV than for an s1 with 5 eV, which is considered by substituting $\delta2$ (8.69) by $\delta2_{\mathrm{p1}}$ (8.56). All other dynode gains are identical to the terms of the RP in the denominator and cancel out. Therefore the uncertainties for calculating the relative $q_{\gamma1}$ result from the errors of only three variables instead of ten.

Charge range

For effects with loss terms (Λ, B), gain reduction factors rn (from Ef, Enn, Lnn), or caused by sn (which includes higher energetic shn with a certain fraction) some gain factors show a range instead of a well-defined peak, so the resulting total charge also has a range. This is taken into account by calculating the minimum and maximum charge of a pulse.

A Λ pulse (energy of the primary is reduced) can have a gain on the Λ-stage ranging from one (required for observability) to the normal gain from the impact energy.

On d2+ in addition an RP is present, whose charge is reduced by the charge, which would normally be created on later stages by the SE which is subject to the Λ effect, i.e.:

$$q_{\mathrm{RP}} = 1 - \left(\frac{1}{\prod_{i=1}^{\mathrm{n}-1} \delta \mathrm{i}} \right) \tag{2.167}$$

with n the loss stage.

For branching effects (B), it is necessary that both pulses – the B pulse from the scattered primary and the RP created from regular SE creation – are started.
On d1 it is therefore assumed that at least one SE is emitted for the RP (else no RP would be produced and this would be a Λ effect) and for the B pulse, that the primary is emitted with at least the energy of a typical SE – a natural but arbitrary choice; the minimum emission energy of an electron could be less than the mean SE energy. The initial energy of a electron, which is B-scattered on d1 therefore is at least identical to an SE (corresponding to a gain of 1 at the B stage) and at most the total primary energy minus the energy required to produce one SE for the RP. The gain is created at the impact site of the B pulse (for e.g. BE1 on d2). Similarly, the RP-type pulse has a gain on d1 from one up to the normal gain minus the energy required to emit the primary with SE energy.
On d2+, other SEs are present besides the one of the B process, so no production of SEs is required to create the RP-type pulse. However, the B pulse still can produce SEs on the B stage, which merge with the RP and increase its charge. The B pulse energy thus ranges from 1 SE to the total primary impact energy (the gain is produced on the stage which is hit by the backscattered primary; for e.g. BL2 on d2). The RP charge is obtained by *subtracting* the charge normally created by the full impact energy of the SE subject to B (no SEs from the B pulse), and *adding* the gain on the B stage resulting from the impact energy minus the energy carried away by the scattered SE. This scattered energy ranges from the SE emission energy on the scattering stage (just enough to start the B pulse) to the full impact energy (no added gain).

If a gain reduction factor rn is present in addition to the δn, the gain is reduced by a factor of $^1/_{\mathrm{rn}}$ (worst case reduction) to 1 (regular hit with no δ reduction).

For sn effects, there actually is no charge range. Rather, the mean charges for sn and shn are calculated which produces a larger and a smaller charge value. For the accompanying RPs the missing charge has to be subtracted accordingly. sn effects could also be split up in separate sn and shn effects, but this would introduce further datapoints and labels into the plots and was thus avoided.

Quantization

The quantization of the charge carrier – here electrons – plays a crucial role to determine the charge for many effects. This influences either the mean charge, if only one electron is in transit, or the minimum charge, defined by the prerequisite of pulse production which is based on the emission (and amplification) of at least one electron.

This also sets a lower limit for the charge of effects with *regular* or close-to-regular amplification (i.e. producing many SE on d1) – for example for γk, the RPs. There

always is a statistical spread of the number of emitted SEs which extends down to 0 SE with a probability $p_{1,0\mathrm{SE}}$. In these cases, however, no pulse can be observed, which limits the minimum charge of γk to $1/\delta 1$. Of course, the statistic spread of subsequent stages will broaden the charge response of an RP with only 1 s1, but still a gap in which no pulses occur is expected in the CTD below the RP peak. Comparing this expectation with the CTD in fig. 2.38 and 2.45, this is in fact the case. However, h_s sets a detection limit of 4.5% pe which is comparable in size to $1/\delta 1 = 6.7\%$, which can also be the reason for the missing events beneath the RP peak. It is difficult to separate both effects due to the charge spread from the 2 s1-contribution. Therefore the missing events below the RP peak cannot be taken as proof of this postulate.

In general, charge quantization has the consequence that on the charge axis no values can lie in between integer numbers of sn emitted from the first stage of amplification, multiplied by the gain of all subsequent stages. However, multiplication smearing smoothes out the expected peaks that much that observation should only be possible if not only d1 but also the following few stages have a large δ. This could in fact lead to discernible small overlapping peaks in the SER, which are centered on the number of s1 underlying the respective γk pulses.

For the example of $\gamma 1$, the p1 can get lost on its way to d2, which enters the probability of observation as a CE term, since no pulse is produced if the p1 is lost. This is the reason, why this loss fraction is *not* included in q, since the production of a pulse is a prerequisite for a value different from zero – pulse production is assumed implicitly in the q values.
This is different if a number of SEs are produced on a stage, of which a fraction is lost on their way to the next stage. This is included in the charge through a reduction factor quantifying the average losses.

For charge ranges (indicated by brackets in the notation), a single emitted electron from the particular stage was always taken to determine the lower boundary. The maximum value, however, was not quantized if it was between integer numbers i and $i+1$ of emitted sn, so as to allow to estimate the probability for i resp. $i+1$ sn emission.

Examples

E2:	$q_{\mathrm{PP}} = \frac{1}{\delta 1 \delta 2} \cdot \frac{\delta 3_{\mathrm{s1}\varepsilon 2}}{\delta 3}$	$q_{\mathrm{AP}} = 1 - \frac{1}{\delta 1}$
ΛE2:	$q_{\mathrm{PP}} = \frac{1}{\delta 1 \delta 2} \cdot \frac{\delta 3_{\mathrm{s1}\Lambda 2}}{\delta 3} = \frac{1}{\delta 1 \delta 2} \left[1,\ \frac{\delta 3_{\mathrm{s1}\varepsilon 2}}{\delta 3}\right]$	$q_{\mathrm{AP}} = 1 - \frac{1}{\delta 1}$
BE2:	$q_{\mathrm{PP}} = \frac{1}{\delta 1 \delta 2} \cdot \frac{\delta 3_{\mathrm{s1B}2}}{\delta 3} = \frac{1}{\delta 1 \delta 2} \left[1,\ \frac{\delta 3_{\mathrm{s1}\varepsilon 2}}{\delta 3}\right]$	$q_{\mathrm{AP}} = 1 - \frac{1}{\delta 1} + \left[0,\ \frac{\delta_{\mathrm{s}} 2}{\delta 1 \delta 2}\right]$
E22:	$q_{\mathrm{PP}} = \frac{1}{\delta 1} \cdot \frac{\delta 2_{\varepsilon 2}}{\delta 2} \cdot [\mathrm{r}2,\ 1]$	$q_{\mathrm{AP}} = 1 - \frac{1}{\delta 1}$
sh1E1E2:	$q_{\mathrm{PP}} = \frac{1}{\delta 1 \delta 2} \cdot \frac{\delta 3_{\mathrm{sh1}\varepsilon 2}}{\delta 3}$	$q_{\mathrm{AP}} = \frac{\delta_{\mathrm{sh}} 1}{\delta 1}$
Rg:	$q_{\mathrm{PP}} = \max(1, \delta \mathrm{g}) \cdot \frac{\delta 1_{\mathrm{g}}}{\delta 1} \cdot \mathrm{r}1_{\mathrm{g}}$	
L4:	$q_{\mathrm{PP}} = 1 - \frac{1}{\delta 1 \delta 2 \delta 3}$	$q_{\mathrm{AP}} = \frac{1}{\delta 1 \delta 2 \delta 3} \cdot \frac{\delta 4_{\varepsilon 4}}{\delta 4}$

Values

The charge values or ranges for the produced pulses of all effects are listed in table 2.2 and 2.3.
When looking at the CTD (figs. 2.45 and 2.47) and the EAP CDD (fig. 2.56)[118], it becomes clear that effects can be grouped in charge clusters according to the missed amplification stages: none ($q \approx 100\,\%$pe); d1 ($q \approx {}^1\!/_{\delta 1} = 6.7\,\%$pe); d1 and d2 ($q \approx {}^1\!/_{\delta 1 \delta 2} = 0.77\,\%$pe), d1, d2 and d3 ($q \approx {}^1\!/_{\delta 1 \delta 2 \delta 3} = 0.081\,\%$pe), ... This progression is simply the inverse of the values in eq. (2.165), which elucidates the rapid disappearance below the detection threshold: Many effects missing $\delta 1$ and $\delta 2$ can still be detected, however, if the first three gain steps are missed, only the most frequently occurring effects are visible (BL4, s3L33, s3L3, L4, ΛL4). No effect missing four steps is expected to be detectable, if only for the sheer mean N which renders them a part of the average pulse shape with only minor statistical fluctuations.
Effects from particles originating on stage n will miss all previous amplification steps in the pulse they create, although the charge can be somewhat higher due to E-effects or higher emission energy (sa and sh). Thus E1 and L2 belong to the same charge cluster – E1 is a pk which misses $\delta 1$ (but has a slightly higher gain on d2), while L2 is an s1 which hits d2 (with a delay).
Loss effects on the other hand reduce the charge. For Enn and Lnn this decrease is moderate, while for B and Λ it bridges the gap between one charge cluster and the next lowest – the lowest possible charge required for pulse creation corresponds to 1 SE emission which is equivalent to one missed gain step.

Charge function

For effects with charge ranges only the extreme values were calculated. In the CTD plots, the function of charge over time is approximated by a linear interpolation between the extrema. This, however, is only meant to give an idea of the general trend of the charge–time dependence. The exact curve can be obtained by calculating all relevant gain factors depending on impact energy.

This was carried out exemplarily for the very prominent and important effect BL1, which spans most of the EAP CDD, see fig. 2.56. The BL1 AP originates from the pk (primary) which is inelastically backscattered with SE production on d1, and hits d1 again after a loop towards k.
The lateness of 2.6 to 45.3 ns (which is equivalent to the delay here since the PP is an RP) is calculated using the equations from 2.3.5.5. It can be almost as large as for L1 (50.3 ns) [119], and can almost disappear for the minimum backscattered energy corresponding to 5 eV.
For the smallest energies, the kinetic energy of the backscattered pk which returns to d1 actually would have a $\delta 1 < 1$. However, a fraction of these low-energetic electrons is backscattered or manages to knock out 1 SE, which is why $\delta 1 = 1$ – only the probability is decreased, the charge is quantized.
Around 4 ns delay, $\delta 1$ becomes > 1 and the curve basically rises with growing delays according to the SEY(E); larger delays are due to larger backscattered energies.

[118] Charge or delay range, resp. qualitative charge–delay dependence shown by colored arrows or lines for the individual effects; for BL1 the calculated charge–delay curve is plotted.
[119] It is not *quite* as large, since the RP has to be created by emission of at least 1 s1; the lateness is quite sensitive to even small reductions of the backscattered energy due to the contribution of the quadratic field.

Since the charge cannot become larger than the RP [120], the curve finally levels out to the 100% pe level at large delays. Note the long flat region, which stresses the point that already minor energy decreases cause strong reductions in transit time.

Unfortunately, an exact charge–time calculation is quite time-consuming, and therefore could not be carried out for all effects. Due to the concave SEY shape, similar shapes are expected for other events. So the simple linear interpolations (i.e. exponential for semi-logarithmic axes) should be replaced in the mind's eye by concave curves. This corresponds well to the observed distribution shapes in data.

Charge sum

The charge sum is the addition of the PP charge and the charge of *one* AP. It can be useful to show correlations between delay and charge, e.g. to determine if an AP was branched off from a PP by splitting up the primary energy – with reservations, since the SEY is not linear in energy, which complicates such analyses.
If an effect includes one or more APs, the charge sum Σ_q for an afterpulse equals

$$\Sigma_q = q_{\mathrm{PP}} + q_{\mathrm{AP}} \tag{2.168}$$

If either PP or AP have a charge range, the charges of the extreme manifestations of the effect are added and sorted by size to produce the charge sum range.

Charge fraction

This quantity again is defined only for a single AP and defined as

$$\eta_q = \frac{q_{\mathrm{AP}}}{q_{\mathrm{PP}} + q_{\mathrm{AP}}} = \frac{q_{\mathrm{AP}}}{\Sigma_q} \tag{2.169}$$

For a range of values, the charge fraction range is obtained by calculating η_q from related values and sorting the results. The order can be different than for the charge sum.

Similar to the charge sum, the charge fraction is also useful to shows a dependence between charge and delay, as it redistributes events which are evenly spread out in the charge–delay distribution to narrowly defined clusters.
BL1 is a good example for this accentuation. In the CDD in fig. 2.56, the BL1 function cannot be resolved from the various γnk effects in the range between 30 and 45 ns, as they are nearly congruent. However, in the charge fraction–delay distribution (FDD) in fig. 2.62, BL1 becomes clearly visible. The γnk events are randomly smeared around 50% – their charge is independent from the PP and both have a mean charge of 100% pe – whereas BL1 has a well-defined curve, constricting the few events evenly spread out in the CDD to a few bins at large charge fractions for large delays.
The curve shown in the FDD was calculated similarly to the BL1 curve in the CDD, by splitting up the impact energy on both pulses, retrieving the gain from the SEY(E), and calculating η_q according to (2.169). The resulting shape shows some very interesting behavior, caused by the opposed charge curves of PP and AP. The

[120] Unless $\delta 1$ lies beyond the SEY peak.

AP follows the same curve as shown in the CDD. The PP on the other hand has its largest charge for the shortest delays. So the PP has a charge plateau corresponding to $\delta 1$ fixed at 1 at the largest delays (else it would be unobservable), shows a rising charge towards smaller delays and levels out to its maximum for small delays. For short delays the charge fraction is therefore determined by the shape of the AP charge, visible in the CDD, since the PP charge is nearly leveled out. Around 20 ns delay both charges are already almost at peak value due to the strongly concave SEY, which is why a flat plateau-like region is observed here. The charge fraction only can continue to rise notably, when the PP charge decreases for large delays, and at the largest delays reaches a plateau very close but different from 1 due to the remaining small q_{PP}.
A similar behavior in $\eta_q(d)$ is expected for other effects showing a charge anti-correlation.

Height

The height of an effect can be obtained in the first approximation (see p. 129) by assuming a linear relationship between charge and height and multiplying the relative charge with the measured h_{RP} (here 24.28 mV):

$$h \approx \frac{q}{q_{\mathrm{RP}}} \cdot h_{\mathrm{RP}} \tag{2.170}$$

However, for effects with low charges the altered amplification processes likely lead to small changes in the pulse shape. For example for E1, the first gain is smaller ($\delta 2_{\mathrm{p}\varepsilon 1} < \delta 1$) and one gain step less is used. The former will increase the pulse width, the latter will reduce it a bit. Altogether, a slightly broader pulse is expected for E1 than for γk. The same changes apply in varying degree to other effects with changes in δn and the number of amplification steps. Nevertheless, by and large, the pulse shape will be similar to RPs, so the relative height (quoted in pe) will have approximately the same value as the relative charge.

2.3.5.5 Transit time and delay

Abbreviations

$t$ij Interstage transit time from i to j for an electron; variations described by impact notation in subscript. The standard particle is the one of the start stage. For tf1, the default particle is pk, since no sf can escape.

- **$t_x$ij** Transit time after emission of particle x (s, sh, sa, γ) near surface

$t\gamma$ij Interstage transit time from i to j for a photon.

- **$t\gamma$Mij** Idem when photon is reflected off i onto j; includes total path of reflected photon inside device

T Total transit time, from photon hit to pulse output.

- **$T_0 = T_{\gamma\mathrm{k}}$** Total transit time for an RP

$\boldsymbol{T_r = T - T_0}$ Relative total transit time of an effect, defined as difference to T_0. Negative for early pulses (earliness $|T_r|$), positive for late pulses (lateness T_r)

$\boldsymbol{T_{PP}}$ Relative total transit time of the PP of an event

$\boldsymbol{T_{AP}}$ Relative total transit time of an AP of an event

$\boldsymbol{d = T_{AP} - T_{PP}}$ Delay of the afterpulses after its primary pulse

$\boldsymbol{xij}$, $\boldsymbol{x_{ij}}$ Typical distance between stage i and j which a particle has to traverse to reach one from the other; energy dependent. $x$ij can differ from $x$ji

$\boldsymbol{xij_{\downarrow/\uparrow}}$ Minimum/maximum distance between i and j

$\boldsymbol{xnn_{\downarrow/\uparrow}}$ For intradynode scattering, the subscript $\downarrow$ denotes the distance from the typical impact point to the position on n closest to dynode n+1 which is reachable for the particle (an sn has shorter range than an shn or a scattered s(n−1)), $\uparrow$ the range to the furthest point.

$\boldsymbol{v}$ Velocity of a particle

Total transit time

For the total transit time of a pulse produced in an effect it is also sufficient to only consider relative values compared to the RPs. As for gain and probability, the interstage transit times only differ for some stages from the regular case. As soon as the irregular behavior of an effects ends with SE production, the rest of the interstage transit times is normal, since the energy of the SEs is independent of the history of the event [121] and then the SE transit time to the next stage is identical. Only the number of produced SEs varies. So when the total transit time T_0 of the RP is subtracted from the transit time T of an effect pulse, only a few terms remain in the relative transit time T_r. This simplifies calculation, reduces errors and shortens the notation, since only the differences to the RP have to be noted down.
The regular transit time T_0 is (see eq. (2.110))

$$T_0 = T_{\gamma\mathrm{k}} \approx t\mathrm{k}1 + t12 + t23 + t34 + t45 + t56 + t67 + t78 + t89 + t9,10 \tag{2.171}$$

For example for $\gamma 1$ the relative transit time T_r then amounts to

$$\begin{aligned} T_{r,\gamma 1} &= T_{\gamma 1} - T_{\gamma\mathrm{k}} \\ &= (t\gamma\mathrm{k}1 + t12_{\mathrm{p}1} + t23 + t34 + t45 + t56 + t67 + t78 + t89 + t9,10) \\ &\quad - (t\mathrm{k}1 + t12 + t23 + t34 + t45 + t56 + t67 + t78 + t89 + t9,10) \\ &= t\gamma\mathrm{k}1 - t\mathrm{k}1 + t12_{\mathrm{p}1} - t12 \\ &= (0.40 - 25.11 + 5.29 - 4.35)\mathrm{ns} = -23.77\,\mathrm{ns} \end{aligned} \tag{2.172}$$

For reasons of clarity, in the following only T_r formulas shall be used. The relative transit time of a PP resp. an AP are denoted by T_{PP} and T_{AP}.

[121] There should be a slight sn energy dependence on impact energy, which is unknown to the author, however.

Similar to the charge, T can be plotted as absolute values, or relative to T_0 by simply subtracting T_0 from all time axis values.

Delay

The delay d of an afterpulse is defined as the difference in total transit time to its primary pulse

$$d = T_{\mathrm{AP}} - T_{\mathrm{PP}} \tag{2.173}$$

It makes no difference whether absolute or relative values are used for PP and AP transit times in this formula, since an offset of T_0 cancels out anyway.
The vast majority of effects have $T_{\mathrm{AP}} > 0$, though T_{AP} can also be negative if an early pulse creates an AP with very short delay, which still occurs before T_0. The delay, however, can only be positive due to the definition of PP and AP, which labels the first pulse as primary pulse and all following as afterpulses.
Since for many effects the PP or APs are congruent with an RP or other prominent effects such as L1 or E1 – possibly with reduced charge but always at the same T – the delay distributions are far more crowded: In the TTD and CTD pulses which are identical in time to dominant effects (γk and L1) are omitted, since no information gain would result. However, in the DD, CDD and FDD the T differences between an AP and PP, of which one corresponds to an RP or L1, are typically *not* identical and thus split up from degenerate values.

Interstage transit time formulas

In the equations potential differences and effective potentials are used. Energies are converted to effective potentials through $U = E/e$. The initial energy E_0 thus becomes U_0.

Linear field

For a linear potential, the interstage transit time $t_{l,a}$ for electrons emitted with U_0 and *accelerated* in a field while crossing the potential difference U_{ij} was derived on p. 175, resulting in (2.135):

$$t_{l,a} = \sqrt{\frac{2\,m}{e}}\,\frac{x_{ij}}{U_{ij}}\left(\sqrt{U_{ij}+U_0} - \sqrt{U_0}\right) \tag{2.174}$$

For many effects, however, particles are backscattered and traverse the potential not in forward direction (acceleration), but in reverse and are *decelerated*. When their kinetic energy reaches zero (if they do not hit a structure first), they turn around and are accelerated towards the backscattering stage again (which they can hit or miss).
Two cases must be distinguished. If the initial energy of the particle is larger than the potential difference U_{ij} to the previous stage, the full distance and potential difference is crossed and the particle hits the previous stage. If it is smaller, the particle stops short of the previous stage after crossing only a part of the distance and potential, and then returns to the scattering stage. In the latter case it is sufficient to calculate the time to the return point, since the return path is identical in the 1D model used here and requires the same time.

The deceleration formula for $U_0 \geq U_{ij}$ (greater energy) can be derived using

$$U(x) = -U_{ij}\,\frac{x}{x_{ij}}$$

and eq. (2.132), which results in the transit time $t_{l,dg}$:

$$\begin{aligned} t_{l,dg} = \int_0^{x_{ij}} \frac{dx}{v} &= \sqrt{\frac{m}{2e}} \int_0^{x_{ij}} \frac{1}{\sqrt{U_0 - U_{ij}\frac{x}{x_{ij}}}}\,dx \\ &= \sqrt{\frac{2m}{e}}\,\frac{x_{ij}}{U_{ij}}\left(\sqrt{U_0} - \sqrt{U_0 - U_{ij}}\right) \end{aligned} \tag{2.175}$$

This is equivalent to substituting U_{ij} in eq. (2.174) by $-U_{ij}$. As a cross-check, if the backscattered energy is identical to the impact energy $t_{l,dg}$ should equal $t_{l,a}$ which is the case.
$t_{l,dg}$ applies when the initial energy equals the impact energy (e.g. when εn is neglected) or for L-type scattering.

The deceleration equation for $U_0 \leq U_{ij}$ (smaller energy) is obtained by using the same potential as $t_{l,dg}$, which results in the same integrand, but the integration is stopped at a turning point $x_t = -U_0\,\frac{x_{ij}}{-U_{ij}}$ which follows from the intercept theorem. The transit time then reads

$$t_{l,ds} = \sqrt{\frac{m}{2e}} \int_0^{x_t} \frac{1}{\sqrt{U_0 - U_{ij}\frac{x}{x_{ij}}}}\,dx = \sqrt{\frac{2m}{e}}\,x_{ij}\,\frac{\sqrt{U_0}}{U_{ij}} \tag{2.176}$$

For $U_0 = U_{ij}$ the result is identical to $t_{l,dg}$.
$t_{l,ds}$ is used for inelastic backward scattering.

The reverse potential equations apply for backward scattering whenever $i > j$ in $t$ij (e.g. t21) and from f to d1 since $U_\mathrm{f} > U_1$.

Quadratic field

The interstage transit time $t_{q,a}$ of an electron in an accelerating quadratic potential was derived on p. 176 resulting in (2.141):

$$t_{q,a} = \sqrt{\frac{m}{2\,e}}\,\frac{x_{ij}}{\sqrt{U_{ij}}}\,\ln\left(\frac{\sqrt{1+\frac{U_0}{U_{ij}}}+1}{\sqrt{\frac{U_0}{U_{ij}}}}\right) \tag{2.177}$$

As for the linear field, two cases for a decelerating reverse potential have to be discussed. However, a value of $U_0 = U_{ij}$ is equivalent to a particle without emission energy in the forward potential, which is not accelerated away from its start point; the same applies in the reverse potential for a particle coming to rest and never reaching its target, with the transit time approaching infinity. Thus $U_0 = U_{ij}$ is excluded in the calculations.
The potential can be described by

$$U(x) = U_{ij}\left(\left(\frac{x - x_{ij}}{x_{ij}}\right)^2 - 1\right)$$

For $U_0 > U_{ij}$ follows $t_{q,dg}$ using eq. (2.132):

$$t_{q,dg} = \int_0^{x_{ij}} \frac{dx}{v} = \sqrt{\frac{m}{2e}} \int_0^{x_{ij}} \frac{1}{\sqrt{U_0 + U_{ij}\left(\left(\frac{x-x_{ij}}{x_{ij}}\right)^2 - 1\right)}}\, dx$$
$$= \sqrt{\frac{m}{2e}}\, \frac{x_{ij}}{\sqrt{U_{ij}}}\, \ln\left(\frac{\sqrt{\frac{U_0}{U_{ij}} - 1}}{\sqrt{\frac{U_0}{U_{ij}}} - 1}\right) \tag{2.178}$$

The results are identical to $t_{q,a}$ if $U_0 = U_i$.

If $U_0 < U_{ij}$, the integrand is the same as in $t_{q,dg}$, but the integration is stopped at the turning point x_t. This has to be derived by demanding $U(x_t) \stackrel{!}{=} -U_0$, from which follows

$$x_t = x_{ij}\left(1 \pm \sqrt{\frac{U_{ij} - U_0}{U_{ij}}}\right) = x_{ij}\left(1 - \sqrt{\frac{U_{ij} - U_0}{U_{ij}}}\right)$$

If the root would be added instead of subtracted, $x_t > x_{ij}$ would follow, which is unphysical. With this $t_{q,ds}$ can be calculated:

$$t_{q,ds} = \sqrt{\frac{m}{2e}} \int_0^{x_t} \frac{1}{\sqrt{U_0 + U_{ij}\left(\left(\frac{x-x_{ij}}{x_{ij}}\right)^2 - 1\right)}}\, dx = \sqrt{\frac{m}{2e}}\, \frac{x_{ij}}{\sqrt{U_{ij}}}\, \ln\left(\frac{\sqrt{1 - \frac{U_0}{U_{ij}}}}{1 - \sqrt{\frac{U_0}{U_{ij}}}}\right) \tag{2.179}$$

The use cases are identical to the linear reverse potential equations.
Both $t_{q,dg}$ and $t_{q,ds}$ approach infinity when U_0 converges to U_{ij}. On the other hand, when U_0 is decreasing from U_{ij}, $t_{q,dg}$ *quickly* becomes smaller: For tk1 at $0.9\,U_{\mathrm{k1}}$ $t_{q,dg}$ is already only 0.48 as long as for $U_{0,\mathrm{pk}} + U_{\mathrm{k1}}$. When moving in forward direction, most of the transit time accumulates in the region close to the origin, because the quadratic potential accelerates very slowly at the start and the emission energy is small. Therefore, a particle moving in reverse direction which does not completely traverse the potential (due to energy lost in the scattering or converted to lateral momentum) skips the part which is most time-consuming and gets accelerated back far quicker, since the field gradient at its stopping point is much higher than at the origin of the quadratic field.

Constant velocity

For certain circumstances, the electron velocity can be approximated as constant. This is the case for intradynode scattering, where the electron trajectory is close to the dynode surface and the traversed potential equals more or less the dynode potential. It also applies for t_{kg} since $U_{\mathrm{kg}} = 0$ and the path is close to k and g, which are both at cathode potential.
In these cases the electron velocity v equals the initial velocity v_0 which can be obtained from (2.133), resulting in

$$v = \sqrt{\frac{2e}{m}\, U_0}$$

The transit time then is

$$t = \frac{x_{ij}}{v} = \sqrt{\frac{m}{2e}}\,\frac{x_{ij}}{\sqrt{U_0}} \tag{2.180}$$

Photon transit time

The time for a photon to cross a distance x_{ij} is simply

$$t\gamma\mathrm{ij} = \frac{x_{ij}}{c} \tag{2.181}$$

where c is the speed of light.

Calculation

Field shape

For the potential between k and d1, a linear combination with even weighting of the transit times from linear and quadratic potential yielded a better match of t_{k1} with measured values (γ1 and L1) than the classically used quadratic potential [170]. For the k–d1 region thus the combined potential was used for accelerating and decelerating paths:

$$t_a = \frac{t_{l,a} + t_{q,a}}{2} \tag{2.182}$$

$$t_{dg} = \frac{t_{l,dg} + t_{q,dg}}{2} \tag{2.183}$$

$$t_{ds} = \frac{t_{l,ds} + t_{q,ds}}{2} \tag{2.184}$$

The interdynode fields show complicated distributions, which neither resembles a quadratic nor a linear field (fig. 2.20). A linear combination of transit times from linear and quadratic potential as for the k–d1 region also resulted in values closer to reality than from a single potential and was thus used here as well.
This approach could be refined in a next step by either introducing weighting coefficients different from 1 for both k–d1 and interdynode potentials or derive transit time formulas for a combined potential (linear plus quadratic with weighting coefficients). While this might further improve accuracy, analytically calculable potentials can only serve as order of magnitude approximations; only a full 3D electron optics simulation is expected to yield accurate predictions.

Initial energy

Depending on the interaction either the emission energy or the backscattered energy was plugged in as U_0 in the interstage transit time equations. The values for U_0xn and U_0n were already presented in the section on gain and charge.

Distances

As already described on p. 180, the distances between stages were estimated in a first iteration based on the box-and-line schematic in fig. 2.15c, the PMT dimensions from the datasheet [315; p. 35] and a real device (for dk1; the electron multiplier dimensions can not be assessed without “disassembly”).

These values for the distances between the stages in the RP path were adapted using the measured $T_0 = 59.89\,\text{ns}$. The mean value for T_0 from the manufacturer's datasheet is 55 ns, but this number strongly depends on HV and specimen. The interstage transit times were calculated for linear and quadratic potentials using the estimated distances and emission energies, from which the mean values were obtained using (2.182). The resulting calculated tk1 was compared to the value from the measured $T_{\gamma 1}$ by rearranging (2.172) to

$$t\text{k1} = t\gamma\text{k1} + t12_{\text{p1}} - t12 - (T_{\gamma 1} - T_{\gamma\text{k}}) \tag{2.185}$$

and calculating $t\gamma$k1, $t12_{\text{p1}}$ and $t12$, whose values and thus systematic uncertainties are small compared to tk1. From this a correction factor of 1.0082 for tk1 was obtained (24.90 ns from estimate, 25.09 ns from data). Through comparison of the computed T_0 using (2.171) to the measured T_0, a uniform correction factor for the dynode interstage transit times of 0.976 resulted (electron multiplier transit time 35.64 ns estimated, 34.78 ns measured). The resulting $t$ij values are

$$\begin{aligned}\{t\text{k1},\, t12,\, t23,\, \ldots\, t910,\, t10\text{a}\} = \{&25.1,\, 4.3,\, 2.7,\, 3.5,\, 3.4,\\ &4.0,\, 4.0,\, 4.0,\, 4.0,\, 4.0,\, 0.8\}\,\text{ns}\end{aligned} \tag{2.186}$$

The distances in the RP path were then adapted according to the corrected $t$ij values:

$$\begin{aligned}\{x\text{k1},\, x12,\, x23,\, \ldots\, x910,\, x10\text{a}\} = \{&11.9,\, 1.7,\, 1.1,\, 1.3,\, 1.0,\\ &1.0,\, 1.0,\, 1.0,\, 1.0,\, 1.0,\, 0.2\}\,\text{cm}\end{aligned} \tag{2.187}$$

Since the anode a lies in front of d10, $x9\text{a} \approx x910 - x10\text{a}$, which is relevant for photon transit times from the anode and d10 to a previous stage.

Other distances were estimated based on the layout and these tuned values: $x\text{kg} = 4.0\,\text{cm}$, $x\text{kg}_{\uparrow} = 23.8\,\text{cm}$, $x\text{g1} = 11.9\,\text{cm}$, $x\text{ks} = 11.9\,\text{cm}$, $x\text{km} = 9.9\,\text{cm}$, $x\text{f1} = 1.7\,\text{cm}$, $x\text{kf} = 10.2\,\text{cm}$. For the effect p3, the pk trajectories d1–d2 and d2–d3 are shorter: $x12_{\text{pk}} = 0.8\,\text{cm}$, $x23_{\text{pk}} = 0.5\,\text{cm}$.

Distance variation

For some interactions (E/Lnn, Ef, effects involving t_{m1}) not the typical distance between stages could be used for x, but a spread had to be incorporated to account for the distribution of emission angles, which leads to variations in the trajectory length to the next stage *and* changes to the path length from the impact site on the next stage to the stage after that. By analogy with the cathode time spread, this dependence on the dynode (hit) position is labeled dynode time spread.

For forward scattering from f onto d1, xf1 was estimated to range from 0 to 2.5 cm due to the close proximity of f to the end of d1 close to k.
The path length for pm from m to f was assessed based on photos to lie between 0.5 cm ($x\text{mf}_{\downarrow}$) and 3.2 cm ($x\text{mf}_{\uparrow}$).
For intradynode scattering, the extreme paths were analyzed, i.e. when the electron hits its origin dynode on the position furthest and closest to the next dynode. This intradynode distance depends on particle energy – an sn can only cover a shorter distance than an s(n−1) before it hits dn or is drawn away to d(n+1). In addition, the intradynode impact point changes the path length to the next dynode for the SEs emitted from the impact. Intradynode distances $x\text{nn}_{\downarrow/\uparrow}$ to the point closest/furthest

to d(n+1) were estimated to be: For large kinetic energy: d1 0.8/1.9 cm, d2 0.4/1.2, d3 0.6/0.8. For sn: d1 0.3/0.2, d2 0.2/0.6, d3 0.3/0.4. And for san and shn: d1 0.5/0.4, d2 0.4/1.2, d3 0.6/0.8.
The minimum/maximum interdynode distances $x\mathrm{n}(\mathrm{n}+1)_{\downarrow/\uparrow}$ were assumed with uniform values for all energies: $x12$ 0.7/2.0 cm, $x23$ 0.2/1.6 cm, $x34$ 0.6/1.6 cm.

Intradynode scattering transit times

For intradynode drift times, the electron velocity was assumed to be constant since the path is close to the dynode surface. The traversed distance varied between a minimum and maximum value as detailed above, resulting in a transit time range.
For intradynode Λ scattering, the initial energy was set to the value required to produce 1 SE on impact. For d1 this is 11.8 eV, for d2+ 19.9 eV, where the difference originates from the higher SEY for d1.

The subsequent interdynode transit times to the next stage (and in a few cases to the previous, e.g. E1L22L2R1) of the SE produced on impact (resp. the backscattered primary) could be obtained by linearly scaling the regular tn(n+1) with the modified distance, since eqs. (2.174)–(2.179) are linear in x_{ij} and the traversed potential U_{ij} is the same. Of course, this is only valid on the assumption that the potential is stretched uniformly.

Transit time versus distance

In fact *all* transit time equations (2.174)–(2.181) have a linear dependence on the interstage distance x_{ij} – including constant velocity (no field gradient), linear and quadratic fields, and photons – which is noteworthy in itself. The studied fields are general and not limited to PMTs, so this applies to *all* sensors. The linear dependence of transit time on distance appears to be independent of the field shape. This means that the total transit time T_0 scales directly with the dimensions of the device. For large PMTs xk1 increases with cathode diameter, while the distances in the dynode stack remain similar. The linear dependence of T_0 on the diameter in fig. 2.35 as well as the y-axis offset of the fit function at a diameter of zero are a direct result of this – even a small PMT has a dynode stack with finite size.
Only if also the interdynode distances are reduced, such as in metal channel PMTs, fine mesh PMTs or MCP-PMTs, T_0 can be reduced further (see table 2.1).

ΛL1

For ΛL1, the minimum energy is set to the value creating 1 SE on d1 impact, 11.8 eV. This raises the question of how to calculate the transit time, since this energy is smaller than the potential difference to f, and the electron is thus moving solely within the region between f and d1. Furthermore, the potential will be strongly inhomogeneous in this space due to the influence of d1, f and d2. If the transit time is calculated using the reverse potential k–d1 and the potential gradients from f are ignored, a transit time of $t1\mathrm{k}_{\Lambda 1} = 2.0$ ns results. An attractive potential to f leads to $t1\mathrm{f}_{\Lambda 1} = 6.1$ ns, but this would pull the electron to f and not d1.
In the end, $t1\mathrm{k}_{\Lambda}1$ was used for the lower limit of $T_{\Lambda \mathrm{L}1}$. Given that the fields from both d2 and f pull electrons away from d1, it is likely, however, that for such small energies ΛL1 cannot occur and the minimum energy lies in fact higher, namely large enough to enter the f–k space *and* pass f on the return path. This would result in a larger minimum transit time.

Selection of sh1 energy

As already mentioned in 2.3.5.4, the energy of sh1 was estimated by comparing the transit times $t12$ resulting from varying s1 energies while considering the fraction of s1 electrons with higher energies and their expected number per pulse. This was done for s1 electrons with energies from 5 to 25 eV.
For 5 eV the regular $t12 = 4.35$ ns results. 70% of s1 electrons have approximately this energy or more [122] (see fig. 2.12a and p. 224) and from $\delta 1 = 15.0$ thus 10.5 electrons with larger energy are expected per pulse. At 10 eV these values are 3.88 ns, 25% and 3.8; for 15 eV 3.60 ns, 10% and 1.5; for 20 eV 3.39 ns, 3.2% and 0.5, and finally for 25 eV 3.23 ns, 1.1% and 0.165.
s1 electrons with 20 eV therefore are only 0.2 ns earlier than those with 15 eV, which with an expectancy value of 1.5 will be present in most pulses. This makes a separation of both nigh impossible.
Electrons with 25 eV have a lower occurrence probability than for 20 eV, but are better separated from 15 eV (0.4 ns). This is why this value was chosen as U_0s1 and the shn energy of all other dynodes.

Values

The $t$ij in the RP path (see (2.171)) were already listed in eq. (2.186).
The intradynode transit times for E-scattering amount to 0.3 to 2.1 ns depending on impact energy and dynode, for L-scattering 0.5 to 4.7 ns
Intradynode inelastic scattering for the minimum initial energy has transit times from 1.6 to 3.8 ns for E-scattering and 3.1 to 9.3 ns for L-scattering.
Regarding interdynode transit times, $t12$ can vary between 0.8 and 8.7 ns, $t23$ between 0.1 and 4.4 ns, and $t34$ between 0.6 and 5.7 ns.
The transit time from f to d1 after forward scattering on f ranges from < 0.1 to 1.8 ns (see fig. 2.40a).
The range of the difference of $tg1_{\downarrow/\uparrow}$ to tk1) for now was estimated to be about $\pm$ 2 CTS $= \pm 2.0$ ns.
tk3 was 26.0 ns.
Photon transit times to the cathode range from 0.13 to 0.79 ns (both extremes from g), $t\gamma$nk times from the RP path lie between 0.40 (d1) to 0.73 ns (d10). The times to reach d1 are between 0.06 ns (d2 and f) and 0.40 ns (g), 0.33 ns from d10.

Values from data

The extraction of transit time values from unambiguously identified effects was only possible for a select few cases, where the effects are strong and isolated enough: γk, E1, L1, and γ1.

$T_{\gamma k}$ and $T_{\gamma 1}$ were taken to calculate the distance correction factors. Thus the position of the γ1 peak in fig. 2.44 and 2.45 agrees with the calculations by definition. But already the value calculated ab initio for tk1 (24.90 ns) shows remarkable agreement with the final value of 25.11 ns.

For L1, $T_{r,\mathrm{L1}} = 2 \cdot t1k_{\varepsilon 1}$ was measured at 48.81 ns. Since εn is negligible, this should also provide a value of tk1 $\approx t1k_{\varepsilon 1}$ in good approximation. However, with

[122] This suggests that the mean SE energy is higher than 5 eV as assumed here. However, the empirical mean value given by [170] is deemed more reliable than a posterior graphical evaluation of the SE energy schematic from the same source.

24.41 ns the k–d1 transit time which is obtained from L1 is a little too low compared to the value of 25.11 ns from $T_{\gamma 1}$.
Since the field layout assumed for k–d1 produces results adequately representing the observed transit times, this offset could have other reasons. Instead the discrepancy could be due to a slight energy loss of the pk in d1 before or after backscattering, or due to a lateral momentum if the scattered pk path is not congruent with the field gradient. In fact the pk would be expected to hit the cathode for an unreduced backscattered energy and no lateral momentum, since $U_0 = U_{0,\mathrm{pk}} + U_{\mathrm{k1}} > U_{\mathrm{k1}}$.
If an energy loss in d1 (or an energy transfer to lateral movement) is introduced and adapted to match the lateness, a loss/transfer of 2.89 eV results.
The broad L1 peak in the TTD (fig. 2.46) could partially be caused through a spread in energy reduction which translates to transit times: The position of the calculated L1 peak is to the right of the peak in data, which corresponds to higher energies. Around the calculated value, however, almost as many events as at the peak are found. This could result from lossless scattering (besides the tiny $\varepsilon 1$) resp. no momentum transfer combined with elastic backscattering on the cathode.
Alternatively, the transit times calculated with the approximate analytic formulas could be slightly too high.

Fig. 2.44 shows that the calculations for E1 underestimate the earliness $T_{r,\mathrm{E1}} = t12_{\mathrm{p}\varepsilon 1} - t12$, which is computed to −3.19 ns (data −4.20 ns). This mismatch is in all likelihood due to the simplified fields used here and is visible for all early pulse effects, whose earliness is systematically too low by several ten percent.
The E1 peak was fitted in data to obtain its probability, which was used to determine the direction fractions, but its relative transit time was not used to adapt the interdynode transit times. This could be done in a next step to further refine the predictions.

Curve shapes

In the CTD (fig. 2.45 and 2.47), CDD (fig. 2.56) and FDD (fig. 2.61) the qualitative dependence of charge (charge fraction) on time (delay) was made visible through lines connecting the end points of each range.
If no strong correlation is expected, this results in a horizontal bar for effects covering time/delay ranges, a vertical bar for charge/charge fraction ranges, and a cross if both x and y have ranges.
In case of a notable correlation, the behavior was indicated through a rising or falling line connecting point pairs. This only shows the tendency of the correlation, however; the exact curve shapes would have to be calculated as for BL1 or, better still, obtained from field simulations.

Nevertheless, already the general tendency (rising or falling) of similar effects shows interesting behavior patterns and differences.
For interdynode L effects (e.g. ΛL1, s1L1) the charge is rising with time/delay, because a higher initial energy lets the electron traverse a larger distance (refocused, indirect dynode hit) and creates more SEs, whereas it is falling for intradynode L effects (L11, s1L11), since a larger starting energy causes the electron to hit the dynode earlier (direct dynode hit).
On the other hand, for interdynode E effects (ΛE1) the charge falls with time, as a larger initial energy lets the particle traverse the distance faster and creates more SEs on impact.

For intradynode E effects matters are more complicated and no general tendency can be stated, because both the intradynode transit time and the subsequent interdynode transit time to the next stage enter the equation. In some cases their sum results in earliness, in some in lateness, and for several effects it can be both, which requires to split up the effect and treat all subsets separately.

For intradynode effects charge–time correlations occur for the scattered pulse, since the other pulse is of RP-type and has no T variation. For effects on d2+ this translates to partial charge–delay and charge fraction–delay correlations. For lossy scattering (Λ/B) in general only *partial* correlations are observed. These dependencies originate from several effects resulting from variations in emission angle and energy, which in addition can oppose each other.
For all intradynode effects, the gain reduction through the edge-effect plays a role. The emission angle of the backscattered particle affects its trajectory and determines the hit position on the same dynode and the transit time to it. Impact positions close to the dynode edge have possibly reduced SEY from suboptimal material properties, and the produced SEs also have a lower probability to reach the next stage. This correlation between impact position and charge becomes a $q(T)$ dependence through the transit time difference introduced by the altered path compared to RP interdynode trajectories. This difference can be negative (many Enn-based effects) or positive (Lnn-effects *and* some Enn-based types), resulting in q either rising or falling with T.
For Λ/B scattering, in addition the backscattered energy affects the flight time to remote impact positions on the same dynode. Emission energy and angle are assumed to be mostly independent of each other. Also, since the minimal Λ/B energy is still sufficient to reach all parts of the same dynode and should not alter the trajectories much, the hit positions and path lengths will be similar to elastic intradynode scattering. However, a lower scattered energy leads to longer flight times and smaller impact gains. This reduces the earliness / increases the lateness of the pulse, and introduces a partial falling $q(T)$ correlation for the backscattered pulse. This dependence is partial, since the variation in time is not exclusively caused by the energy fluctuation, but also results from the variance in hit position.
For intradynode effects based on sn, partial correlations result from the emission energy variation between sn and shn in a similar fashion.
The $q(T)$ correlations from edge effect and energy variations are opposed for lossy E-scattering, which complicates the dependencies.
For effects on d2+, the discussed $q(T)$ correlations also transfer to the charge fraction; the other RP-type pulse has no charge-time dependence. The tendency of the correlation depends on whether the scattered pulse is the PP or AP, the effect is E- or L-type and whether the scattering is lossless. For branching effects (B) the correlation is additionally boosted by the charge correlation between PP and AP (PP smaller $\rightarrow$ AP larger). To give an example, for ΛE22 there is a partial falling $q(T)$ correlation for the PP, which results in a partial falling $\eta_q(d)$ correlation, since a later PP is smaller and thus η_q is larger for a smaller delay. For BE22, in addition a smaller PP results in a larger AP, which further increases η_q with the same tendency.

Definition of early and late pulsing

Early pulses (EPs) and late pulses (LPs) here are used as phenomenological groups of effects, which shift the transit time to earlier or later values compared to γk. A pulse is early if its $T_r < 0$ (earliness $|T_r|$) and late if $T_r > 0$ (lateness T_r).

Several effects produce pulses with a time range, for which T_r can be *both* < 0 and > 0; this pulse type then is an EP if it has earliness and an LP if it shows lateness. Most EPs are PPs, but for some effects an early PP can produce an AP with such short delay that also the AP is still an early pulse.
This manner of grouping in early and late is chosen, since for the detector in which the PMTs are supposed to be used, the underlying effect causing the time shift is irrelevant, but the sign of the delay plays a crucial role. Especially EPs are critical for event time reconstruction, since in a scintillator-based detector this can be done most reliably from the first hits.
In the literature often all effects occurring before the main peak are dubbed prepulses (e.g. [170; pp. 435–439] [307]) instead of early pulses, whereas in the nomenclature of the Borexino collaboration [275, 336] the term prepulse is used solely for $\gamma 1$. The effects usually called "early pulses" and "late pulses" in the literature here are designated by their physical process: E1 and L1.

When discussing early and late pulses, immediately the question arises how to define both pulse classes.
Take, for example, the early pulses (most notably E1, BE1, sh1E1, and sa1E1). E1, sh1E1 and sa1E1 should be well describable by Gaussian distributions, but BE1 is already more complicated since the probability dependence on T is not known. However, even if it were known, the mean earliness of the effects amounts to only a few nanoseconds, so the distributions overlaps with the main peak, with severity depending on the PMT series and HV. The same applies for other EP effects. It is thus impossible to match pulses to effects event by event. So the population size of the early pulse classes can only be determined statistically. This can be done in several ways:

a) **σ_T-based cut:** Everything $n \cdot \sigma_T$ before T_0. This just gives a number for the sum of *all* EP effects.

b) **Fitting the EP type distributions:** The area under the respective fit functions.

c) **Fitting the main peak:** Everything lying before T_0 after subtracting a main peak fit function. This also summarizes all EP-effects to one number.

Physically speaking, b) is the most natural choice, since early pulses are caused by different processes than regular pulses and events should be sorted by type.
Variant c) assumes the main peak to follow the shape of the (usually Gaussian) fit function, which is not guaranteed (variations from CTS) and therefore should be reserved for PMTs with a clear central Gauss peak to avoid introducing large errors.

One should, however, keep in mind that the ultimate goal of the characterization of the PMT is to use it in some kind of detector, and for this, the sensor properties should be defined with regard to *event reconstruction*. The temporal resolution of the detector is determined by the TTS (i.e. σ_T). Early pulses cause additional errors for the event time reconstruction, whose size depends on the earliness measured in the expected time resolution of the PMT – and this is just how a) is defined.
However, it is not sufficient to simply set a cut at a fixed multiple of σ_T (e.g. $-3\sigma_T$): If e.g. the E1 peak is close enough to the main peak that a substantial fraction of

events lies within this cut, this will lead to non-Gaussian distortions of the main peak, although these events are classified as RPs by the cut. If this occurs, the impact on event reconstruction through the EPs can be assessed by including the TTD in a detector Monte Carlo or by fitting the event classes.

Since both a) and b) have their advantages, values for both definitions will be presented in the measurements with suitably chosen cut position and fit functions. The Borexino collaboration came to a similar conclusion: In the paper describing bulk PMT characterization every pulse earlier than $-3\sigma_T$ before the main peak was designated as EP, and every pulse later than $+3\sigma_T$ after the peak as LP [336]; in the paper analyzing the PMT timing, EP and LP types were classified by fits [275].

The same arguments apply for the late pulses. The late flank of the main peak overlaps with several effects, which have to be separated from regular pulses. Therefore both a σ_T-based limit and cluster analyses will be employed.

2.3.5.6 Directionality

For many effects a strong dependence on the direction of photon incidence is expected. This is well-known for γk, as discussed in 2.1.2, and γ1 [170; pp. 418–419] but is expected also for e.g. γMgk, γMmk, γMg1 and γMm1 (see figs. 2.39c, 2.39d). For the latter two, the influence should be especially pronounced. In addition, the photon hit position on the cathode influences the pk trajectory, which affects Rg and Rm, but also the probability of backscattering on d1 as well as the fraction of backward scattered pk which are collected on d1 again (L1), both due to hit position and angle. For this reason it is actually *not* sufficient to illuminate the sensor homogeneously from the front to completely comprehend its behavior in a detector where it is also illuminated from the side. For a comprehensive understanding it would be necessary to use homogeneous illumination from front and sides (resp. with an intensity profile emulating the surroundings in the planned application), or perform several measurements for varying incidence angles θ and ϕ – i.e. a surface scan (see p. 334).

2.3.5.7 Discussion

Uncertainty of values

Wherever possible, values were taken from literature or calculated based on other values. Nevertheless, due to the unknown exact layout and material properties of the electron multiplier, it was necessary to estimate many variables. While it is expected that too large and too low estimates can balance statistically if several such variables contribute to the calculation of a value, its systematic errors can be large.
All distances and transfer probabilities were based on figs. 2.15c, the only available layout of a box-and-line PMT. However, this is only a schematic and not a real layout, much less one of the correct PMT.

All estimates were performed ab initio, without knowledge of the measured values, with the exception of $p_{c,\mathrm{E1}}$, $p_{c,\mathrm{L1}}$, the p_c values of γnk and γn1 (coarse adjustment), δ1, g, R_σ, tk1, T_0, and σ_T which were used to tune the magnitude of probabilities, dynode gains and transit times.

Furthermore, the exact field configurations were not known and the trajectories were calculated analytically and not simulated. The analytic calculation should yield acceptable results between k and d1 due to the rather close match of field shape and the normalization with tk1. This is observed for the late pulses in the TTD, which mostly stem from interactions in the space between k and d1.
The fields are more complicated between the dynodes, and knowledge of the exact distances is crucial, so a mismatch in transit time by a factor of up to two seems plausible. The calculated dynode transit times seem to be systematically undersized and scaling them by a factor of about 1.3–2 leads to a good agreement with results. For the first three dynodes the interdynode fields are different in geometry from the rest, so individual rescaling factors would need to be introduced.

The uncertainty of results is highest for the probability. For values, whose calculation involves many unknowns, an uncertainty of an order of magnitude seems realistic. On the other hand, for example γ1, γMg1 and L1L1 are predicted very accurately.

The charge uncertainties are lowest, since g and δ1, which are most crucial, could be determined from the measurement. This explains the very good match with observations. The γ1 peak has a perfect match; however, calculations for the mean charge of γ41, γ51 and the population with lateness 5–25 ns filled by γm, BEf, Rm and other effects, are about 15% too high. This might be due to the edge effect, non-ideal d2 hit positions, or late-dynode nonlinearities from a shortly preceding PP. γ1, γMg, γMk and E1 are more reliable estimators for δ1, since here no additional electron interactions are involved.

Considering these points, the consistency with measured values is remarkable (see figs. 2.44–2.47, 2.54, 2.56, 2.57 and 2.60–2.62). Even for the probabilities, the trends and the relations between values fit observations, e.g. for the photonic APs (γnk, γn1 and γn2): The concave behavior with n is reproduced, the probabilities generally match expectations, and the upturn for d10 and the anode might be present, although suppressed. The remaining discrepancies can easily be resolved by slightly tuning the transfer variables, since for e.g. γ10k no less than 24 variables are involved (9 interdynode transfer probabilities, most assumed to be equal due to the identical design of late dynodes; 9 dynode reflectances, using identical values; 6 possible photon hit targets to determine the fraction of each trajectory, 5 of which can lead to a cathode hit).

Generalization to other photosensors

These time shift effects were derived for PMTs, but they are applicable to many photosensors with multiple stages (and even single stages for many effects), such as all devices incorporating dynodes or micro-channel plates (LAPPD, MCP-PMT, micro PMT, metal channel PMT, ...; see 4.3.1). To some extent, some type of late pulsing might also occur in semiconductor-based sensors such as SiPMs, if a charge carrier is backscattered on e.g. an impurity or phonon.

With adequate adaptations, the developed theory should therefore be universally applicable to photosensor timing.

Comparison to data

In the following section 2.3.6, the TSEs of the *primary pulses*, which form the TTD, will be detailed. The delays of the afterpulses, which are created in the *same* effects and constitute the DD are analyzed in 2.5.2. Finally, section 3.2.2 analyzes the dependence on HV and different PMT series as well as the difference between a classic and a pulse fitting analysis for the TTD, CTD, EP and LP probability, DD, CDD, FDD, EAP rate, and individual effects (see p. 197 for a disambiguation of these plot types).

Transit time distribution

If an interaction produces several pulses, this inherently sets a focus on early effects in the TTD since the delayed pulse becomes the AP. Late effects only determine the behavior of the primary pulses if they affect the whole pulse, i.e. when they occur at or before the first stage of amplification; often this is d1.
The TTD, however, is not only populated by PPs: If a PP is too small for the threshold, it will be skipped in detection, in which case a subsequent larger AP can be falsely detected as PP. This is of special relevance for BL1 and γAP, which frequently slip into the TTD at transit times between main peak and L1 if the threshold is too large and thus also contribute to the observed late effects.

Applications of time shift theory

The knowledge of these TSEs in the next step allows to use it for application. Usually the only information employed in event reconstruction in a detector is the charge produced by a photosensor within a time window. If in addition the sensor waveforms are recorded with a flash analog-to-digital converter (FADC), the pulse shape parameters can be obtained. Through a fit-based pulse search then also the presence and properties of APs can be obtained, even if they are present in form of pileup. The relative time T_r – which would greatly simplify analysis – however, typically is *not* available for individual photons, since the photon arrival time is not known for randomly occurring events in the detector.
With this extra information it might be possible to improve the TTS through EP and LP exclusion. EP discrimination is especially important for event reconstruction. Compared to a simple charge cut, the knowledge about afterpulses and pulse shape could allow to increase the photodetection efficiency (PDE) by lowering the threshold while maintaining the same exclusion efficiency. By determining the pulse shape of e.g. E1, γ1 and early γk events with reduced charges, it could be possible to distinguish these events from normal γk events through fits and improve the TTS even without PDE loss.
In addition, the photon counting capabilities could be improved. The pileup recognition gained through pulse fitting allows to better recognize two close γk pulses and improve the separation of 1 and 2 photons.
Furthermore, occurring TST effects could be identified from PP and AP properties, which would allow to reconstruct the time and charge as if they were from an average γk. This would improve the sensor weighting used in event reconstruction, but this will only work reliably for single incident photons.

2.3.6 Time shift effects

In the following, the time shift effects (short TSEs, effects or interactions) predicted by TST (see 2.3.5) are calculated and compared with measurements for a Hamamatsu R5912 at +1425 V with a Hamamatsu E7694-01 voltage divider.
In order to test whether the theory is stable and complies with data, the calculations were based on as few input parameters extracted from data as possible. Only a couple of general values from the SER (total gain g, gain on first dynode $\delta 1$, charge resolution R_σ) and TTD (main peak transit time T_0, transit time spread σ_T) and several values for important effects (transit time of $\gamma 1$, probability of E1, L1 and delay of some γnk and γn1 effects) were used to calculate a total of 185 effects. Every other value required for the calculations was taken from literature, calculated from other parameters, or estimated. All relevant values were fixed *ab initio* – i.e. *before* comparison with data – and no fine-tuning was performed, with the exception of a coarse adjustment of two parameters for photon transport between stages, which affects only γAP effects and has no influence on any plot shown in the present section.

Here the time distributions for PPs (TTD and CTD) and the effects causing EPs and LPs for PPs will be discussed. However, since these interactions are identical or similar to the ones causing EAPs, the latter are also already introduced and analyzed. The EAP distributions themselves, which consist of these effects, will be discussed in 2.5.2.2.
The present section first addresses γk (the standard pulse generation process) and other RPs in 2.3.6.1, followed by the effects forming the main peak in the TTD in 2.3.6.2, before covering interactions which can create EPs (mostly for PPs, since APs can only be EPs, if the PP is also an EP and the delay is very short) in 2.3.6.3. The section concludes with the effects causing LPs for either the PP or an AP in 2.3.6.4; for many important EAP effect groups the PP is an RP ($T_0 = 0$), in which case any occurring AP automatically has lateness.

2.3.6.1 Regular pulses (γk)

A regular pulse (RP) is simply one which undergoes normal amplification without any special effects (fig. 2.39a). Without additional interactions these pulses are described by $\gamma\text{Pk} \rightarrow \text{pR1}$, written in the short form as γRk or γk.
Regular pulses leave the PMT output after the "standard" transit time T_0 (the peak in the TTD), which is exactly what was calculated in 2.3.4.
γk pulses also have normal charge, height, rise and fall time, pulse width, and so on. In short, γk is the standard event against whom all other event types are measured, and its mean properties are used as references: 1 pe for charge or height, and T_0 for the transit time, where T_0 is often set to zero to allow to easier define the earliness and lateness of event types.

This in no way means that γk pulses are not subject to fluctuations.

For their jitter this was extensively discussed in the TTS section 2.3.3 (this passage dealt *solely* with effects causing γk jitter) and calculated in 2.3.4 for the

various contributions. As a result of the transit time jitter of γk alone, some pulses have negative or positive T_r. Especially the CTS can cause large systematic time shifts of several nanoseconds depending on the position on the cathode.

For the γk charge – the spe charge – the sources of gain fluctuations were detailed on p. 96 and in 2.2.2.

Due to the quantized number of SEs leaving d1, the γk charge should not extend continuously all the way down to zero, but show a stop at $^1/_{\delta 1} = 6.7\%$, which is statistically smeared from multiplication noise of the later dynodes. The observability of this effect is examined in the following.
The spe charge distribution of the R5912 was analyzed in fig. 2.22c on the assumption of Poissonian amplification, which produced excellent agreement with data. From the fit shown there, the mean value of the compound-Poisson distribution should lie slightly above its peak due to the positive skew of the distribution. In a real SER, however, the mean invariably lies *below* the peak due to the strong contribution from underamplified pulses of other effects. As a first-order approximation, the peak can be taken as mean value of the true γk distribution. In order to obtain a reliable mean γk-charge, the pulse-shape analysis of PEST [123] (see p. 481) was used to exclude waveforms with APs or unsatisfactory pulse subtraction (residual height and charge after subtraction not within noise band). This results in a clean PP distribution with a reduced contribution from underamplified pulses. The fitted peak charge was 2.403 pC.
According to fig. 2.22c the γk events should drop below the detection threshold of one event around 8.3% pe. This value is actually *above* $\frac{1}{\delta 1}$, which is a theoretical threshold; production of only one s1 is too improbable to be observable. If the TTS and the binning are considered, the detection threshold lies even higher at around 16% pe. Since the distribution drops below 1% of the peak value at 27.5%, it can be concluded that everything filling the CTD in fig. 2.45 at $T_r = 0$ at charges below about 30% pe is predominantly *not* from γk pulses. This is in fact a far larger gap between γk pulses and zero charge than originally estimated from the production of one s1 (6.7% pe).

Nevertheless, in the CTD many pulses are observed around T_0 with charges below 30%, which raises the questions of their origins. Branching effects can lower the charge of (non-γk) RPs down to the $\frac{1}{\delta 1}$ level, while leaving $T_r = 0$ unaffected. Charges below 30% pe can only be caused by B-effects affecting the pk, that is, BE1 (p_r =7.4%), BE11 (1.3%), BL1 (4.2%), and BL11 (2.5%). These effects have large p_r values which in the measurement in total account for about 15% p_c (for PPs only BL1 and BL11 contribute (in sum 6.6%), since the undersized RPs of BE1 and BE11 are afterpulses), and provides an explanation for the large number of RPs found at reduced charges.

The large majority of PPs from early pulse effects and some from LP effects, however, has lower charges than γk can reach. This allows to separate them this way to an extent – at the cost of losing some reduced-charge RPs. Coincidentally, the value of 27.5% pe lies close to the valley in the SER and near the value of 25% pe, which are both often recommended as instrumental thresholds. Values in

[123] Photosensor Evaluation Software Toolkit

this magnitude allow to improve the time resolution by effectively eliminating most TSEs with acceptable RP losses, but at the cost of PDE, which would be gained through inclusions of the underamplified pulses.

2.3.6.2 Main peak

The main peak (see figs. 2.36, 2.44 and 2.46) for the most part is made up of regular pulses, whose transit time variations determine its shape. Generally, the peak has Gaussian shape close to the center, but the non-Gaussian systematic contributions to jitter often quickly lead to deviations. These variations mostly arise from the fluctuations of the long k–d1 transit time (tk1): CTS, emission energy and spot size. In addition, there are several time shift effects overlapping with the RPs, which thus contribute to the shape of the main peak.

The CTS (see p. 179) stems from inhomogeneities in the path length or field from k to d1 over the cathode surface. Since for no PMT the cathode is a perfect hemisphere and the fields are not concentrically focused on d1, this invariably introduces systematic broadening. Often early and late spots are observed close to the cathode brink (e.g. [190]), which contributes to the events at regular charges shortly before and after γk (see figs. 2.45 and 2.47).
The pk emission energy resembles a Gaussian with negative skew (fig. 2.4). The stronger than linear increase of tk1 with falling U_0 (see eqs. (2.135) and (2.141)) further tilts the transit time distribution, which results from emission energy, towards longer times. This might be another reason for the strong right flank of the main peak.
The spot size on d1 (see p. 179) has an effect similar to intradynode forward or backward scattering, since the pk hits a position on d1 which is closer to d2 or further away from it. The transit time difference between d1 hit positions for the fast pk is far smaller than the variance in $t12$ this introduces for the newly emitted s1, which should dominate time shifts from spot size. Since $t12_{\uparrow} - t12$ (4.3 ns) $> t12 - t12_{\downarrow}$ (1.4 ns), here also a skewed distortion in time is expected, which is inclined towards lateness. This further contributes to the late flank.

Significant TSEs which contribute to the main peak for EP effects come from: a) interdynode forward scattering or emission from d1 or d2 (E1, BE1, ΛE1, sh1E1, sa1E1, E2, BE2, ΛE2, sh2E2, sa2E2), b) intradynode forward scattering or emission from d1 to d3 (E11, BE11, ΛE11, s2E22, sa2E22, E33, BE33, ΛE33), c) photoeffect on the deck or aluminized glass (γm, γg), and d) possibly the Shockley-Ramo pulse (see p. 171). Late pulse effects with significant contributions to the main peak comprise single and multiple reflections such as γMgk and γMmk, and intradynode backward scattering (L11, ΛL11). This will be discussed in 2.3.6.4.
The systematic effects on pk transit time for γk (CTS, U_0 and spot size) are expected to yield the largest contribution to the main peak distortion, while the largest influence from time shift effects is from sh1E1, γMgk and sa1E1.
[336] already recognized that distortions of the right flank of the main peak can occur due to photon scattering inside the PMT with lateness in the order of 1 ns. [301] attributed this behavior to CTS from field inhomogeneities and d1 orientation. [275] and [301] phenomenologically described the falling flank with a convolution of an exponential decay with a Gaussian.

Despite all these effects, often a large region of the main peak can be well described with a single Gaussian curve (e.g. [275] for the ET 9351) which goes to show the tremendous efforts the manufacturers have put into optimizing timing. For perfectly matched path lengths and fields the CTS disappears, and transit time compensation as well as adapted d1 shape can battle the jitter from spot size.
Therefore normally a single Gaussian function is employed as fit functions of the main peak. This is often combined with Gaussian distributions for other close-by effects to allow to select a broader fit region not only limited to a few bins close to the peak.

2.3.6.3 Early pulses

In the following section, all studied TSEs are discussed, which are able to produce early pulses ($T_r < 0$) in the TTD, that is, early *primary* pulses (EPP).
Only effects which have a $p_c \geq 5 \cdot 10^{-5}$, i.e. which can possibly be observed, and a few noteworthy exceptions are included in the study. Effects for which the first amplification occurs at d4 or later were not covered since their charges are too low for observation.

The findings are summarized in a plot of the measured PP-TTD (fig. 2.44) in which the calculated p_m and T of the effects are drawn in, a plot of the measured PP-CTD where the predicted q and T are marked (fig. 2.45), and a table of all analyzed EP effects which lists the topology, probability equations, charge and transit time for *both* PPs and APs as well as the resulting values (table 2.2). To give an overview, first the plots and the table will be explained, before the individual effects are analyzed closer.
Since EP effects also can produce APs, not all resulting pulses are visible in the EPP-TTD and EPP-CTD. In addition, the AP plots should be consulted during lecture of the table and the text: the DD in fig. 2.54, the afterpulse charge–time distribution (AP-CTD) [124] in fig. 2.60, the CDD in fig. 2.56 and 2.57, and the charge fraction–delay distribution (FDD) in fig. 2.62 and 2.61.

Transit time distribution

When referring to the TTD, generally the transit time distribution of the primary pulses is meant. Although the TTD could also be plotted for APs, here the DD is more helpful (see 2.5.2.2).
In the following, only the early range of the TTD (early primary pulse transit time distribution (EPP-TTD)) is discussed. The range after the RP peak is examined in the section on the late primary pulse transit time distribution (LPP-TTD) 2.3.6.4, including the EP effects extending into this range. The EPP-TTD is shown in fig. 2.44. Effects which can produce both EP and LP are plotted here, since earliness is more critical for detector operation.

[124] The AP-CTD shows APs with both earliness and lateness. For APs the terms "early" (EAPs) and "late" (LAPs) are defined differently, namely regarding *delay* after the PP compared to the iAPs, and not the time relative to γk.

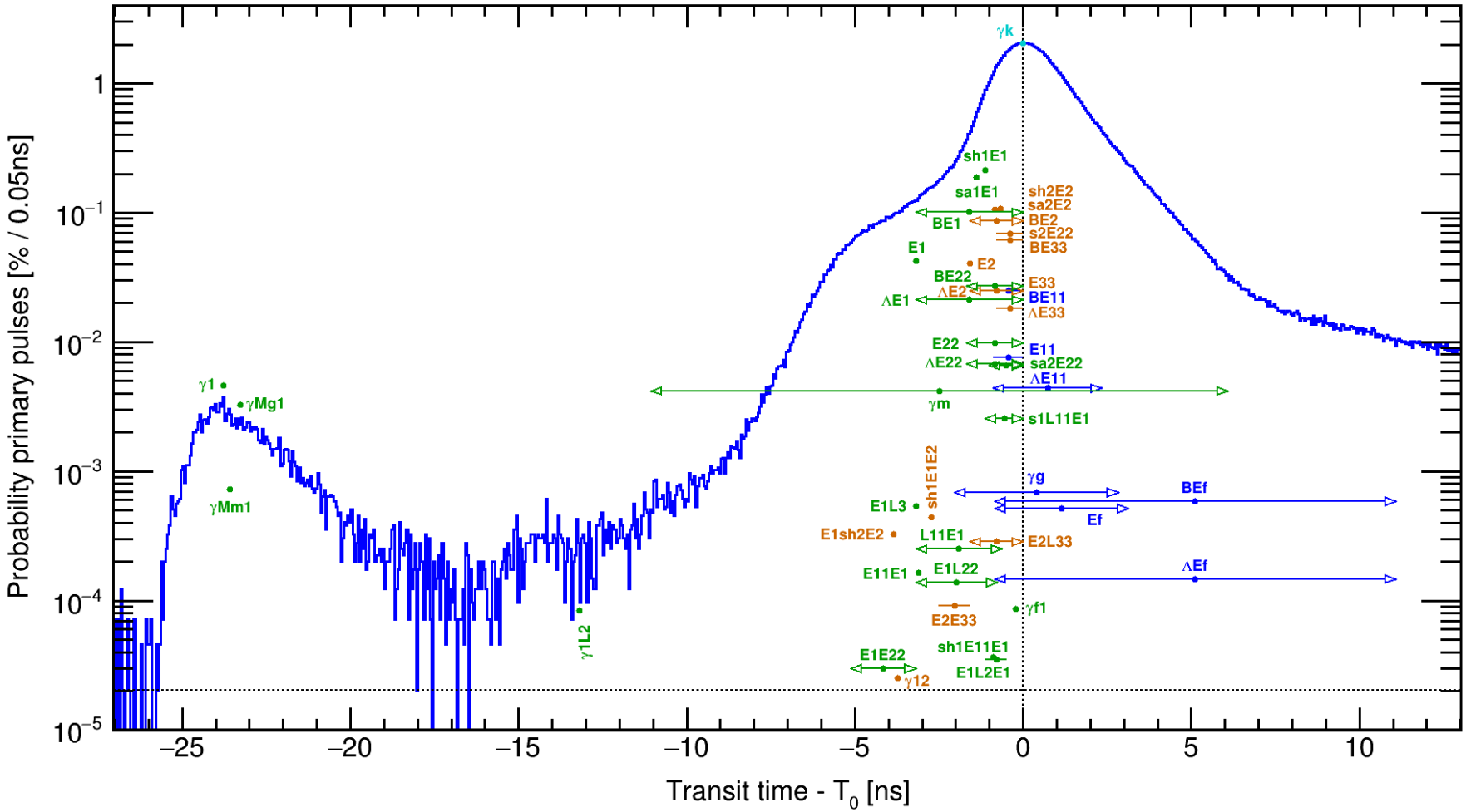


Figure 2.44: Transit time distribution (TTD) of early primary pulses (EPP-TTD) from measurement and calculated with time shift theory (TST). The distribution shows the detectable probability p_c over the relative transit time $T_r = T - T_0$, where T_0 is the mean transit time of γk, which denotes regular pulse production without additional interactions. For late pulses (LPs) see fig. 2.46 (LPP-TTD); effects with both earliness and lateness are shown in the present plot.
Data for a Hamamatsu R5912 at gain $g = 1.3 \cdot 10^7$, occupancy $o = 2.9\%$, measured with SPAX (see p. 460) using pulse fitting with instrumental threshold $h_s = 4.5\%$ pe and pulse fit threshold $h_f = 1.44\%$ pe for $4.9 \cdot 10^6$ waveforms, which results in $T_0 = 59.9$ ns and a transit time spread (TTS) $T_{\mathrm{FWHM}} = 2.26$ ns. $T_r = 0$ and detection limit from 1 bin entry indicated by dotted lines.
Calculated TST values drawn in for binned peak probability p_m (dots) resp. assuming a flat probability distribution (arrows/lines), both smeared with measured σ_T (standard deviation of main peak) where no σ could be obtained from fits. Effect colors reflect primary pulse (PP) charge: blue/cyan $> {}^2/_{\delta 1} = 13.3\%$, green $> {}^2/_{\delta 1 \delta 2} = 1.5\%$, orange $> {}^2/_{\delta 1 \delta 2 \delta 3} = 0.16\%$. γk value matched to peak, all other probabilities are relative to γk.
The calculated earliness for electronic effects (e.g. E1) is systematically underestimated by the analytic computations. The probability values have large error bars due to many unknown factors, which had to be estimated. Considering this, the TTD can be reproduced remarkably well. Discussion see text.
Also compare the TTD in fig. 2.36 which was measured for the same PMT at identical HV but with a different setup (LNGS) and without pulse fitting (classic measurement); fig. 3.13d depicts the TTD for a classic analysis of the same measurement as shown here. In fig. 2.45 the charge-resolved EPP-TTD (EPP-CTD) is shown for the same data.

Calculated probability values

The calculated p_c are drawn at their binned peak values p_m, as described on p. 215. For effects with time ranges for simplicity a flat distribution in time was assumed. However, this needs not be the case, as is apparent for the earliest part of γm, where the true distribution lies below the uniformly distributed probability.
An additional statistical spread in T from amplification on later stages occurs for effects with both point-like T and T ranges. This was approximated by a Gaussian distribution, which is in very good agreement with data where reliable fits of individual effects were possible (γk, E1, γ1, and L1 in fig. 2.46). Therefore where no fits were possible, a Gaussian smearing using the fitted σ_T of γk was used to obtain the PDF of each effect and p_m.
Due to the many probability factors involved in the calculation of p_c, of which a large number had to be estimated, the peak values have large uncertainties and can differ from the real value possibly by up to a factor of ten or more if many unknowns are involved. For the dominant effects, which unsurprisingly are mostly from single interactions [125], in most cases only a moderate number of probability factors is involved, and the predictions are quite accurate. The p_m for γ1 for example is predicted (only) about 45% too high, which is not much when considering that no less than nine factors (none of which were taken from data) are involved in the calculation of p_t, as well as seven factors with uncertainties (some from data, some estimates) in the computation of p_m from p_t.

Effects

The individual EP effects shall be examined next.

As expected, γk is the most distinctive feature in the TTD. It shows a mostly Gaussian shape close to the peak, with a fitted mean of 59.85 ns and $\sigma = 0.799$ ns, which points to a well-controlled CTS for the majority of early RP.

In between the RP and E1 a cluster of several strong effects is found (most notably sh1E1, sa1E1, BE1). Since their exact times and probabilities are too dependent on field geometry details to be accurately predictable (especially the curve shape of BE1) it is difficult to fit this region reliably.

E1 on the other hand can be accurately fitted with a Gaussian on the left flank. However, its T is reconstructed with too small earliness from analytic calculations (−3.19 ns) compared to the fitted peak position (−4.20 ns), likely due to inaccurate knowledge of the field geometry [126]. The calculated peak value lies lower than the curve in data due to strong contributions of other effects to the bin count (see CTD in fig. 2.45). If heights $< 3.1\%$ are excluded, the bin entries are reduced by a factor of about 2, and the curve matches the calculated value. This way the value for the forward scattered fraction E1 was retrieved as described in 2.3.5.3.

Most effects lie close to T_0, between E1 and γk, with only small contributions to the overall TTD curve and are covered by more prevalent effects. It is nigh impossible

[125] With the exception of γMg1 and γ1L2 – but these can hardly be considered dominant, as their p_c is small and they would be unobservable if they were not so isolated.
[126] Assuming a purely quadratic field, a value of −4.01 ns follows, which is far closer to data, but still not early enough.

to ascertain their exact p and T values due to the multitude of overlapping effects within a small T range and the presence of several stronger effects at comparable earlinesses.

Not many effects can occur earlier than E1. Most of these lie close to E1 and are covered by its rising flank by several order of magnitude.
The cluster around $-10\,\mathrm{ns}$ most likely is caused by γm – a photon producing a photoelectron on the deck, which is then accelerated to d1. It is expectable that the probability of pm production and capture on d1 strongly depends on the pm emission position, leading to a non-flat distribution in time. The largest earliness is expected for pm emitted close to the aperture to d1 (confer fig. 2.39b). While this should have a large CE, the concentric area around the gap increases quadratically with radius, which results in a p_c drop towards earlier times, which is smeared by dynode time resolution.
However, there might be alternative explanations for the events around $-10\,\mathrm{ns}$. p_c strongly depends on h_s, and a small variation can greatly change p_c for small charge effects. Therefore the cluster might be from a known effect, which was inadvertently excluded. p_r sets a threshold-independent upper limit for the achievable probability of an effect. Looking at the p_r of effects in table 2.2, e.g. p3 (2.9%, $-6.2\,\mathrm{ns}$) or E1E2 (0.061%, $-5.2\,\mathrm{ns}$) would be candidates, if a systematic underestimation of earliness is assumed, as is the case for the other electronic EPs (caused by interactions of electrons, not photons). However, when looking at the CTD it quickly becomes clear that the charges of these effects is too small, which excludes them as possible explanations. The events could also be from an unknown effect.

Even earlier lies the cluster around $-14\,\mathrm{ns}$, which can be attributed to γ1L2, with a predicted mean time which is slightly too late, and a probability underestimated by roughly a factor of three (12 probability factors $\rightarrow \approx 9.6\%$ mean deviation of factors from true value).
For effects between the γ1-based cluster and E1 only a few effect types are possible. Purely electronic EPs cannot extend to such early times with large enough probabilities, since the earliness can only be gained from reduced interdynode transit times, which requires to chain En effects. Only E1E2E3 comes even close ($-8.1\,\mathrm{ns}$), but its pulses have a mean charge of 0.09% pe, rendering them undetectable both in h_s (only one pulse is produced) and h_f; p_r would be only $1.7 \cdot 10^{-5}$ anyway. Photonic EP effects on the other hand are either too early (from p1) or too late (pm). Only photoelectron creation on d1 can be early enough, but large subsequent delays have to be introduced somehow. Multiple reflections are not enough due to the short photon transit times – this produces the trailing flank of the γ1 peak. The solution is the introduction of electronic delays after a photonic effect, e.g. through γ1L2.
Another candidate could be γ1L11 (not calculated). L2 and s1L11 have similar U_0 as the p1 emitted in γ1L2 resp. γ1L11 and produce maximum delays of 8.7 and 5.8 ns. γ1L11 could thus maybe account for the earlier events in the cluster around $-14\,\mathrm{ns}$, and γ1L2 for the later. γ1L1 should be similar to γ1L11, with a fluent transition between both effects, and could also play a role.

Finally, the earliest observable grouping is caused by γ1 and its higher-order effects, e.g. from reflections, such as γMg1. Calculations are in good agreement with TTD data; the remaining differences can easily be explained by small errors in the estimated variables and especially h_s.

The contributions to the trailing flank after $\gamma 1$ can not be from CTS alone, since this introduces far less fluctuations than for γk, because $t\gamma$k1 is much shorter than tk1. The statistical spread in time can be estimated from the left flank of $\gamma 1$, which as expected is smaller than for γk, and cannot be the sole reason for the right flank. On the other hand, multiple reflections on k, g, and m will certainly contribute to the trailing flank of $\gamma 1$.
Furthermore, the photon hit position on d1 will introduce a dynode time spread due to increased ($x_{12\uparrow}$, L11-like) or decreased ($x_{12\downarrow}$, E11-like) traversed distances of the photoelectron to d2.
Finally, second order electronic effects such as γ1E11, γ1L11 (if occurring earlier than the -14 ns cluster), γ1-p1L1R2, γ1-p1L11 or others should play a role. The difference between the left and right flank of $\gamma 1$ thus can be explained through a number of TSEs.

All in all, the EPP-TTD can be said to be well-understood through TST.

Charge–time distribution

The charge–time distribution from EP effects for primary pulses (EPP-CTD) is shown in fig. 2.45.
As already mentioned in the discussion of the TTD, the calculated earliness is systematically too low for electronic EPs.

Thresholds

Down to the instrumental threshold h_s (4.5% pe), indicated by dashes, all effects trigger acquisition and can be observed [127]. Below the data analysis threshold h_f (1.44% pe), shown as dotted line, no effect can be measured, since the pulses become indistinguishable from noise spikes.
In between h_s and h_f, *not all* effects will be visible in data, since a pulse with height $> h_s$ must be present in the waveform to trigger acquisition, thus allowing to find the smaller pulse later in data analysis. That means that all observed EP effects $< h_s$ must have a larger AP; small effects which produce only PPs are unobservable. This is an important caveat for this and the following figures where the charge is the ordinate: Not all occurring effects are observable with this measurement method down to h_f; there will be even more than plotted. This dependence on both h_s and h_f was taken into account by the introduction of p_c, which is required to be $> 10^{-5}$ for effect inclusion in the CTD: Only observable effects are plotted. The unobservable effects which occur in the transition zone (either p_s or $p_f > 10^{-5}$, the other below) are listed in table 2.2 and discussed later.

The instrumental threshold of the setup was measured to 9.9% pe (2.4 mV), but the minimum charge observed for single-pulse effects in data reaches far lower values, which points to unaccounted small offsets introduced by the electronics. h_s was therefore estimated from the distributions of $\gamma 1$ and Rm (see p. 295) in the CTD,

[127] Save for statistical peak variations in the region near h_s due to noise and varying amplification, and fluctuations of h_s itself. However, too low pulses can be raised above the threshold in the same manner, which compensates the effects for constant contributions; only the first and higher derivatives in charge of the CTD lead to changes.

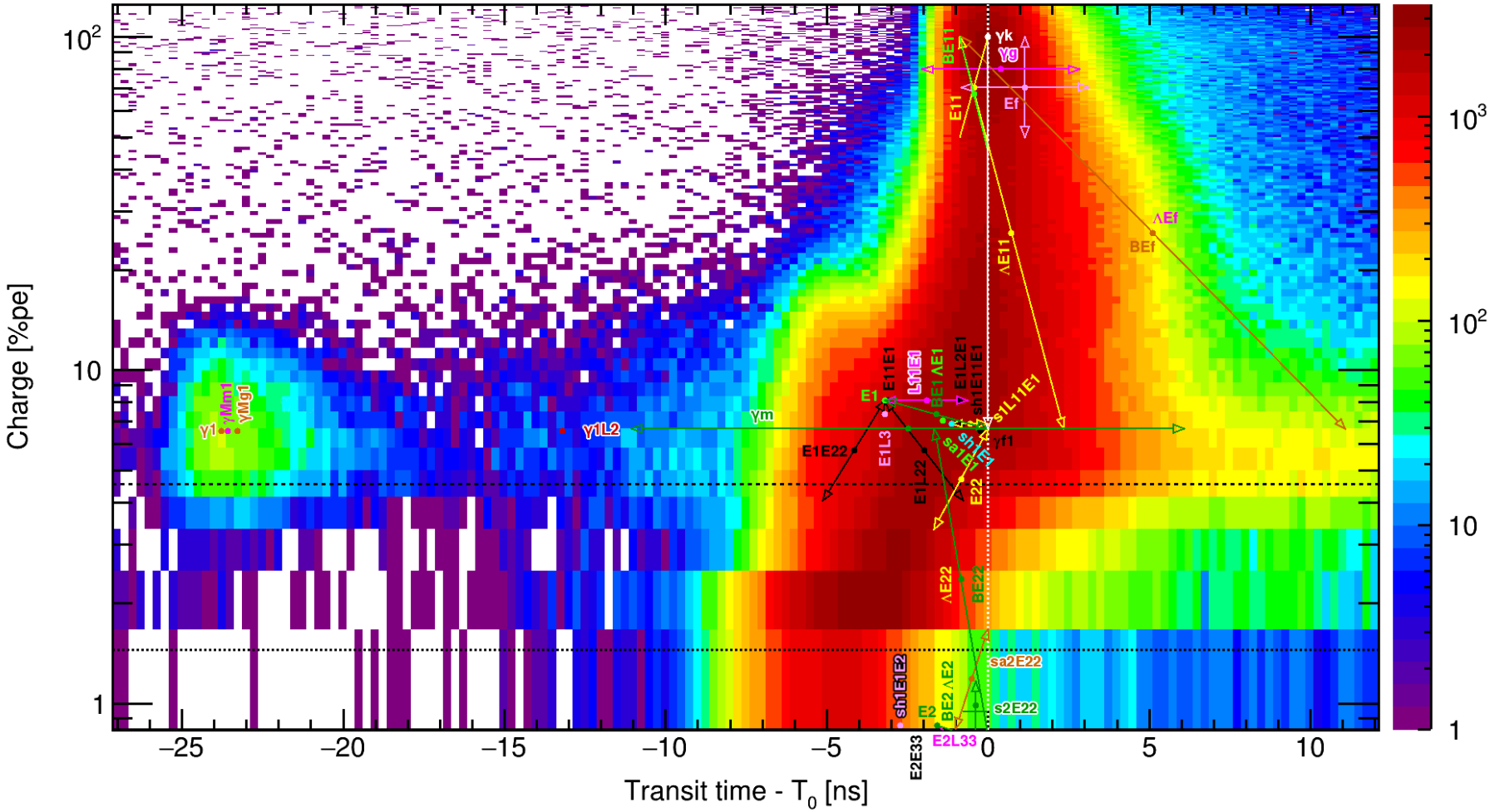


Figure 2.45: Charge–time distribution (CTD) for early primary pulses (EPP-CTD) from data and calculated with TST. The plot shows the charge q (y-axis) and detectable probability p_c (z-axis) of early primary pulses vs. their relative transit time T_r (x-axis). — Measured distribution (SPAX, $o = 2.9\%$, pulse fitting, $h_s = 4.5\%$ pe, $h_f = 1.44\%$ pe, bin widths 0.25 ns and 0.02 pC) for a Hamamatsu R5912 (+1425 V, g = $1.3 \cdot 10^7$). 1 pe = 2.40 pC (charge of pure γk pulses selected with stringent cuts). Number of pulses per bin indicated by z-axis (colors). Instrumental threshold h_s dashed line, PEST (see p. 481) fit routine threshold h_f dotted line; below h_s effects can only be detected if a pulse $> h_s$ triggered acquisition. The lowest bins are only partially filled and therefore have reduced bin contents. The dotted white line shows the RP time (γk and others); everything before is by definition "early". Some effects can also cause lateness ($T_r > 0$, e.g. BEf or γm). — Effects calculated with TST which can cause earliness, have a probability $p_c > 10^{-5}$ and produce measurable charges are drawn in with labels, markers and arrows/lines, whose color depicts the detectable probability: $> 65\%$ white/violet (above this value $N > 1$ occurrences per PP, except for γk), $> 10\%$ cyan/blue, $> 1\%$ bright/dark green, $> 0.1\%$ yellow/orange, $> 0.01\%$ pink/magenta, $< 0.01\%$ red/black. Statistic spreads in time and charge of effects not shown. Datapoints without lines denote the expected distribution centers. Systematic ranges in time or charge are shown by arrows or lines. Diagonal lines imply the tendency of charge–time correlations; the exact curves are laborious to calculate. The shown effects were calculated analytically ab initio, *without* using input values from the shown measurement, with the few exceptions given in the text. — All in all the effects predicted by TST can explain all clusters of the observed early behavior, although the calculated earliness for some effects is systematically too low (e.g. E1). This is likely because analytic calculations were performed without knowledge of the exact field layout. Individual effects are discussed in the text. — For the corresponding TTD see fig. 2.44 (EPP-TTD), for the CTD of late pulses fig. 2.47 (LPP-CTD). The complete CTD for a classic evaluation (without pulse fitting) of the same measurement is shown in fig. 3.14d.

which are well suited for this task, since they are isolated from other effects and only produce a single pulse. Both are expected to show a Poissonian variation in charge due to fluctuations in $\delta 2$ and should extend down to values in the order of $1/\delta 1 \delta 2 = 0.77\%$; likely somewhat higher, as discussed for γk before, since the probability of producing only one s2 is greatly suppressed. Both distributions, however, abruptly disappear at far higher, comparable values, which allows to determine an effective threshold of about 4.5% pe, equivalent to 1.1 mV.

Regarding the choice of instrumental threshold, a value of 20% pe effectively eliminates most EPs (for this particular PMT). This shows once again the benefits of maximizing $\delta 1$, since higher values better suppress the charge of almost all EP events, and then the instrumental threshold can be shifted lower while achieving the same EP suppression. This way less PDE is lost from RP exclusion. As an added benefit, a higher $\delta 1$ reduces the charge spread of RP events, so at low charges less and less RP pulses remain, which further improves separability. It would be interesting to study the TTD and CTD for a PMT with a GaP first dynode.

Small charges

In the plot, effects with mean charges or charge ranges lower than 0.8% pe are not shown. Nevertheless, effects beneath h_f can still contribute to the TTD if their charge distribution extends beyond the threshold, either from statistical spread or due to a high mean occurrence N per pulse.

The lowest shown charge bins (0.8–1.7% pe) are only partially filled since they lie only partly above h_f. Therefore their bin content is reduced compared to the second-lowest bins, even though in the range from about −5 to 0 ns many effects should predominantly fill the lowest bins.

On the other hand, effects with small charge result in larger bin contents around the effect center than effects with identical p_c occurring at larger charge. This is because of the linear binning, as a charge distribution extends over less bins at small charges. The charge approaches zero in a logarithmic manner from additional skipped amplification steps. If the bin height were scaled exponentially to counter this effect, it would become impossible to estimate probabilities from the plot by eye due to the varying bin size, so this was avoided.

The region directly below the RP at small charges could be unresolvable, if an EP/LP with small charge and low T_r as well as an RP-type pulse occur. This is because the fit cannot reliably distinguish a very small pulse close to a large pulse from statistical charge fluctuations of the latter. It is difficult to determine the exclusion area in the CTD resulting from this effect. The region with missing events at low charges up to 3 ns in the AP-CTD (fig. 2.60) and the CDD (fig. 2.56) could indicate its extent, but could also be simply from the fact that there are no effects present in this region.

Another evidence for the large fraction of pulses which are underamplified comes from the dark count pulses. These are visible in fig. 2.38 before 34 ns (and after 164 ns with relatively low “background” from the correlated events). While the low number

of recorded dark noise events does not permit to obtain an accurate dark count SER, it is clearly visible that underamplified pulses constitute a large fraction of events from the accumulation of pulses at low charges. As already discussed on p. 108, the underamplified charge fraction is expected to be even larger for dark counts than for photon pulses.

Influence of average pulse shape

The pulse search routine used to obtain all plots is based on fitting one or more overlapping pulses with an APS, which was created from RP events selected with rigorous cuts to ascertain that there was no contamination with secondary pulses in form of pileup. After a successful fit, all reliably found pulses are subtracted to reveal smaller overlapping pulses, which were covered before.
However, if pulses resulting from TSEs have pulse shapes differing from the RP, the fit model does not match and their charge or time will be reconstructed wrongly, which can create artificial pulses when subtracted.
To avoid this, the residuals remaining after subtraction of each fitted pulse were studied extensively. The mismatch is expected to rise with decreasing charge, since the pulse gets broader in time; the relative offset in time plays no role. The observed residuals were within expectations, so it can only be concluded that the pulse shape of all studied effects with comparatively large charge is described well by the used average pulse shape.

Effects

In the following, individual effects are compared with data (fig. 2.45).

γk shows the expected spread to higher and lower charges, as well as to earlier and later T due to the effects discussed in 2.3.3 and 2.3.6.2. A reduction can be seen down to about 30 % pe, when contributions from other effects become dominant.

At regular charges and slightly earlier than the RPs, minor contributions from BE11, E11 and γg add to the distribution, which still originates mainly from the TTS of γk.

For the events which lie between -6 and -2 ns before γk at q between 20 and 100%, and which show a rising charge–time dependence, there are two possible explanations:
The time range of γg was only coarsely estimated based on a multiple of the CTS. The traversed voltage should be equal for all emission sites on g. However, if the pg is produced at a location closer to the dynode mount, the extraction fields could be stronger, which would boost the pg in the most time-consuming early phase and could lead to an earliness of several nanoseconds. At the same time, the further the emission position is from k, the less accurate the focusing optics will work, which leads to suboptimal hit position on d1 with decreased δ1, as observed in data.
An alternative explanation could be a combination of CTS and edge effect. In surface scans of the CTS, for some PMTs regions near the cathode brink with large earliness are observed (R7081 (10") [342], R11780 (12") [190]), which result from non-optimal pk trajectories. It stands to reason that similar regions with shortened tk1 are

present in the R5912 as well, and that such badly focused trajectories are coupled to suboptimal hit positions on d1, which can decrease $\delta 1$ through the edge effect (CE_{21} and material properties). The origin of the decrease in tk1 could be from stronger extraction fields due to the proximity of the dynode mount (which typically is at d1 potential) or that the region on d1 close to d2 is hit, which would decrease s2 transit time as in the E11 effect. The latter seems more likely due to the abrupt shifts in earliness observed in the abovementioned publications.
To obtain more information about the pulses in this region, they were isolated by cuts in time (−7.5 to −2.4 ns) and charge (0.6 to 2.4 pC). For the remaining PPs, the charge distribution follows an exponential decay and the charge is reduced compared to γk. For the attendant AP, T lies at T_0 or before, but the AP charge for most events is far lower than for γk and shows the behavior of some frequent AP effects. Thus the APs are not of RP-type and the events must be caused by effects affecting the photoelectron, either by a changed generation site or altered pk trajectories. Judging from the obtained CDD and FDD only a few dominant AP effects take place: BE11 or BE1 and s1L1 or s1L11, where the low charge part is preferably filled, and the AP delay is slightly altered, since these effects are the second interaction. So afterpulses can be created on d1, but with smaller charges. Effects on d2, however, are suppressed in probability compared to the regular distributions. Both is likely because less SEs are produced on d1. BE1 as secondary effect is unlikely, as it would reduce the PP charge to the E1 level, which is not the case, which leaves BE11, s1L1 and S1L11 as dominant AP effects. In sum, afterpulses occur with very high probability (80%), which for the mentioned effects is far higher than normally. This points to changed d1 hit positions, which could also explain the reduced PP charge. No strong correlations in charge or time between PP and AP are observed, except for one slight correlation of BE1/BE11 type; the AP effects are mostly centered on small AP charges. This is therefore not caused by branching behavior, although this can be present as a secondary effect. The sum charge of PP and AP is lower than for RPs.
In conclusion, the effect causing the observed pulse cluster likely stems from photoelectrons with changed origin or emission position, which hit d1 not on the regular impact region which results in a reduced $\delta 1$. No candidate effect (γg or CTS) can thus be excluded, but the observations limit other possible causes.

The region from about 10–30% pe near T_0 (−2 to 0 ns), which shows a rising charge–time behavior, also is not covered by predicted TST effects.
Events here could be caused either by the CTS/TTS in combination with edge effect as discussed aboves since the $q(T)$ behavior is similar and there is a smooth transition between both regions.
What *is* expected from TST but not drawn into the CTD, are the RP with reduced charge which result from B-type effects. From BL11 and BL1 about 7% of PP are expected to have reduced charges ranging from 100% pe down to $1/\delta 1$, which can explain a great amount of the pulses found at low charges, but not the rising $q(T)$ dependence found for early pulses.
Alternatively, BE1 and ΛE1 might be the origin. These effects are expected to occur more frequently for short earliness (see p. 284). The E1 charge distribution is quite broad and reaches up to 20% pe, so similar broadening is expected for BE1 and ΛE1 whose minimal charge is not much lower. The charge spread together with rising probabilities could explain the observed $q(T)$ behavior.

Furthermore, events in this region could be caused by two simultaneous occurrences of frequent early effects (sh1E1 or sa1E1), which cannot be resolved by the fit routine due to their proximity in time and thus are treated as one pulse with double charge. This explains the region above those effects, but not the slope towards higher q with rising T.
Instead, it could also be that a frequent EP effect which also produces an RP (sh1E1, BE1, sa1E1, BE22, sh2E2, sa2E2, etc.) is not recognized by the fit. In this case the EP is a pileup pulse on the rising flank of the RP, which shifts the RP time (determined by a software CFD, see pp. 155, 335) earlier. For BE1 and BE22 an earlier PP results in a smaller RP-AP *and* a reduced charge sum, which would explain the observed $q(T)$ dependence for merged pulses. This might also be an alternate explanation for the events a few nanoseconds before γk at high charges, although for such early pulses the fit routine is expected to reliably recognize pileup.
Finally, what is visible here might in fact be the Shockley-Ramo pulse (p. 171), as one would expect to be able to resolve it with such a low threshold and good time resolution.
In conclusion, the events shortly before T_0 at charges between 10 and 30% could be from a time dependence of the probability of BE1 and ΛE1, unresolved pileup of BE1 and BE22, or the Shockley-Ramo pulse of γk.

E1 has a smaller calculated earliness than in data, but otherwise the predictions match the observed cluster. The statistic spread in time and charge is well visible. It should be noted that the region above E1 can *not* be from multiple simultaneous occurrences of E1, since there typically is only one pk unless the occupancy is large. Since o was 2.9%, and $p_{c,\mathrm{E1}} = 2.0\%$ a small population of 2-E1, and even 3-E1 events is expected, but this cannot explain the events lying before γk at high charges.

The effects E2, BE2, ΛE2 likely occur a factor of 2 earlier than predicted, which when considered produces a good match with the cluster around −4 ns and 2% pe, since the N of these effects is rather high. The cluster could, however, also originate from s1E13, sh1E13, and sa1E13 (not shown), which should have charges comparable to E2, larger earliness and sufficiently high p_c, especially s1E13.

γm is in very good agreement with the cluster visible from −13 to −7 ns and 6% pe; at lower earliness γm is covered by stronger effects.

γ1L2 is also somewhat late compared to data, but the predicted charge is identical to the cluster found around −14 ns and 6.5% pe.

Finally, γ1 and its secondary effects are in perfect agreement with data, which indicates that δ1 was predicted correctly.

For some EP effects, the covered time range is that great that their primary pulses even can lie after γk (i.e. have lateness), even though they are caused by forward (E-type) scattering or arise from photons hitting structures other than k. This is the case for Ef, BEf, ΛEf, ΛE11, γm and γg; and of course γk in the context of TTS. The LP-behavior of these effects is discussed in the LP section together with other effects occurring at similar lateness.

Altogether, the CTD can be said to be well understood. While some matters are still unresolved, for each there are possible explanations.

Table of effects

In table 2.2 for all EP effects the topological notation, the equations for p_r, q and T for both PP and APs, if present, as well as the values calculated for the R5912 are given, including N (where different from p_r), p_s and p_f.
Sub-effects (from PP–AP crossover or sn + shn, e.g. BE11, s1L11E1) and additional APs beyond the first (for example E2sh3E3) are listed on separate rows. p_s and p_f are adapted for each sub-effect / AP.
Often p_s is smaller than p_f for effects with small charges, since h_s is larger than h_f and a smaller fraction of the charge distribution lies above the threshold. However, p_f can also be *lower* than p_s if very small pulses are present which are hard to detect even for the fit. This is because p_f requires the detection of *all* pulses, whereas for p_s only the largest pulse has to be registered.
The table should be self-explanatory, all terms and calculations are explained in the section on TST 2.3.5. The results can best be visualized and understood by comparison with the TTD and CTD shown in this section and 2.3.6.4, and the AP distributions shown in 2.5.2.2. Not all effects listed in the table are observable.

Individual effects of note are discussed in the following. First, single-interaction effects are analyzed, starting with incident photons, followed by interdynode and intradynode effects, scattered stage skipping, SE interdynode and intradynode effects, SE stage skipping, pk interactions on structures other than d1 (including stage skipping), and concluding with EPs from luminescence. Afterwards dual, triple and quadruple interactions are discussed where relevant in the same order of effect types.
Effects which are statistic contributions to the APS and indistinguishable from it are marked by placing their title in parentheses "( )". Those which are physically impossible (no electrons are produced or can escape at one stage, or the arrival factor is zero) are indicated by square brackets "[]". And effects which have too low probabilities p_c for observation are denoted by curly brackets "{}".
In the table the effects are sorted in the same manner as in the discussion.

Incident photon

Photoeffect

This event class contains effects where the incident photon is subject to photoeffect on a structure in the device. Since necessarily nearly all structures in a PMT are closer to the anode than the cathode, in most cases this introduces earliness [128], because the transit times are far shorter for photons than for electrons. Thus most of these effects have an earliness beyond of what is reachable with electron-based effects.
Only interactions on structures other than k are discussed, but strictly speaking, the early fraction of γk is part of this group as well.

[128] γm and γg can also be late.

Table 2.2: Effects causing early pulses that were derived with TST. Values for a Hamamatsu R5912 at $g = 1.3 \cdot 10^7$, $T_0 = 59.85\,\mathrm{ns}$, $T_{\mathrm{FWHM}} = 2.26\,\mathrm{ns}$, measured with $h_s = 4.5\%$ pe, $h_f = 1.44\%$ pe. Statistical spread not included, ranges from systematic causes denoted by square brackets. Numbers typically with two significant digits, within 10% of 100% with one digit after decimal point. To condense the topological notations, initial regular amplification steps can be left out and down left arrows are the same as down right arrows. N is given in brackets, where the rounded value is different from p_r. $T_{\mathrm{PP}} = 0$ for γk by definition. AP entries only filled if effect creates APs. For comparison the regular pulse (γk) is listed first; its charge is on average 1 pe, at least $1/\delta 1$, and can also be >100% pe due to fluctuations (not included). Table continued on next pages. Discussion see text.

Effect	**Topology**	$\boldsymbol{p_r}$	$\boldsymbol{p_r(N)}$	p_s	p_f	$\boldsymbol{q_{PP}}$		$\boldsymbol{T_{PP}}$		$\boldsymbol{q_{AP}}$		$\boldsymbol{\Sigma_q}$	$\boldsymbol{\eta_q}$	$\boldsymbol{T_{AP}}$		$\boldsymbol{d}$
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
γ**k**	γPk→pR1	1	100	99.12	99.73	1	[(6.7,) 100.0]	0	0							
γ**1**	γP1→p1R2	$\mathcal{T}\frac{\mathrm{A_1}}{\mathrm{A_k}}\frac{\mathrm{QE_1}}{\mathrm{QE_k}}\frac{\mathrm{CE_{12}}}{\mathrm{CE_{k1}}}\cdot\frac{\mathrm{u2_{p1}}}{\mathrm{u1}}$	0.31	0.22	0.29	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	$t\gamma\mathrm{k1}-t\mathrm{k1}+t12_{\mathrm{p1}}-t12$	−23.8							
γ**g**	γPg→pgR1	$\mathcal{T}\frac{\mathrm{A_g}}{\mathrm{A_k}}\frac{\mathrm{QE_g}}{\mathrm{QE_k}}\frac{\mathrm{CE_{g1}}}{\mathrm{CE_{k1}}}\cdot\frac{\mathrm{u1_{pg}}}{\mathrm{u1}}$	0.069	0.068	0.069	$\frac{\delta 1_{\mathrm{pg}}}{\delta 1}\mathrm{r1_g}$	80	$[0, 2t\gamma\mathrm{k1}]+[t\mathrm{g1}_\downarrow, t\mathrm{g1}_\uparrow]-t\mathrm{k1}$	[−2.0, 2.8]							
γ**m**	γPm→pR1	$\mathcal{T}\frac{\mathrm{A_m}}{\mathrm{A_k}}\frac{\mathrm{QE_m}}{\mathrm{QE_k}}\frac{\mathrm{CE_{m1}}}{\mathrm{CE_{k1}}}\cdot\frac{\mathrm{u1_{pm}}}{\mathrm{u1}}$	1.9	1.4	1.8	$\frac{\delta 1_{\mathrm{pm}}}{\delta 1}$	6.7	$t\gamma\mathrm{km}+[t\mathrm{m1}_\downarrow, t\mathrm{m1}_\uparrow]-t\mathrm{k1}$	[−11.1, 6.1]							
γ**f**	γPf→pfTf	$\mathcal{T}\frac{\mathrm{A_f}}{\mathrm{A_k}}\frac{\mathrm{QE_f}}{\mathrm{QE_k}}\frac{\mathrm{CE_{f1}}}{\mathrm{CE_{k1}}}\cdot\frac{\mathrm{u1_{pf}}}{\mathrm{u1}}$	0	0	0	-		-								
γ**Mg1**	γMg1→γP1→p1R2	$\mathcal{T}\frac{\mathrm{A_{Mg1}}}{\mathrm{A_k}}\mathrm{R_g}\frac{\mathrm{QE_1}}{\mathrm{QE_k}}\cdot\frac{\mathrm{CE_{12}}}{\mathrm{CE_{k1}}}\frac{\mathrm{u2_{p1}}}{\mathrm{u1}}$	0.22	0.16	0.20	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	$t\gamma\mathrm{Mg1}-t\mathrm{k1}+t12_{\mathrm{p1}}-t12$	−23.3							
γ**Mm1**	γMm1→γP1→p1R2	$\mathcal{T}\frac{\mathrm{A_{Mm1}}}{\mathrm{A_k}}\mathrm{R_m}\frac{\mathrm{QE_1}}{\mathrm{QE_k}}\cdot\frac{\mathrm{CE_{12}}}{\mathrm{CE_{k1}}}\frac{\mathrm{u2_{p1}}}{\mathrm{u1}}$	0.048	0.035	0.045	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	$t\gamma\mathrm{Mm1}-t\mathrm{k1}+t12_{\mathrm{p1}}-t12$	−23.6							
γ**M12**	γM12→γP2→pR3	$\mathcal{T}\frac{\mathrm{A_{M12}}}{\mathrm{A_k}}\mathrm{R_d}\frac{\mathrm{QE_2}}{\mathrm{QE_k}}\cdot\frac{\mathrm{CE_{23}}}{\mathrm{CE_{k1}}}\frac{\mathrm{u3_{p2}}}{\mathrm{u1}}$	0.10	0	$7.4 \cdot 10^{-3}$	$\frac{1}{\delta 1 \delta 2}\frac{\delta 3_{\mathrm{p2}}}{\delta 3}$	0.8	$t\gamma\mathrm{k1}-t\mathrm{k1}+t\gamma 12-t12+t23_{\mathrm{p2}}-t23$	−28.4							
E1	pE1→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{u2_{p\varepsilon 1}}$	2.5	2.0	2.4	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1}}{\delta 2}$	8.1	$t12_{\mathrm{p}\varepsilon 1}-t12$	−3.2							
ΛE1	pΛE1→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{i1}\cdot\Lambda 1\cdot\mathrm{E}\Lambda 1\cdot\mathrm{u2}$	1.9	1.5	1.8	$\frac{1}{\delta 1}\left[1, \frac{\delta 2_{\mathrm{p}\varepsilon 1}}{\delta 2}\right]$	[6.7, 8.1]	$[t12_{\mathrm{p}\varepsilon 1}, t12]-t12$	[−3.2, 0]							
BE1	pBE1→a1R2 ↘ pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\cdot\mathrm{i1}\cdot\mathrm{B1}\cdot\mathrm{EB1}\cdot\mathrm{u2}$	7.4	7.4	7.1	$\frac{1}{\delta 1}\left[1, \frac{\delta 2_{\mathrm{p:\varepsilon s:1}}}{\delta 2}\right]$	[6.7, 8.1]	$[t12_{\mathrm{p}\varepsilon 1}, t12]-t12$	[−3.2, 0]	$\frac{[1, \delta_{\mathrm{s}} 1]}{\delta 1}$	[6.7, 100.0]	[14.8, 106.7]	[45, 93.8]	0	0	[0, 3.2]

Effect	**Topology**	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
E2	pR1→r1R2 ↘ s1E2→s1R3	$1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{E2}\cdot\mathrm{u3}_{\mathrm{s1}\varepsilon 2})^{\delta 1}$	14 (0.15)	14	2.0	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{s1}\varepsilon 2}}{\delta 3}$	0.9	$t23_{\mathrm{s1}\varepsilon 2}-t23$	−1.6	$1-\frac{1}{\delta 1}$	93.3	94.2	99.1	0	0	1.6
ΛE2	pR1→r1R2 ↘ s1ΛE2→s1R3	$1-(1-\mathrm{b2}\cdot\mathrm{i2}\cdot\Lambda 2\cdot\mathrm{E}\Lambda 2\cdot\mathrm{u3})^{\delta 1}$	9.3 (0.097)	9.2	1.3	$\frac{1}{\delta 1\delta 2}\left[1,\frac{\delta 3_{\mathrm{s1}\varepsilon 2}}{\delta 3}\right]$	[0.77, 0.86]	$[t23_{\mathrm{s1}\varepsilon 2},t23]-t23$	[−1.6, 0]	$1-\frac{1}{\delta 1}$	93.3	[94.1, 94.2]	[99.1, 99.2]	0	0	[0, 1.6]
BE2	pR1→r1R2 →a2jR3 ↘ s1BE2 ↗ →s1R3	$1-(1-\mathrm{b2}\cdot\mathrm{i2}\cdot\mathrm{B2}\cdot\mathrm{EB2}\cdot\mathrm{u3})^{\delta 1}$	33 (0.39)	32	4.7	$\frac{1}{\delta 1\delta 2}\left[1,\frac{\delta 3_{\mathrm{s1}\varepsilon 2}}{\delta 3}\right]$	[0.77, 0.86]	$[t23_{\mathrm{s1}\varepsilon 2},t23]-t23$	[−1.6, 0]	$1-\frac{1}{\delta 1}+\frac{[0,\delta_{\mathrm{s}}2]}{\delta 1\delta 2}$	[93.3, 99.9]	[94.2, 100.6]	99.1, 99.2]	0	0	[0, 1.6]
E3	pR1→a1R2→r2R3 ↘ s2E3→s2R4	$1-(1-\mathrm{b3}\cdot\mathrm{e3}\cdot\mathrm{E3}\cdot\mathrm{u4}_{\mathrm{s2}\varepsilon 3})^{\delta 1\delta 2}$	87 (2.0)	87	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{\mathrm{s2}\varepsilon 3}}{\delta 4}$	0.11	$t34_{\mathrm{s2}\varepsilon 3}-t34$	−2.3	$1-\frac{1}{\delta 1\delta 2}$	99.2	99.3	98.6	0	0	2.3
ΛE3	pR1→a1R2→r2R3 ↙ s2ΛE3→s2R4	$1-(1-\mathrm{b3}\cdot\mathrm{i3}\cdot\Lambda 3\cdot\mathrm{E}\Lambda 3\cdot\mathrm{u4})^{\delta 1\delta 2}$	76 (1.4)	76	0	$\frac{1}{\delta 1\delta 2\delta 3}\cdot\left[1,\frac{\delta 4_{\mathrm{s2}\varepsilon 3}}{\delta 4}\right]$	[0.08, 0.11]	$[t34_{\mathrm{s2}\varepsilon 3},t34]-t34$	[−2.3, 0]	$1-\frac{1}{\delta 1\delta 2}$	99.2	[99.3, 99.3]	[99.9, 99.9]	0	0	[0, 2.3]
BE3	a1R2→r2R3 →a3jR4 ↘ s2BE3 ↗ →s2R4	$1-(1-\mathrm{b3}\cdot\mathrm{i3}\cdot\mathrm{B3}\cdot\mathrm{EB3}\cdot\mathrm{u4})^{\delta 1\delta 2}$	100 (5.7)	100	$3.2\cdot10^{-6}$	$\frac{1}{\delta 1\delta 2\delta 3}\cdot\left[1,\frac{\delta 4_{\mathrm{s2}\varepsilon 3}}{\delta 4}\right]$	[0.08, 0.11]	$[t34_{\mathrm{s2}\varepsilon 3},t34]-t34$	[−2.3, 0]	$1-\frac{1}{\delta 1\delta 2}+\frac{[0,\delta_{\mathrm{s}}3]}{\delta 1\delta 2\delta 3}$	[99.2, 100.0]	[99.3, 100.1]	[99.9, 99.9]	0	0	[0, 2.3]
E11	pE11→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E11}\cdot\mathrm{u1}_{\varepsilon 1}$	0.38	0.38	0.38	$\frac{\delta 1_{\varepsilon 1}}{\delta 1}[\mathrm{r1},1]$	[50, 100.0]	$[t11_{\downarrow}+t12_{\downarrow}-t12,0]$	[−0.9, 0]							
ΛE11	pΛE11→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{i1}\cdot\Lambda 1\cdot\mathrm{E}\Lambda 11\cdot\mathrm{u1}_{\Lambda 1}$	0.31	0.31	0.31	$\left[\frac{1}{\delta 1},\frac{\delta 1_{\varepsilon 1}}{\delta 1}\right]$	[6.7, 100.0]	$[t11_{\downarrow},t11_{\Lambda\downarrow}]+t12_{\downarrow}-t12$	[−0.9, 2.3]							
BE11	pBE11→a1R2 ↘ pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{i1}\cdot\mathrm{B1}\cdot\mathrm{EB11}\cdot\mathrm{u1}_{\mathrm{B1}}$	1.3	1.2	1.3	$\left[\frac{\delta_{\mathrm{c}}1_{\mathrm{B1E}}}{\delta 1},\frac{\delta 1_{\varepsilon\mathrm{s}:1}}{\delta 1}\right]$	[45.4, 100.0]	$[t11_{\downarrow}+t12_{\downarrow}-t12,0]$	[−0.9, 0]	$\frac{[1,\delta_{\mathrm{c}}11_{\mathrm{B1E}}]}{\delta 1}$	[6.7, 99.0]	[106.7, 144]	[6.2, 69]	0	0	[0, 0.9]
					1.2	$\left[\frac{\delta_{\mathrm{c}}11_{\mathrm{B1E}}}{\delta 1},\frac{\delta_{\mathrm{s}}1}{\delta 1}\right]$	[99.0, 100.0]	0	0	$\frac{[1,\delta_{\mathrm{c}}1_{\mathrm{B1E}}]}{\delta 1}$	[6.7, 45]	[106.7, 144]	[6.2, 31]	$[0,t11_{B\downarrow}+t12_{\downarrow}-t12]$	[0, 2.3]	[0, 2.3]
E22	pR1→r1R2 ↘ s1E22→s1R2	$1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{E22}\cdot\mathrm{u2}_{\varepsilon 2})^{\delta 1}$	0.56	0.56	0.53	$\frac{1}{\delta 1}\frac{\delta 2_{\varepsilon 2}}{\delta 2}[\mathrm{r2},1]$	[3.3, 6.7]	$[t22_{\downarrow}+t23_{\downarrow}-t23,0]$	[−1.7, 0]	$1-\frac{1}{\delta 1}$	93.3	[96.7, 100.0]	[93.3, 96.6]	0	0	[0, 1.7]
ΛE22	pR1→r1R2 ↘ s1ΛE22→s1R2	$1-(1-\mathrm{b2}\cdot\mathrm{i2}\cdot\Lambda 2\cdot\mathrm{E}\Lambda 22\cdot\mathrm{u2}_{\Lambda 2})^{\delta 1}$	0.39	0.39	0.37	$\frac{1}{\delta 1}\left[\frac{1}{\delta 2},\frac{\delta 2_{\varepsilon 2}}{\delta 2}\right]$	[0.77, 6.7]	$\left[[t22_{\downarrow},t22_{\Lambda\downarrow}]+t23_{\downarrow}-t23,0\right]$	[−1.7, 0]	$1-\frac{1}{\delta 1}$	93.3	[94.1, 100.0]	[93.3, 99.2]	0	0	[0, 1.7]
BE22	pR1→r1R2 →a2jR3 ↘ s1BE22 ↗ →s1R2	$1-(1-\mathrm{b2}\cdot\mathrm{i2}\cdot\mathrm{B2}\cdot\mathrm{EB22}\cdot\mathrm{u2}_{\mathrm{B2}})^{\delta 1}$	1.5 (0.016)	1.5	1.5	$\frac{1}{\delta 1}\left[\frac{1}{\delta 2},\frac{\delta 2_{\varepsilon 2}}{\delta 2}\right]$	[0.77, 6.7]	$\left[[t22_{\downarrow},t22_{\mathrm{B}\downarrow}]+t23_{\downarrow}-t23,0\right]$	[−1.7, 0]	$1-\frac{1}{\delta 1}+\frac{[0,\delta_{\mathrm{s}}2]}{\delta 1\delta 2}$	[93.3, 99.9]	[100.0, 100.6]	[93.3, 99.2]	0	0	[0, 1.7]

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	$q_{\mathbf{PP}}$		$T_{\mathbf{PP}}$		$q_{\mathbf{AP}}$		Σ_q	η_q	$T_{\mathbf{AP}}$		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
E33	pR1→a1R2→r2R3 s2E33→s2R3	$1-(1-b3\cdot e3\cdot E33\cdot u3_{\varepsilon 3})^{\delta 1\delta 2}$	15 (0.17)	15	1.2	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\varepsilon 3}}{\delta 3}$ [r3,1]	[0.38, 0.77]	$[t33_{\downarrow}+t34_{\downarrow}-t34,0]$	[−0.8, 0]	$1-\frac{1}{\delta 1\delta 2}$	99.2	[99.6, 100.0]	[99.2, 99.6]	0	0	[0, 0.8]
ΛE33	pR1→a1R2→r2R3 s2ΛE33→s2R3	$1-(1-b3\cdot i3\cdot \Lambda 3\cdot E\Lambda 33\cdot u3_{\Lambda 3})^{\delta 1\delta 2}$	11 (0.12)	11	0.91	$\frac{1}{\delta 1\delta 2}\left[\frac{\delta_c 3_{\Lambda 3E}}{\delta 3}, \frac{\delta 3_{\varepsilon 3}}{\delta 3}\right]$	[0.22, 0.77]	$[t33_{\downarrow}+t34_{\downarrow}-t34,0]$	[−0.8, 0]	$1-\frac{1}{\delta 1\delta 2}$	99.2	[99.5, 100.0]	[99.2, 99.8]	0	0	[0, 0.8]
					0	$1-\frac{1}{\delta 1\delta 2}$	99.2	0	0	$\frac{1}{\delta 1\delta 2}\cdot\frac{[1,\delta_c 3_{\Lambda 3E}]}{\delta 3}$	[0.08, 0.22]	[99.3, 99.5]	[0.08, 0.22]	$[0, t33_{\Lambda\downarrow}+t34_{\downarrow}-t34]$	[0, 0.7]	[0, 0.7]
BE33	a1R2→r2R3 →a3jR4 s2BE33→s2R3	$1-(1-b3\cdot i3\cdot B3\cdot EB33\cdot u3_{B3})^{\delta 1\delta 2}$	38 (0.48)	38	3.1	$\frac{1}{\delta 1\delta 2}\left[\frac{\delta_c 3_{B3E}}{\delta 3}, \frac{\delta 3_{\varepsilon 3}}{\delta 3}\right]$	[0.22, 0.77]	$[t33_{\downarrow}+t34_{\downarrow}-t34,0]$	[−0.8, 0]	$1-\frac{1}{\delta 1\delta 2}+\frac{[0,\delta_c 33_{B3E}]}{\delta 1\delta 1\delta 3}$	[99.2, 99.9]	[100.0, 100.2]	[99.2, 99.8]	0	0	[0, 0.8]
					0	$1-\frac{1}{\delta 1\delta 2}+\frac{[\delta_c 33_{B3E},\delta_s 3]}{\delta 1\delta 2\delta 3}$	[99.9, 100.0]	0	0	$\frac{1}{\delta 1\delta 2}\cdot\frac{[1,\delta_c 3_{B3E}]}{\delta 3}$	[0.08, 0.22]	[100.1, 100.2]	[0.08, 0.22]	$[0, t33_{B\downarrow}+t34_{\downarrow}-t34]$	[0, 0.7]	[0, 0.7]
sh1E1	pR1→r1R2 sh1E1→sh1R2	$1-(1-s1_h\cdot E1_{sh1}\cdot u2_{sh1})^{\delta 1}$	11	11	10	$\frac{1}{\delta 1}\frac{\delta 2_{sh1}}{\delta 2}$	6.9	$t12_{sh1}-t12$	−1.1	$\frac{\delta_{sh}1}{\delta 1}$	100.0	106.9	93.5	0	0	1.1
sa1E1	pR1→r1R2 sa1E1→sa1R2	$1-(1-s1_a\cdot E1_{sa1}\cdot u2_{sa1})^{\delta 1}$	9.5 (0.099)	9.4	9.0	$\frac{1}{\delta 1}\frac{\delta 2_{sa1}}{\delta 2}$	7.1	$t12_{sa1}-t12$	−1.4	$\frac{\delta_{sa}1}{\delta 1}$	100.0	107.1	93.4	0	0	1.4
sh2E2	pR1→a1R2→r2R3 sh2E2→sh2R3	$1-(1-s2_h\cdot E2_{sh2}\cdot u3_{sh2})^{\delta 1\delta 2}$	60 (0.91)	59	5.2	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{sh2}}{\delta 3}$	0.8	$t23_{sh2}-t23$	−0.7	$1-\frac{1}{\delta 1}\frac{1-\delta_{sh}2}{\delta 2}$	99.6	100.4	99.2	0	0	0.7
sa2E2	pR1→a1R2→r2R3 sa2E2→sa2R3	$1-(1-s2_a\cdot E2_{sa2}\cdot u3_{sa2})^{\delta 1\delta 2}$	56 (0.81)	55	5.1	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{sa2}}{\delta 3}$	0.8	$t23_{sa2}-t23$	−0.8	$1-\frac{1}{\delta 1}\frac{1-\delta_{sa}2}{\delta 2}$	99.6	100.4	99.2	0	0	0.8
sh3E3	pR1→a1R2→a2R3→r3R4 sh3E3→sh3R4	$1-(1-s3_h\cdot E3_{sh3}\cdot u4_{sh3})^{\delta 1\delta 2\delta 3}$	100 (8.6)	100	$1.2\cdot 10^{-4}$	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{sh3}}{\delta 4}$	0.08	$t34_{sh3}-t34$	−1.0	$1-\frac{1}{\delta 1\delta 2}\cdot\frac{1-\delta_{sh}3}{\delta 3}$	100.0	100.0	99.9	0	0	1.0
sa3E3	pR1→a1R2→a2R3→r3R4 sa3E3→sa3R4	$1-(1-s3_a\cdot E3_{sa3}\cdot u4_{sa3})^{\delta 1\delta 2\delta 3}$	100 (7.3)	100	$2.2\cdot 10^{-6}$	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{sa3}}{\delta 4}$	0.09	$t34_{sa3}-t34$	−1.2	$1-\frac{1}{\delta 1\delta 2}\cdot\frac{1-\delta_{sa}3}{\delta 3}$	100.0	100.1	99.9	0	0	1.2
s2E22	pR1→a1R2→r2R3 s2E22→s2R2	$1-(1-c2\cdot E22_{s2}\cdot u2_{s2}\cdot c2_{s2})^{\delta 1\delta 2}$	9.3 (0.097)	9.2	3.4	$\frac{1}{\delta 1\delta 2}\max\Big(1, [\delta 2_{s2},\delta 2_{sh2}]\cdot[r2,1]\Big)$	[0.8, 1.2]	$\Big[[t22_{s2\downarrow},t22_{sh2\downarrow}]+t23_{\downarrow}-t23,0\Big]$	[−0.8, 0]	$1-\frac{1}{\delta 1}\cdot\frac{[1,1-\delta_{sh}2]}{\delta 2}$	[99.2, 99.6]	[100.0, 100.8]	[98.8, 99.2]	0	0	[0, 0.8]
sa2E22	pR1→a1R2→r2R3 sa2E22→sa2R2	$1-(1-s2_a\cdot E22_{sa2}\cdot u2_{sa2}\cdot c2_{sa2})^{\delta 1\delta 2}$	0.54	0.53	0.33	$\frac{1}{\delta 1\delta 2}\max(1, \delta 2_{sa2}[r2,1])$	[0.8, 1.7]	$[t22_{sa2\downarrow}+t23_{\downarrow}-t23,0]$	[−1.0, 0]	$1-\frac{1}{\delta 1}\frac{1-\delta_{sa}2}{\delta 2}$	99.6	[100.5, 101.3]	[98.3, 99.2]	0	0	[0, 1.0]

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
Ef	pEf→pR1	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\frac{\mathrm{bf}}{\mathrm{u1}}\mathrm{ef}\cdot\mathrm{Ef}\cdot\mathrm{u1}_{\varepsilon\mathrm{f}}$	0.043	0.043	0.043	$\frac{\delta 1_{\varepsilon\mathrm{f}}}{\delta 1}[\mathrm{r1_f},1]$	[50, 100.0]	$[tf1_{\varepsilon f\uparrow}+t12_{\downarrow},tf1_{\varepsilon f\downarrow}+t12_{\uparrow}]-tf1-t12$	[−0.8, 3.1]							
ΛEf	pΛEf→pR1	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\frac{\mathrm{bf}}{\mathrm{u1}}\mathrm{if}\cdot\Lambda\mathrm{f}\cdot\mathrm{E}\Lambda\mathrm{f}\cdot\mathrm{u1}_{\Lambda\mathrm{f}}$	0.035	0.035	0.035	$\left[\frac{1}{\delta 1},\frac{\delta 1_{\varepsilon f}}{\delta 1}\right]$	[6.7, 100.0]	$\left[[tf1_{\varepsilon f\uparrow},tf1_{\Lambda f\uparrow}]+t12_{\downarrow},[tf1_{\varepsilon f\downarrow},tf1_{\Lambda f\downarrow}]+t12_{\uparrow}\right]-tf1-t12$	[−0.8, 11.1]							
BEf	pBEf→afTf↘pR1	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\frac{\mathrm{bf}}{\mathrm{u1}}\mathrm{if}\cdot\mathrm{Bf}\cdot\mathrm{EBf}\cdot\mathrm{u1}_{\mathrm{Bf}}$	0.14	0.14	0.14	$\left[\frac{1}{\delta 1},\frac{\delta 1_{\varepsilon f}}{\delta 1}\right]$	[6.7, 100.0]	$\left[[tf1_{\varepsilon f\uparrow},tf1_{\mathrm{Bf}\uparrow}]+t12_{\downarrow},[tf1_{\varepsilon f\downarrow},tf1_{\mathrm{Bf}\downarrow}]+t12_{\uparrow}\right]-tf1-t12$	[−0.8, 11.1]							
p3	pR3	$\frac{\mathrm{CE_{k3}}}{\mathrm{CE_{k1}}}\frac{\mathrm{u3_p}}{\mathrm{u1}}$	2.9	$5\cdot10^{-14}$	0.30	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}}}{\delta 3}$	0.80	$tk3-tk1-t12-t23$	−6.2							
γ12	pRγ1→a1R2↘γP2→pR3	$\gamma c1\cdot\gamma e1\cdot\gamma t12\cdot\mathrm{QE_2}\cdot\mathrm{CE_{23}}\cdot\mathrm{u3_{p2}}$	0.017	0.016	$1.2\cdot10^{-3}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p2}}}{\delta 3}$	0.8	$t\gamma 12-t12+t23_{\mathrm{p2}}-t23$	−3.7	1	100.0	100.8	99.2	0	0	3.7
γ23	pR1→a1Rγ2→a2R3↘γP3→pR4	$1-(1-\gamma c2\cdot\gamma e2\cdot\gamma t23\cdot\mathrm{QE_3}\cdot\mathrm{CE_{34}}\cdot\mathrm{u4_{p3}})^{\delta 1}$	0.18	0.18	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{\mathrm{p3}}}{\delta 4}$	0.08	$t\gamma 23-t23+t34_{\mathrm{p3}}-t34$	−1.9	1	100.0	100.1	99.9	0	0	1.9
γ13	pRγ1→a1R2↘γP3→pR4	$\gamma c1\cdot\gamma e1\cdot\gamma t13\cdot\mathrm{QE_3}\cdot\mathrm{CE_{34}}\cdot\mathrm{u4_{p3}}$	$4.6\cdot10^{-3}$	$4.5\cdot10^{-3}$	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{\mathrm{p3}}}{\delta 4}$	0.08	$t\gamma 12-t12+t\gamma 23-t23+t34_{\mathrm{p3}}-t34$	−6.2	1	100.0	100.1	99.9	0	0	6.2
γ1L2	γP1→p1L2→p1R2	$\mathcal{T}\frac{\mathrm{A_1}}{\mathrm{A_k}}\frac{\mathrm{QE_1}}{\mathrm{QE_k}}\frac{\mathrm{CE_{12}}}{\mathrm{CE_{k1}}}\cdot\frac{\mathrm{b2_{p1}}}{\mathrm{u1}}\mathrm{e2_{p1}L2_{p1}}\cdot\mathrm{u2_{p1\varepsilon 2}}$	$5.6\cdot10^{-3}$	$4.1\cdot10^{-3}$	$5.3\cdot10^{-3}$	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1\varepsilon 2}}}{\delta 2}$	6.6	$t\gamma k1-tk1+t12_{\mathrm{p1}}-t12+2t21_{\mathrm{p1\varepsilon 2}}$	−13.2							
E1E2	pE1→pE2→pR3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2_{p\varepsilon 1}}\cdot\mathrm{e2_{p\varepsilon 1}}\cdot\mathrm{E2_{p\varepsilon 1}}\cdot\mathrm{u3_{p\varepsilon 1\varepsilon 2}}$	0.061	$1\cdot10^{-15}$	$6.2\cdot10^{-3}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p\varepsilon 1\varepsilon 2}}}{\delta 3}$	0.80	$t12_{\mathrm{p\varepsilon 1}}-t12+t23_{\mathrm{p\varepsilon 1\varepsilon 2}}-t23$	−5.2							
E1L2R2	pE1→pL2→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2_{p\varepsilon 1}}\cdot\mathrm{e2_{p\varepsilon 1}}\cdot\mathrm{L2_{p\varepsilon 1}}\cdot\mathrm{A2_{p\varepsilon 1\varepsilon 2}}\cdot\mathrm{u2_{p\varepsilon 1\varepsilon 2}}$	0	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p\varepsilon 1\varepsilon 2}}}{\delta 2}$	8.1	$t12_{\mathrm{p\varepsilon 1}}-t12+2t21_{\mathrm{p\varepsilon 1\varepsilon 2}}$	−0.9							

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
E2E3	pR1→r1R2 ↘ s1E2→s1E3→s1R4	$1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{E2}\cdot\mathrm{b3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{e3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{E3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{u4}_{\mathrm{s1}\varepsilon 2\varepsilon 3})^{\delta 1}$	0.41	0.41	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{\mathrm{s1}\varepsilon 2\varepsilon 3}}{\delta 4}$	0.11	$t23_{\mathrm{s1}\varepsilon 2}-t23+t34_{\mathrm{s1}\varepsilon 2\varepsilon 3}-t34$	−4.1	$1-\frac{1}{\delta 1}$	93.3	93.4	99.9	0	0	4.1
E1E3	pE1→pR2→r2R3 ↘ s2E3→s2R4	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{u2}_{\mathrm{p}\varepsilon 1}\cdot\left(1-(1-\mathrm{b3}\cdot\mathrm{e3}\cdot\mathrm{E3}\cdot\mathrm{u4}_{\mathrm{s2}\varepsilon 3})^{\delta 2_{\mathrm{p}\varepsilon 1}}\right)$	0.39 (0.0042)	0.30	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{\mathrm{s2}\varepsilon 3}}{\delta 4}$	0.11	$t12_{\mathrm{p}\varepsilon 1}-t12+t34_{\mathrm{s2}\varepsilon 3}-t34$	−5.5	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1}}{\delta 2}-\frac{1}{\delta 1\delta 2}$	7.4	7.5	98.6	$t12_{\mathrm{p}\varepsilon 1}-t12$	−3.2	2.3
E1L3	pE1→pR2→r2R3 ↘ s2L3→s2R3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{u2}_{\varepsilon 1}\cdot\left(1-(1-\mathrm{b3}\cdot\mathrm{e3}\cdot\mathrm{L3}\cdot\mathrm{u3}_{\varepsilon 3})^{\delta 2_{\mathrm{p}\varepsilon 1}}\right)$	0.32 (0.0034)	0.25	0.026	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1}}{\delta 2}-\frac{1}{\delta 1\delta 2}$	7.4	$t12_{\mathrm{p}\varepsilon 1}-t12$	−3.2	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\varepsilon 3}}{\delta 3}$	0.77	8.1	9.4	$t12_{\mathrm{p}\varepsilon 1}-t12+2\,t32_{\varepsilon 3}$	2.2	5.4
E11E1	pE11→pE1→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E11}\cdot\mathrm{b1}_{\varepsilon 1}\cdot\mathrm{e1}_{\varepsilon 1}\cdot\mathrm{E1}_{\varepsilon 1}\cdot\mathrm{u2}_{\mathrm{p}\varepsilon 1^2}$	$9.7\cdot10^{-3}$	$7.9\cdot10^{-3}$	$9.4\cdot10^{-3}$	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1^2}}{\delta 2}$	8.1	$[t12_{\mathrm{p}\varepsilon 1^2},t11_{\downarrow}+t12_{\mathrm{p}\varepsilon 1^2\downarrow}]-t12$	[−3.2, −3.0]							
L11E1	pL11→pE1→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L11}\cdot\mathrm{b1}_{\varepsilon 1}\cdot\mathrm{e1}_{\varepsilon 1}\cdot\mathrm{E1}_{\varepsilon 1}\cdot\mathrm{u2}_{\mathrm{p}\varepsilon 1^2}$	0.019	0.016	0.019	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1^2}}{\delta 2}$	8.1	$[t12_{\mathrm{p}\varepsilon 1^2},t11_{\uparrow}+t12_{\mathrm{p}\varepsilon 1^2\uparrow}]-t12$	[−3.2, −0.6]							
E22E2	pR1→r1R2 ↘ s1E22→s1E2→s1R3	$1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{E22}\cdot\mathrm{b2}_{\varepsilon 2}\cdot\mathrm{e2}_{\varepsilon 2}\cdot\mathrm{E2}_{\varepsilon 2}\cdot\mathrm{u3}_{\mathrm{s1}\varepsilon 2^2})^{\delta 1}$	$6.8\cdot10^{-3}$	$6.7\cdot10^{-3}$	$9.7\cdot10^{-4}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{s1}\varepsilon 2^2}}{\delta 3}$	0.9	$[t22_{\downarrow}+t23_{\mathrm{s1}\varepsilon 2^2\downarrow},t23_{\mathrm{s1}\varepsilon 2^2}]-t23$	[−2.0, −1.6]	$1-\frac{1}{\delta 1}$	93.3	94.2	99.1	0	0	[1.6, 2.0]
E33E3	pR1→a1R2→r2R3 ↙ s2E33→s2E3→s2R4	$1-(1-\mathrm{b3}\cdot\mathrm{e3}\cdot\mathrm{E33}\cdot\mathrm{b3}_{\varepsilon 3}\cdot\mathrm{e3}\varepsilon 3\cdot\mathrm{E3}_{\varepsilon 3}\cdot\mathrm{u4}_{\mathrm{s2}\varepsilon 3^2})^{\delta 1\delta 2}$	0.33	0.33	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{\mathrm{s2}\varepsilon 3^2}}{\delta 4}$	0.11	$[t34_{\mathrm{s2}\varepsilon 3^2},t33_{\downarrow}+t34_{\mathrm{s2}\varepsilon 3^2\downarrow}]-t34$	[−2.3, −2.2]	$1-\frac{1}{\delta 1\delta 2}$	99.2	99.3	99.9	0	0	[2.2, 2.3]
E1E22	pE1→pE22→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{E22}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{u2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$	$2.1\cdot10^{-3}$	$1.7\cdot10^{-3}$	$2.0\cdot10^{-3}$	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1\varepsilon 2}}{\delta 2}\cdot[\mathrm{r2},1]$	[4.1, 8.1]	$t12_{\mathrm{p}\varepsilon 1}-t12+[t22_{\mathrm{p}\varepsilon 1\varepsilon 2\downarrow}+t23_{\downarrow}-t23,0]$	[−5.1, −3.2]							
E1L22	pE1→pL22→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{L22}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{u2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$	0.010	$8.4\cdot10^{-3}$	0.010	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1\varepsilon 2}}{\delta 2}\cdot[\mathrm{r2},1]$	[4.1, 8.1]	$t12_{\mathrm{p}\varepsilon 1}-t12+[0,t22_{\mathrm{p}\varepsilon 1\varepsilon 2\uparrow}+t23_{\uparrow}-t23]$	[−3.2, −0.8]							

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
E2E33	pR1→r1R2 ↘ s1E2→s1E33→s1R3	$1-(1-b2\cdot e2\cdot E2\cdot b3_{s1\varepsilon2}\cdot e3_{s1\varepsilon2}\cdot E33_{s1\varepsilon2}\cdot u3_{s1\varepsilon2\varepsilon3})^{\delta1}$	0.032	0.032	$4.6\cdot10^{-3}$	$\frac{1}{\delta1\delta2}\frac{\delta3_{s1\varepsilon2\varepsilon3}}{\delta3}\cdot[r3,1]$	[0.43, 0.86]	$t23_{s1\varepsilon2}-t23+[t33_{s1\varepsilon2\varepsilon3\downarrow}+t34_{\downarrow}-t34,0]$	[−2.5, −1.6]	$1-\frac{1}{\delta1}$	93.3	[93.8, 94.2]	[99.1, 99.5]	0	0	[1.6, 2.5]
E2L33	pR1→r1R2 ↘ s1E2→s1L33→s1R3	$1-(1-b2\cdot e2\cdot E2\cdot b3_{s1\varepsilon2}\cdot e3_{s1\varepsilon2}\cdot L33_{s1\varepsilon2}\cdot u3_{s1\varepsilon2\varepsilon3})^{\delta1}$	0.11	0.11	0.015	$\frac{1}{\delta1\delta2}\frac{\delta3_{s1\varepsilon2\varepsilon3}}{\delta3}\cdot[r_c3_{\varepsilon2E\varepsilon3L},1]$	[0.62, 0.86]	$[t23_{s1\varepsilon2}-t23,0]$	[−1.6, 0]	$1-\frac{1}{\delta1}$	93.3	[94.0, 94.2]	[99.1, 99.3]	0	0	[0, 1.6]
					$1.8\cdot10^{-3}$	$1-\frac{1}{\delta1}$	93.3	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{s1\varepsilon2\varepsilon3}}{\delta3}\cdot[r3, r_c3_{\varepsilon2E\varepsilon3L}]$	[0.43, 0.62]	[93.8, 94.0]	[0.46, 0.66]	$[0,t23_{s1\varepsilon2}-t23+t33_{s1\varepsilon2\varepsilon3\uparrow}+t34_{\uparrow}-t34]$	[0, 1.3]	[0, 1.3]
sh1E1E2	pR1→r1R2 ↙ sh1E1→sh1E2→sh1R3	$1-(1-s1_h\cdot E1_{sh1}\cdot b2_{sh1}\cdot e2_{sh1}\cdot E2_{sh1}\cdot u3_{sh1\varepsilon2})^{\delta1}$	0.15	0.15	0.021	$\frac{1}{\delta1\delta2}\frac{\delta3_{sh1\varepsilon2}}{\delta3}$	0.86	$t12_{sh1}-t12+t23_{sh1\varepsilon2}-t23$	−2.7	$\frac{\delta_{sh}1}{\delta1}$	100.0	100.9	99.1	0	0	2.7
E1sh2E2	pE1→pR2→r2R3 ↙ sh2E2→sh2R3	$\frac{b1}{u1}e1\cdot E1\cdot u2_{p\varepsilon1}\cdot\left(1-(1-s2_h\cdot E2_{sh2}\cdot u3_{sh2})^{\delta2_{p\varepsilon1}}\right)$	0.18 (0.0019)	0.15	0.016	$\frac{1}{\delta1\delta2}\frac{\delta3_{sh2}}{\delta3}$	0.78	$t12_{p\varepsilon1}-t12+t23_{sh2}-t23$	−3.9	$\frac{1}{\delta1}\frac{\delta_{sh}2_{p\varepsilon1}}{\delta2}$	8.1	8.9	91.3	$t12_{p\varepsilon1}-t12$	−3.2	0.7
E2sh3E3	pR1→r1R2 ↘ s1E2→s1R3→r3R4 ↙ sh3E3→sh3R4	$\left(1-(1-b2\cdot e2\cdot E2\cdot u3_{s1\varepsilon2})^{\delta1}\right)\cdot\left(1-(1-s3_h\cdot E3_{sh3}\cdot u4_{sh3})^{\delta3_{s1\varepsilon2}}\right)$	0.98 (0.011)	0.97	0	$\frac{1}{\delta1\delta2\delta3}\frac{\delta4_{sh3}}{\delta4}$	0.08	$t23_{s1\varepsilon2}-t23+t34_{sh3}-t34$	−2.5	$1-\frac{1}{\delta1}$	93.3	93.4	99.9	0	0	2.5
					0					$\frac{1}{\delta1\delta2}\frac{\delta_{sh}3_{s1\varepsilon2}}{\delta3}$	0.9	0.9	91.0	$t23_{s1\varepsilon2}-t23$	−1.6	1.0
sh1E11E1	pR1→r1R2 ↙ sh1E11→sh1E1→sh1R2	$1-(1-s1_h\cdot E11_{sh1}\cdot b1_{sh1}\cdot e1_{sh1}\cdot E1_{sh1}\cdot u2_{sh1\varepsilon1})^{\delta1}$	$1.8\cdot10^{-3}$	$1.8\cdot10^{-3}$	$1.7\cdot10^{-3}$	$\frac{1}{\delta1}\frac{\delta2_{sh1\varepsilon1}}{\delta2}$	6.9	$[t12_{sh1\varepsilon1},t11_{sh1\downarrow}+t12_{sh1\varepsilon1\downarrow}]-t12$	[−1.1, −0.5]	$\frac{\delta_{sh}1}{\delta1}$	100.0	106.9	93.5	0	0	[0.5, 1.1]
s1L11E1	pR1→r1R2 ↘ s1L11→s1E1→s1R2	$1-(1-c1\cdot L11_{s1}\cdot b1_{s1}\cdot e1_{s1}\cdot E1_{s1}\cdot u2_{s1\varepsilon1})^{\delta1}$	0.14	0.14	0.13	$\frac{1}{\delta1}\frac{\delta2_{sh1\varepsilon1}}{\delta2}$	6.9	$[t12_{sh1\varepsilon1}-t12,0]$	[−1.1, 0]	$\frac{\delta_{sh}1}{\delta1}$	100.0	106.9	93.5	0	0	[0, 1.1]
					0.13	$1-\frac{1}{\delta1}$	93.3	0	0	$\frac{1}{\delta1}\frac{\delta2_{s1\varepsilon1}}{\delta2}$	6.7	[6.7, 100.0]	[6.7, 100.0]	$[t12_{\varepsilon1},t11_{s1\uparrow}+t12_{\varepsilon1\uparrow}]-t12$	[0.0, 5.8]	[0, 5.8]
					0.13	$\frac{\delta_{sh}1}{\delta1}$	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{sh1\varepsilon1}}{\delta2}$	6.9	106.9	6.5	$[0,t11_{sh1\uparrow}+t12_{sh1\varepsilon1\uparrow}-t12]$	[0, 3.4]	[0, 3.4]

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
E1sh2E22	pE1→pR2→r2R3 sh2E22→sh2R2	$\frac{b1}{u1}$e1·E1·u2$_{p\varepsilon 1}$ ·$\left(1-(1-s2_h\cdot E22_{sh2}\cdot u2_{sh2})^{\delta 2_{p\varepsilon 1}}\right)$	$1.1\cdot10^{-3}$	$8.9\cdot10^{-4}$	$4.1\cdot10^{-4}$	$\frac{1}{\delta 1}\frac{\max(1,\delta 2_{sh2})}{\delta 2}$ ·[r2,1]	[0.58, 1.2]	$t12_{p\varepsilon 1}-t12$ +[$t22_{sh2\downarrow}+t23_{\downarrow}$ $-t23,0$]	[−4.0, −3.2]	$\frac{1}{\delta 1}\frac{\delta_{sh}2_{p\varepsilon 1}}{\delta 2}$	8.1	[8.7, 9.3]	[87, 93.3]	$t12_{p\varepsilon 1}-t12$	−3.2	[0, 0.8]
E1γ23	pE1→pRγ2→a2R3 γP3→pR4	$\frac{b1}{u1}$e1·E1·u2$_{p\varepsilon 1}$ ·γc2$_{\varepsilon 1}$·γe2·γt23 ·QE$_3$·CE$_{34}$·u4$_{p3}$	$5.4\cdot10^{-4}$	$4.4\cdot10^{-4}$	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{p3}}{\delta 4}$	0.08	$t12_{p\varepsilon 1}-t12+t\gamma 23$ $-t23+t34_{p3}-t34$	−5.1	$\frac{1}{\delta 1}\frac{\delta 2_{p\varepsilon 1}}{\delta 2}$	8.1	8.2	99.0	$t12_{p\varepsilon 1}-t12$	−3.2	1.9
γgf	pRγg→agR1 γPf→pfTf	$\frac{CE_{kg}}{CE_{k1}}\frac{ug}{u1}$γcg·γeg ·γtgf·QE$_f$·CE$_{f1}$ ·u1$_{pf}$	0	0	0	-		-								
γf1	pRγf→afTf γP1→pR2	$\frac{CE_{kf}}{CE_{k1}}\frac{uf}{u1}$γcf·γef ·γtf1·QE$_1$·CE$_{12}$ ·u2$_{p1}$	$5.7\cdot10^{-3}$	$4.1\cdot10^{-3}$	$5.4\cdot10^{-3}$	$\frac{1}{\delta 1}\frac{\delta 2_{p1}}{\delta 2}$	6.6	$t\gamma f1-tf1$ $+t12_{p1}-t12$	−0.2							
pE$_\gamma$f1	pE$_\gamma$f→pR1 γP1→pR2	$\frac{CE_{kf}}{CE_{k1}}\frac{bf}{u1}$ef·Ef ·u1$_{\varepsilon f}$·γcf·γef·γtf1 ·QE$_1$·CE$_{12}$·u2$_{p1}$	$2.2\cdot10^{-5}$	$1.6\cdot10^{-5}$	$2.1\cdot10^{-5}$	$\frac{1}{\delta 1}\frac{\delta 2_{p1}}{\delta 2}$	6.6	$t\gamma f1-tf1$ $+t12_{p1}-t12$	−0.2	$\frac{\delta_{\varepsilon\gamma f}1}{\delta 1}$[r1$_f$,1]	[50, 100.0]	[57, 106.6]	[88, 93.8]	[$tf1_{\varepsilon\gamma f\uparrow}+t12_{\downarrow}$, $tf1_{\varepsilon\gamma f\downarrow}+t12_{\uparrow}$] $-tf1-t12$	[−0.8, 3.1]	[0, 3.4]
pRγ32	pRγ3→a3R4 γP2→pR3	$\frac{CE_{k3}}{CE_{k1}}\frac{u3_p}{u1}$γc3 ·γe3·γt32·QE$_2$ ·CE$_{23}$·u3$_{p2}$	$2.4\cdot10^{-4}$	$4\cdot10^{-18}$	$1.8\cdot10^{-5}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_p}{\delta 3}$	0.80	tk3−tk1−$t12$−$t23$	−6.2	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{p2}}{\delta 3}$	0.76	1.6	49	tk3−tk1−$t12$ $-t23+t\gamma 32$ $+t23_{p2}$	−2.9	3.3
E1E2E3	pE1→pE2→pE3→pR4	$\frac{b1}{u1}$e1·E1·b2$_{p\varepsilon 1}$ ·e2$_{p\varepsilon 1}$·E2$_{p\varepsilon 1}$ ·b3$_{p\varepsilon 1\varepsilon 2}$·e3$_{p\varepsilon 1\varepsilon 2}$ ·E3$_{p\varepsilon 1\varepsilon 2}$·u4$_{p\varepsilon 1\varepsilon 2\varepsilon 3}$	$1.7\cdot10^{-3}$	0	0	$\frac{1}{\delta 1\delta 2\delta 3}$ ·$\frac{\delta 4_{p\varepsilon 1\varepsilon 2\varepsilon 3}}{\delta 4}$	0.09	$t12_{p\varepsilon 1}-t12$ $+t23_{p\varepsilon 1\varepsilon 2}-t23$ $+t34_{p\varepsilon 1\varepsilon 2\varepsilon 3}-t34$	−8.1							
E1L2E1	pE1→pL2→pE1→pR2	$\frac{b1}{u1}$e1·E1·b2$_{p\varepsilon 1}$ ·e2$_{p\varepsilon 1}$·L2$_{p\varepsilon 1}$ ·A1$_{\varepsilon 1\varepsilon 2}$·b1$_{\varepsilon 1\varepsilon 2}$ ·e1$_{\varepsilon 1\varepsilon 2}$·E1$_{\varepsilon 1\varepsilon 2}$ ·u2$_{p\varepsilon 1\varepsilon 2\varepsilon 1}$	$2.2\cdot10^{-3}$	$1.8\cdot10^{-3}$	$2.1\cdot10^{-3}$	$\frac{1}{\delta 1}\frac{\delta 2_{p\varepsilon 1\varepsilon 2\varepsilon 1}}{\delta 2}$	8.1	$t12_{p\varepsilon 1}-t12$ $+t21_{p\varepsilon 1\varepsilon 2}$ $+t12_{p\varepsilon 1\varepsilon 2\varepsilon 1}$	−0.9							

Effect	**Topology**	$\boldsymbol{p_r}$	$\boldsymbol{p_r(N)}$	p_s	p_f	$\boldsymbol{q_{PP}}$		$\boldsymbol{T_{PP}}$		$\boldsymbol{q_{AP}}$		$\boldsymbol{\Sigma_q}$	$\boldsymbol{\eta_q}$	$\boldsymbol{T_{AP}}$		$\boldsymbol{d}$
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
E1L2E2	pE1→pL2→pE2→pR3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{L2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{A2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{E2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{u3}_{\mathrm{p}\varepsilon 1\varepsilon 2^2}$	0	0	0	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}\varepsilon 1\varepsilon 2^2}}{\delta 3}$	0.80	$t12_{\mathrm{p}\varepsilon 1}-t12$ $+2t21_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $+t23_{\mathrm{p}\varepsilon 1\varepsilon 2^2}-t23$	−2.9							
E1E2L33	pE1→pE2→pL33→pR3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{E2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{b3}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{e3}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{L33}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{u3}_{\mathrm{p}\varepsilon 1\varepsilon 2\varepsilon 3}$	$5.3\cdot10^{-4}$	$9\cdot10^{-18}$	$5.4\cdot10^{-5}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}\varepsilon 1\varepsilon 2\varepsilon 3}}{\delta 3}$ $\cdot[\mathrm{r}3,1]$	[0.40, 0.80]	$t12_{\mathrm{p}\varepsilon 1}-t12$ $+t23_{\mathrm{p}\varepsilon 1\varepsilon 2}-t23$ $+[0,t33_{\mathrm{p}\varepsilon 1\varepsilon 2\varepsilon 3\uparrow}$ $+t34_{\uparrow}-t34]$	[−5.2, −2.6]							
E1L22E2	pE1→pL22→pE2→pR3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{L22}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{E2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{u3}_{\mathrm{p}\varepsilon 1\varepsilon 2^2}$	$2.5\cdot10^{-4}$	$4\cdot10^{-18}$	$2.5\cdot10^{-5}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}\varepsilon 1\varepsilon 2^2}}{\delta 3}$	0.80	$t12_{\mathrm{p}\varepsilon 1}-t12$ $+[t23_{\mathrm{p}\varepsilon 1\varepsilon 2^2},$ $t22_{\mathrm{p}\varepsilon 1\varepsilon 2\uparrow}$ $+t23_{\mathrm{p}\varepsilon 1\varepsilon 2^2\uparrow}]-t23$	[−5.2, −4.0]							
L11E1E2	pL11→pE1→pE2→pR3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L11}\cdot\mathrm{b1}_{\varepsilon 1}$ $\cdot\mathrm{e1}_{\varepsilon 1}\cdot\mathrm{E1}_{\varepsilon 1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1^2}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1^2}\cdot\mathrm{E2}_{\mathrm{p}\varepsilon 1^2}$ $\cdot\mathrm{u3}_{\mathrm{p}\varepsilon 1^2\varepsilon 2}$	$4.7\cdot10^{-4}$	$8\cdot10^{-18}$	$4.7\cdot10^{-5}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}\varepsilon 1^2\varepsilon 2}}{\delta 3}$	0.80	$[t12_{\mathrm{p}\varepsilon 1^2},t11_{\uparrow}$ $+t12_{\mathrm{p}\varepsilon 1^2\uparrow}]-t12$ $+t23_{\mathrm{p}\varepsilon 1^2\varepsilon 2}-t23$	[−5.2, −2.7]							
L11E1-L22E2	pL11→pE1→pL22→ →pE2→pR3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L11}\cdot\mathrm{b1}_{\varepsilon 1}$ $\cdot\mathrm{e1}_{\varepsilon 1}\cdot\mathrm{E1}_{\varepsilon 1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1^2}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1^2}\cdot\mathrm{L22}_{\mathrm{p}\varepsilon 1^2}$ $\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1^2\varepsilon 2}\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1^2\varepsilon 2}$ $\cdot\mathrm{E2}_{\mathrm{p}\varepsilon 1^2\varepsilon 2}$ $\cdot\mathrm{u3}_{\mathrm{p}\varepsilon 1^2\varepsilon 2^2}$	$1.9\cdot10^{-6}$	$3\cdot10^{-20}$	$1.9\cdot10^{-7}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}\varepsilon 1^2\varepsilon 2^2}}{\delta 3}$	0.80	$[t12_{\mathrm{p}\varepsilon 1^2},t11_{\uparrow}$ $+t12_{\mathrm{p}\varepsilon 1^2\uparrow}]-t12$ $+[t23_{\mathrm{p}\varepsilon 1^2\varepsilon 2^2},$ $t22_{\mathrm{p}\varepsilon 1^2\varepsilon 2\uparrow}$ $+t23_{\mathrm{p}\varepsilon 1^2\varepsilon 2^2\uparrow}]$ $-t23$	[−5.2, −1.5]							

$\gamma 1$

A γ1 pulse (fig. 2.39b), also sometimes called "prepulse" in the literature [275, 336], occurs when the photon passes the semitransparent photocathode, hits d1 and emits a photoelectron there, since d1 is coated with a material with low photoemission threshold W_{ph} as well. The pulse is early by mainly the difference of tk1 and $t\gamma$k1 and the charge is reduced by a factor between 10 and 20 vs. a RP according to [170; p. 438], which mostly comes from skipping δ1.
This effect was first reported by [331], depends on visible d1 size [129], incident angle and wavelength (the QE curve of the dynode material, here Cs_3Sb) and due to its small charge is only observed if the instrumental threshold is low enough [170; pp. 74, 224, 435–439]. In [336; p. 692] a formula for p_s is presented, which already includes $\mathcal{T}$, ${}^{A_1}/_{A_k}$, ${}^{QE_1}/_{QE_k}$ and threshold losses; only the CE and u terms are missing, and the charge distribution was assumed as an exponential decay instead of a Gaussian.

$\gamma 2$, $\gamma 3$

Looking at fig. 2.43b, γ2 and even γ3 are possible within some small solid angle. These effects occur even some nanoseconds earlier than γ1, but also have far smaller charges since one resp. two gain steps are skipped. The mean charge therefore lies far below h_s and since only one pulse is produced none of both effects can be observed (see fig. 2.38). In addition, $QE_2 = QE_3 < QE_1$, and the very small solid angle would further reduce p_r.

γg

Photons which hit the aluminized and photosensitive sidewall can produce photoelectrons (fig. 2.39b) [170]. Depending on position, the transit time to d1 can be shorter or longer than tk1, resulting in an EP or LP. Since the focusing optics is not optimized for these trajectories, suboptimal hit positions on d1 should reduce δ1 due to edge effect, although the traversed voltage difference is identical to U_{k1}, because g is at k potential. As discussed for the EPP-CTD, the maximum earliness could be far larger, with charge dropping towards earlier times.

γm

When the deck of the dynode mount is hit, photoelectrons can be emitted with a certain probability (fig. 2.39b), since during cathode evaporation, the deck also is covered in photosensitive material. Since m is at d1 potential, the pm is accelerated by the grid potential only and decelerated towards d1 after passing f, before it either hits d1 with its emission energy and is backscattered (SE creation is almost impossible due to the small impact energy) *or* is directly drawn to d2 (not calculated), which would further increase the earliness. The charge for both variants is the same: In both cases one photoelectron starts at d1 potential, traverses U_{12} and hits d2. Because tk1 is substituted by $t\gamma$km, the pulse initially has a large earliness, but tm1 is large due to the small voltage gradients along the pm path to f and then d1. The variation of the pm starting position and the resulting trajectory length, and the small pm velocities are the reason for the large spread in T. p_c drops with rising earliness since the origin region moves closer to the gap to d1 and the contributing area decreases (see p. 266).

[129] ${}^{A_1}/_{A_k}$ was determined here for the 8" R5912 to 1.2%, Wright gives a value of $\approx$ 1% for a 5" PMT [170; p. 224].

The physics of this effect was deduced from the existence of the cluster in the CTD, which could neither be explained by electronic EPs (the effect must be from one electron hitting d2 with energy $\approx U_{12}$ to explain the charge, but is earlier than E1; earlier electronic EPs have too small charge) nor electronic secondary delays of γ1 (effect can only be later than γ1L2 with additional interactions, but the observed probability is already larger than for the γ1L2 cluster). This left only photon hits on structures with a potential similar to d1 with some subsequent delay as possible causes: m and f. Since f traps all photoelectrons, the effect could only be from γm and subsequent calculations revealed the delay through tm1.

[γf]

For PMTs with accelerator-type focusing grids, no pf can leave the grid to d1 and no pulse results (fig. 2.39b). p_r would actually be observable (0.038%), so the effect occurs, although without being detectable.
This is different for grids with potential between k and d1, where γf should produce pulses with large earliness, but later than γ1, since a part of the distance to d1 is covered by a pf instead of a photon. The charge will depend greatly on the potential of f and could range from $\frac{1}{\delta 1}$ to values *above* 100% pe for large U_{f1} – the grid then acts as an additional gain step, and due to the concave shape of SEY(E), the total gain is increased by splitting up the same voltage onto several steps.

All following effects are "electronic", that means electron-based (photoelectron or secondary electron).

Interdynode scattering

d1-based

E1

E1 pulses – often called "early pulses" in literature – occur when a pk is elastically scattered on d1 before it can produce secondary electrons, and its new trajectory ends on d2, where it starts a cascade (fig. 2.40c) [275, 301] [170; pp. 435–439]. Since the scattered pk continues with undiminished momentum, it crosses x_{12} faster than the s1 electrons normally produced on d1 (U_0s1$\approx$5 eV) and starts the cascade on d2 early, which shifts the complete pulse to earlier transit times. The pulse is smaller due to the missed gain from δ1, which is slightly compensated through the increased impact energy on d2 ($^{\delta 2_{\mathrm{p}\varepsilon 1}}/_{\delta 2} = 1.22$)
Scattering of pk from d1 onto d2 is only the case for a small fraction of events, in the order of 1% of pulses. First, the backscattering probability b1 is "only" about 30%, and the elastic scattering fraction is expected to lie around 25%. In addition, considering pe momentum and field distributions, the pk paths ends on d2 only for a small emission solid angle from d1: A fraction of all backscattered pk turn to E1 (34.2%), another produces L1 (16.4%), others create E11 (5%) or L11 (10%), and in all other cases (34.4%) the pk hits inactive structures and is lost, which is the main reason that the multiplication efficiency (ME) of a PMT is smaller than one.
Because of their small charge, E1 are strongly suppressed for typical h_s of around 25% pe, which is good since they affect time resolution strongly. However, due to the gain fluctuations, which are stronger than for the RP since $\delta 2_{\mathrm{p}\varepsilon 1} < \delta 1$, still a large fraction of E1 manages to exceed the threshold. This is visible in the large charge

spread in fig. 2.45. The suppression can be improved by raising the threshold, but this comes at the cost of PDE, which normally is not an option. Therefore, in spite of large $\delta 1$ values and high h_s, E1 routinely poses a problem for event reconstruction and must be characterized for all employed PMTs.

The distribution in the TTD usually can be well described by a normal distribution. The fitted standard deviation is larger than for the main peak due to the statistical smearing in T from the reduced first gain.

E1 probabilities for large PMTs found in literature range from 0.05% (R12860, $g = 10^7$, $\delta 1$ and threshold unknown [307]) to 0.75% (ET 9351, $g = 2 \cdot 10^7$, $\delta 1 = 5.7$, threshold 0.2 pe [336]); the values strongly depend on HV, $\delta 1$ and the chosen threshold. The measurements performed here for the R5912 blend in well with these results, if a classic pileup-agnostic analysis is used (0.69%, 4.5% pe) as for the measurements in literature. If the advanced analysis developed for this work is employed, the value is far higher (2.54%) since the threshold is lower.

ΛE1

This is the same process as E1, but with inelastic pk scattering on d1 without s1 production (as E1 in fig. 2.40c; the x_{12} trajectory can be different due to the reduced energy). Since the scattered pk energy is smaller than for E1, $\delta 2$ is smaller (but at least 1, since the pk still has to reach d2 to observe a pulse) and $t12$ is larger, which reduces the earliness. Charge and time are coupled through the backscattered energy: The larger the earliness, the larger the charge.
This effect was predicted by [301], where a group of pulses with T between E1 and γk and charges between 30 and 90% pe were observed with a slight rising $q(T)$ correlation for the Hamamatsu R7081 (10”). In the publication the SER was analyzed for EPs in the range from -6 to -2.5 ns and showed two Gaussian functions, one around 15% pe and one centered around 70% pe. ΛE1 should not show this rising behavior and cannot explain such high charges. Likely the observations were either for γg / CTS with edge effect (as discussed for the CTD) or from unresolved pileup from a BE1 PP and its RP-AP.

BE1

BE1 is an E1 with SE production on d1 before pk scattering (fig. 2.40c); thus, the pk produces a PP on d2 with reduced charge and earliness compared to E1, but also an RP on d1 which becomes the AP and has lower charge than γk but unchanged $T_r = 0$.
Since the pk energy is split up on two pulses, their charge is correlated, and charge and time / delay of each pulse are connected ($q_{\mathrm{PP}}(T_{\mathrm{PP}})$ and $q_{\mathrm{AP}}(d)$).
Judging from the nearly flat energy distribution of inelastically backscattered primaries in fig. 2.12a, $p_c(T)$ should be quite evenly distributed in time. However, this can be modified by kinematics. For BE1, larger emission energies lead to increased loss likelihood since the focusing onto d2 is less effective [232]. Thus in sum an increase of events to smaller earliness is expected and this could be one of the causes of the cluster near $T_r = 0$ around 10% pe. Similar arguments apply to BL1 (see there), other BEn and BLn, and Λ effects.

d2-based

E2, ΛE2 and BE2 are equivalent to the effects on d1 (fig. 2.42c), but instead of one pk now all s1 electrons can contribute, which greatly raises p_r. On the other hand, one more amplification step is skipped, which lowers p_s and thus limits p_c. In sum this results in similar probabilities as for the d1-effects *despite* all charges lying below h_s and without considering the increased charge spread compared to γk from the lower δ. The earliness is smaller than for E1 since the backscattered energy is lower ($U_{12} < U_{k1}$). Since the earliness is likely underestimated and the charge distributions will be broad, these effects might explain the cluster in the CTD around −3.5 ns and 2% pe.
All three effects produce RP-afterpulses, since the continuation of the RP which was started on d1 is unaffected by the loss of one s1. This is different from E1 and ΛE1, which do not produce APs, since no SEs create an RP. In the case of BE2 there is a slight correlation between PP and AP charge, because the s2 electrons produced by the scattered s1 merge with the regular cascade.

d3-based

Interdynode forward scattering on d3 occurs even more frequently due to the abundance of s2 electrons (p_r = 76% to 100%), but the pulses are too small for observation ($p_f = 0$; $p_s = p_r$ due to the presence of the RP-AP). Instead, they are expected to influence the average pulse shape since they are present in the majority of pulses (N = 1.4 to 5.7).

Intradynode scattering

d1-based

E11

For E11 a pk is scattered off d1 and then hits the same dynode at a position closer to d2, where it produces SEs (fig. 2.40e). The distance to d2 is therefore shorter for the s1, which lowers $t12$. On the other hand, the extra intradynode transit time of the pk has to be considered. In sum an EP results, but the re-hit position on d1 depends on the emission angle, so the earliness can be as low as zero, when the pk hits d1 close to its original impact site without notable $t11$. The charge also is expected to depend on hit position due to the edge effect, which should results in a rising $q(T)$ dependence, if the first impact location is assumed to be in the optimized region with high $\delta 1$. Since no dynode is skipped, the charge is similar to γk with the mentioned reduction through edge effect.

ΛE11

ΛE11 in addition can have SE-less losses at the first d1 hit. A reduced pk re-emission energy increases $t11$ and can even lead to lateness. The $q(T)$ dependence is falling for the Λ-scattering, but rising for the variation with emission angle which determines the second hit position. Both effects are mostly independent (a smaller emission energy, however, limits the range on d1), so $q(T)$ is not a curve but an area, even without statistic smearing. In sum, the energy dependence is estimated to dominate over the influence from the emission angle.

BE11

BE11 is similar to ΛE11 but even more complicated, since an RP-AP is produced (fig. 2.40e). Thus when the energy of the scattered pk falls below the point where T_r becomes > 0 (crossover energy), the PP and AP switch places and the RP now is the PP. In the TTD, the BE11 curve stops at $T_r = 0$, since after this point the RP is the PP. The crossover leads to *two* different curves in the CDD (sub-effects), both expected to show predominantly falling behavior in $q(d)$: The RP-AP behaves inverse to ΛE11 since a smaller BE11 pulse creates more SE (since the AP charge rises with falling PP charge and thus earliness, the AP charge falls with falling delay). The BE11-AP behaves like a ΛE11 and has larger T_r (and thus d) for smaller charge. This behavior is similar in the FDD.
For BE11 a rather flat $p(T)$ distribution is expected for medium and large charges, since the emission energy distribution is flat and the emission energy is large enough to cross diverting fields. For small energies the latter is not the case anymore, especially for large intradynode distances, and the pk will be drawn to d2 instead. At these energies, the scattered pk produces the AP whose charge thus is reduced due to stage skipping, which can lead to a probability drop towards large delays due to reduced detection. In addition, the Lambertian emission angle distribution will suppress short intradynode drift times and affect the $p(T)$ and $p(d)$ distributions as a secondary effect.

d2-based

E22, ΛE22 and BE22 are similar to the effects on d1, but the maximum earliness is larger due to the layout of d2 and d3, which are directly adjacent. Therefore even for the smallest possible re-emission energy $T_r < 0$, so no crossover is observed for BE22. Since the solid angles for a re-hit of the same dynode are smaller for E22 than E11, and the impact energy is lower which reduces b2, the p_c are similar to the effects on d1 although far more particles can contribute.

d3-based

For ΛE33 and BE33 lateness can occur again since the dynode distances are larger, which causes crossover and two sub-effects for both interactions. However, the sub-effects where the scattered pulse is the AP cannot be observed, since the charge is too low for h_f. The p_r are quite high (11 to 38%), but the small charge reduces p_c due to p_f for the observable (sub)effects almost to d1 intradynode scattering levels.

Scattered stage skipping

Instead of interdynode scattering to the next stage, within a small emission solid angle the stage after next or even beyond that can be hit (effects not calculated). This creates an even larger earliness than E-effects, but with further reduced charge through additional skipped gain steps. En(n+2) is similar to EnE(n+1) in charge since the elastic scattering losses εn are negligible, and occurs even a little earlier than EnE(n+1) since the traversed path is shorter.
Only one-stage skipping is relevant, because two or more skipped stages reduce the charge too much for observability already for E14 (charge similar to E1E2E3 = 0.09% pe). This leaves the following relevant effects:
E13 produces only a PP which is comparable in charge to p3 and E1E2 (0.80% pe) and even a little earlier than E1E2 (-5.2 ns) due to its shorter path, but a little later

than p3 (-6.2 ns) due to the extra path in the E1 scattering. This already sets quite narrow limits for T_r. p3 and E1E2 are already undetectable due to their small charge, so E13 will be as well.
E24 creates a stage skipping EP and an RP from the rest of the s1. The EP is even smaller than E13 and can be neglected (charge comparable to E2E3 = 0.11% pe), but its s1 is missing in the RP cascade, which reduces the RP charge slightly by $1/\delta 1 \delta 2$. Since these are the most likely effects for scattered stage skipping and both are not observable, the omission from the calculations is justified.

Secondary electron interdynode emission

Just as a particle from a previous stage can be forward *scattered* from dn onto d(n+1), the different types of electrons originating from dn can be *emitted* in forward direction.

d1-based – (s1E1), sh1E1, sa1E1

s1E1 denotes an s1 emitted in direction of d2 – which is just a label for the s1 electrons produced in an RP and as such by definition is indistinguishable from an RP and unobservable.
If instead an sh is emitted towards d2 (fig. 2.42a), its higher initial energy of about 25 eV and shorter $t12$ allows to distinguish between the pulse created by sh1 impact on d2 and the subsequent RP. With $T_r = -1.1$ ns the head start is not much, but the probability with $p_c = 10\%$ is very high due to the abundance of SEs created on d1, each of which has a chance of about 1.1% to have an energy above 25 eV. For larger energies the earliness increases further, but the expected frequency quickly drops. Accordingly, sh1E1 should actually cover a range in T with falling $q(T)$ and rising $p_c(T)$, and merge with s1E1 (i.e γk) at $T_r = 0$.
sa1E1 has even more energy (about 36 eV were assumed) and so the PP is a little earlier (-1.4 ns) and larger than sh1E1. With 9.0%, p_c is almost as high as for sh1E1, which makes it a significant EP effect on the left flank of the RP in the TTD.
For both sh1E1 and sa1E1 the relative charge spread should be larger than for γk and E1, since sh1E1 and sa1E1 have a lower first δ and the relative standard deviation (standard deviation divided by mean) of the Poisson amplification process scales with $1/\sqrt{\delta}$.

d2-based – (s2E2), sh2E2, sa2E2

s2E2 again is part of the RP and unobservable. sh2E2 and sa2E2 are not as early as their d1 counterparts due to the shorter interdynode distance and larger potential difference (for the R5912 $U_{23} > U_{12}$), but occur even more frequently ($p_r = 56$ to 60%). Their lower charge, however, reduces their p_c to about 5%. Because the earliness and charge are small and N is already 0.91 resp. 0.81, these effects are on the verge to being a constant contribution to the pulse shape. It is questionable, whether they can actually be resolved that close to their far larger RPs (see p. 272).

d3-based – (s3E3), {sh3E3}, {sa3E3}

s3E3 is part of the RP. The pulses for sh and sa have larger earliness again due to the increased distance and lower fields compared to d2. With $N = 8.6$ and 7.3 they are a constant contribution in every RP and even in underamplified pulses. Their extremely small charge, however, suppresses p_c from 100% to the 10^{-6} level and below.

Secondary electron intradynode emission

As for intradynode scattering of electrons from earlier stages, SEs can be emitted from a dynode under such small angles that they hit the same dynode. Since SE emission is Lambertian, this probability is largely suppressed unless the dynode geometry favors the effect through large curvatures.
Since the emission energies are small compared to the kinetic energy of SEs moving between dynodes, the intradynode transit times are large, which can outweigh the earliness from a reduced interdynode distance in the next step. For the studied box-and-line geometry, SE intradynode scattering on d1 and d3 only results in lateness, so these effects are discussed in the LP section 2.3.6.4.

d2-based – s2E22, sh2E22, sa2E22

For s2E22 (as in fig. 2.42b but on d2) and sh2E22 a small maximum earliness of -0.6 resp. -0.8 ns results. Due to the small charge, the combined probability is reduced from $p_r = 9.3\%$ to $p_c = 3.4\%$. sa2E22 is earlier (T_r down to -1.0 ns) but p_c with 0.33% is far lower. Thus the majority of pulses overlaps with the statistical fluctuations of naturally occurring s2E2 and cannot be resolved.

Secondary electron stage skipping

Just as a pk can hit d3 or d2, stage skipping can also occur for SEs for certain emission positions, angles and energies.
Again, only one-stage skipping is analyzed due to the strong charge reduction for more skipped stages. For sxnEn(n+2) the charge is comparable to sxnEnE(n+1) since the εn are negligible, and the effects occur even earlier since the path is more direct.
s1E13, sh1E13, and sa1E13 are likely more probable to take place from k-near positions on d1 (as shown for BE11 in fig. 2.40e). The charge should be similar to s1E1E2 = E2 resp. sh1E1E2 and sa1E1E2 (all 0.86% pe). For s1E1E2, T_r will be somewhat earlier than E2 (-1.6 ns). sh1E13 and especially sa1E13 should be earlier than sh1E1E2 (-2.7 ns), and can maybe explain the cluster at -3.5 ns and 2% pe in the CTD better than a large earliness correction factor for E2. The probability for sh1E13 and sa1E13 will be lower than sh1E1 and sa1E1 ($p_c = 10$ and 9%). However, for s1E13 it could be quite high due to the many available s1 electrons.
s2E24, sh2E24, and sa2E24 maybe are possible for emission from the d1-near part, but this is difficult to assess without field simulations. In any case, the charges would be comparable to sx2E2E3 (sh2E2E3 = E3 has 0.11% pe, sh/a2E2E3 were not studied) and are too small for detection.

Grid

These effects can only be observed, if a focusing grid is present between k and d1.

Ef

When a pk is scattered on the grid, its emission vector can only point towards d1 (forward scattering) if the electron glances off the side of a wire; backscattering after a frontal hit will result in Lf. Since the particle vector is modified by the scattering, the pk is not focused onto the target region on d1 anymore, but can hit any position on d1 (with varying probability). This is similar to intradynode scattering and can

lead to earliness (shorter s1 path to d2) or lateness (longer path) and should also reduce the charge towards both extremes due to the edge effect. For Ef (fig. 2.40a) the maximum earliness (−0.8 ns) is smaller than for E1 (-3.2 ns), since only a part of x_{12} is traversed by a high-speed pk instead of an s1. The maximum lateness (3.1 ns) is larger than the highest earliness due to the far greater distance the s1 has to cover – the designated d1 hit region lies close to the E1 scattering site in fig. 2.40c.

ΛEf

If instead of elastic scattering Λ-scattering takes place (similar to Ef in fig. 2.40a, but with reduced energy), the lateness is increased due to the slower particle velocity. In addition, the charge is decreased through a lower $\delta 1$ on impact, which results in a falling $q(T)$ dependence.

BEf, {BEF-safR1}

For PMTs with accelerator-grids, BEf (fig. 2.40a) is comparable to ΛEf, since the sf it produces cannot leave the grid and only the pulse from the (quasi SE-less) scattered pk is observed. It is nevertheless an independent effect, which further contributes to the late flank of the main peak.
For the R5912, in case of saf emission (BEf-safR1) the initial energy could actually be large enough to overcome the potential barrier to d1. The resulting APs, however, only would occur for a small fraction of BEf events, have a charge of around $\frac{1}{\delta 1}$ as the d1 impact energy is only a few eV, and are expected to have rather large lateness values, since the saf cross x_{f1} with very low speed. Since the saf can also hit a large region on d1, depending on emission site, tf1 should vary greatly, so the detectable probability is spread out over a large time range. Because p_c is already only 0.14% for BEf, the observability of BEf-safR1 is questionable.

pk amplification not on d1

The effects of this group which produce EPs must result from pk stage skipping, since amplification on structures lying before d1 introduces extra distances, which are traversed by slow SEs, and results in lateness.

{p3}

For a small fraction of pk trajectories with slanted incidence angles (similar to γ2k in fig. 2.43b), the particle is not focused onto d1, but can hit d3 (γPk→pR3 = p3). This would produce a very early electronic EP with -6.2 ns, but the charge is greatly diminished by skipping two stages (0.80% pe), which reduces the probability from a large $p_t = 2.9\%$ to $p_s = 5 \cdot 10^{-16}$ since no other pulse is produced. Because $p_f = 0.30\%$ would not prevent detection, the observability strongly depends on h_s and the charge variance. $\delta 3_{\mathrm{pk}} = 9.87 < \delta 1 = 15.00$, so the relative charge distribution will be broader than for γk. In addition, noise spikes sitting on top of a large p3 might be raised above h_s. Therefore, p3 might be the origin of the few events in the region of −9 to −6 ns at 2% pe and below. Alternatively, these could originate from p2, or γm pulses for which the pm is directly focused onto d2 (γm-pmR2, skips amplification on d1).

For linear-focused PMTs, p3 is possible as well, but in this case the pk hits the *outside* of d3 (p3o, see fig. 2.15a. If any SEs are produced there at all (unlike the

inside, the surface of the outside likely has no large SEY), they cannot overcome the potential barrier to d2, let alone d1, and are subject to a3T3 and create no pulse.

p2

Similar to the rationale for p3, the effect p2 (γPk→pR2, not calculated) might be possible as well, although for a lower fraction of pk due to the smaller cross section of d2 compared to d3 for incident pk. On the other hand, the charges should be comparable to E1 (8.1% pe), maybe somewhat reduced by edge effect, which would make the effect well observable. p2 pulses should be slightly earlier than E1 due to the more direct path.

Luminescence-based

Any cascade can produce photons via e.g. bremsstrahlung as a side effect, so the listed effects can cause additional EPs for all pulses, which undergo amplification at the respective stage. Here only the standard effect γk is examined as the mother pulse.

γ12

A photon can traverse x_{12} much faster than the s1 electrons (fig. 2.43a), so the γ12 pulse is early (−3.7 ns). However, the p2 has a lower emission energy than an s2, so $t23_{\mathrm{p2}} > t23$, which reduces the earliness a bit. Due to skipping two gain steps, the charge is very low (0.80% pe), which lowers the $p_r = 0.017\%$ to $p_f = 1.2 \cdot 10^{-5}$.

{γ23}

γ23 (same emission site as γ21 in fig. 2.43b but towards d3) is not quite as early (−1.9 ns) since $x_{23} < x_{12}$ and $U_{23} > U_{12}$, and is unobservable due to its tiny charge.

{γ13}

γ13 would be quite early (−6.2 ns) thanks to skipping two electron transit times, but has the same, immeasurably low charge as γ23. The additional reflection, respectively that only the k-near part of d1 can cause a direct d3 hit (see fig. 2.43a) also reduce p_c strongly compared to γ12 and γ23.

Dual interactions

Incident photon

Reflection

When a photon passes the cathode without absorption, it can be reflected onto a reflective structure and hit a photosensitive structure (k, g, m, f, dn), where it produces a photoelectron. This is not limited to the device itself – the photon can also leave the cathode and hit *another* photosensor in a detector. Inside the PMT, the reflection can cause earliness or lateness, depending on what is hit afterwards. If the cathode is struck again, this introduces delays through the extra photon paths. If the photon hits a structure other than k, the extra photon path through the reflection does not affect T much and the effect will have a large earliness since tk1 is replaced

by far smaller terms. Three such effects are discussed exemplarily here. There will be more for triple, quadruple and higher-order interactions, especially for reflections on g, m and k, which still will have observable probabilities due to the high reflectance R and large areas of these structures.

γMg1 and γMm1

γMm1 and γMg1 (fig. 2.39d) produce pulses identical to γ1, save for the introduction of small delays through the extra paths. Both effects are only observable for slanted light incident at a large angle around $\approx 60°$ within a small angular phase space. Judging from the TTD in fig. 2.44, the probability of γMg1 is overestimated slightly, by roughly the same factor as γ1. Other, similar effects from multiple reflections on k and g will be present at slightly larger delays. These will likely overlap with γMg1 in their T distributions, which could require to lower the p_c attributed to γMg1.

{γM12}

Since the dynodes have a reflectivity of about 50%, a d1 hit can also be reflected onto d2, which should be quite likely since the shape of d2 favors this. Nevertheless, the charge reduction from gain skipping suppresses the pulse below h_s, and since no other pulse is produced the effect is unobservable.

Electronic interactions

The previously discussed EP effects can be chained to form dual interaction effects. For the large majority of effects, all relevant information is contained in table 2.2 and the plots of the EPP-TTD, EPP-CTD, AP-CTD, DD, and FDD. Therefore effects are only discussed in the text, if they are of special interest. Exemplarily, some dual interaction effects are shown in fig. 2.41b (E11E1) and 2.41c (L11E1).

γ1L2

The combination of a photonic stage skip with an electronic delay (fig. 2.39e) leads to an earliness which is shorter than obtainable with purely photonic effects (γ1, γMg1, ...) and larger than for electronic effects at the same charge (at γ1 level) alone. The effect was already discussed during the analysis of TTD (p. 266) and CTD (p. 272).

Interdynode scattering

{E1E2}

Electronic EPs can reach larger earliness than E1 (-3.2 ns) by chaining forward scattering processes, such as in E1E2 (fig. 2.41a, -5.2 ns) and E2E3 (-4.1 ns). However, despite the huge impact energy at d3, the charge loss through skipping two gain steps outweighs and suppresses p_r (0.061%) far below possible detection ($p_s = 10^{-17}$) for the R5912.

E1E2 was first postulated by the Borexino collaboration in [275] for the ET 9351 (8"), where the early flank of the main peak was attributed to a forward scattering chain of E1, E1E2, E1E2E3 and so forth. For the 9351, E1E2 and even E1E2E3 might actually be observable if the dynode material of d1 has regular gain instead

of high gain and $\delta 2$ and $\delta 3$ are small, since then the relative charge is reduced less. Since the 9351 has 12 dynodes and reaches $g = 2 \cdot 10^7$ the average δ is smaller than for the R5912. The possibility shall be estimated in the following based on the numbers given in Borexino papers [275, 336, 343].

In [336] [130], the assumption is made that all δn for n > 2 are in the linear regime of the SEY. In this case, the values are multiples of the resistance ratios: $\delta 2 = 2 \cdot \delta$n, $\delta 3 = 1.5 \cdot \delta$n, and δn for d4+. From this follow average values of $\delta 1 = 5.719$, $\delta 2 = 7.12$, $\delta 3 = 5.34$, and δn= 3.56. However, using this method the values for $\delta 1$ and $\delta 2$ are certainly overestimated, since they will definitely not be in the linear regime anymore. In fact, the $\delta 2$ from this calculation would be higher than $\delta 1$, although the impact energy on d2 is only about 26% of that on d1, which is not possible. For this reason it is necessary to reestimate the δn using the method derived here (see 2.3.5.4).
The resistance ratios of the VD are given in [343]. The voltage divider is front-tapered with an accelerator at d3 potential and 12 dynodes. U_{k1} is fixed to 600 V with three Zener diodes, the interstage resistances between d1 and a are: $\{2\mathrm{M}, 1.5\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M}, 1\mathrm{M} + 10\mathrm{k} + 33\}\,\Omega$. $g = 2 \cdot 10^7$ (effectively $1 \cdot 10^7$ due to decoupling of the signal [343]) and the typical HV was 1650 V. From this follow the interstage voltages $\mathrm{U_{k1}} \ldots \mathrm{U_{34}} = \{600, 155, 117, 78\}$V. With the Cs_3Sb SEY from [169, 184] the δn are $\delta 1 = 8.89$, $\delta 2 = 6.75$, $\delta 3 = 5.74$, and $\delta 4+ = 4.16$, which results in $g = 1.28 \cdot 10^8$, which is too high. Assuming a regular gain for d1 (best case for E1E2 observability) a uniform correction factor for all dynodes of 0.856 results, leading to the corrected gains $\delta 1 = 7.61$, $\delta 2 = 5.78$, $\delta 3 = 4.92$, and $\delta 4+ = 3.56$.
Using the charge formulas given in table 2.2 for E1,E1E2 and E1E2E3, the relative charge results in $q_{\mathrm{E1}} = 16.8\%$ pe, $q_{\mathrm{E1E2}} = 3.3\%$ pe, and $q_{\mathrm{E1E2E3}} = 0.90\%$ pe, which all are far larger than for the R5912 (8.1% pe, 0.80% pe, 0.09% pe).
However, the threshold of 16% pe employed in [336] was far higher, which requires to also calculate p_s. From the SER fit in [336], $\sigma_{\mathrm{RP}}/\mathrm{q_{RP}} = {}^{1.16\bar{\mathrm{q}}}/_{0.52\bar{\mathrm{q}}}$ is retrieved, which was used in eq. (2.157). From this follows that already for E1 only 54.8% of events are detectable; E1E2 and E1E2E3 are completely unobservable with an instrumental threshold of 16% pe, even using the assumption of a low-gain d1.
Thus the observation of E1E2 and E1E2E3 in [275] and [336] is refuted. Other effects have to be the cause for the measured events, e.g. γm, sx1E13 and γ1L2.
As a side-note, it likely *would* be possible to measure E1E2 for the 9351 with the shielded setup constructed in this work (SPAX [131], see p. 460), assuming a threshold h_s of about 4.5% pe can be reached for this PMT as well and that d1 really is of the low-gain variant. E1E2E3 likely is too far below reachable h_s to be observable at all, since only one pulse is produced.

Secondary electron interdynode emission

sh1E1E2, sa1E1E2

These effects are equivalent to E2 (=s1E1E2) but for sh1 or sa1 (not calculated) instead of s1. Since U_0 is higher, T is notably earlier (sh1E1E2: −2.7 ns) than E2 (−1.6 ns). The earliness should be even larger for sa1E1E2. The charge is only increased minimally compared to E2, since the interdynode voltage dominates the

[130] Newer paper than [275] based on the same PMT measurements.
[131] Shielded Photosensor Analysis eXperiment

impact energy and the SEY is very flat in this region. For these reasons, both effects are closer to the cluster at −3.5 ns and 2% pe than E2, especially when the earliness is scaled by a correction factor. On the other hand, the probabilities are far lower than for E2, which indicates minor contributions to the cluster only, perhaps in the earlier part.

Luminescence

{E1γ23}

The combination of electronic forward scattering and photonic forward emission leads to a large earliness of −5.1 ns; however, the charge of 0.08% is far too low for observation. This is one of the cases, where an AP with earliness (the E1 pulse) would result. This is also the case for E1E3 (unobservable), E1sh2E2 and E1sh2E22 (unobservable).

[γgf]

If a pk drifts onto the aluminized glass, it could produce a photon through bremsstrahlung there, which can hit k, d1 or in this case f (fig. 2.43e). However, the impact energy of 1.1 eV is too small to create photons, so these effects do not occur; all pk which hit g instead undergo Rg (fig. 2.39f).

γf1

If a pk hits the grid and dissipates its energy in SE creation, a photon can be created as a side product and emitted towards d1 (fig. 2.43d). This is only the case for lateral wire hits, since otherwise d1 is obscured by the wire. This produces a pulse with minute earliness (−0.22 ns) since only $t\mathrm{f1}_{\mathrm{pk}}$ is substituted by a photon transit time, and charge at γ1 level. The SEs produced on f cannot escape the grid potential for the R5912, so no AP is observed.

{pE$_\gamma$f1}

Identical to γf1 with the exception that the pk is elastically scattered *after* (inelastic) photon production. In this case a second pulse *can* be observed, which is comparable to Ef with T ranging from −0.8 to 3.1 ns. Which pulse is the PP thus depends on the T_r of the Ef-pulse. The effect is unobservable, however, since two improbable interactions are chained.

(pRγ32)

A p3 can also create photons as a side effect on impact. However, a second, photonic pulse only has any chance of observability if an earlier stage is hit. d1 has a small solid angle from the p3 impact site, which leaves only d2 as likely target. But even in this case, the effect cannot be observed due to being too improbable – p3 is already undetectable.

Triple and quadruple interactions

Only the most probable combinations of frequently occurring effects were studied. As for the dual interactions, only notable effects are discussed in the text.

Interdynode scattering

{E1E2E3}

This was already discussed under E1E2. First postulated by [275], the effect likely is unobservable for all PMTs, even for types with low-gain d1 and when using extremely low instrumental thresholds.

E1L2E1, E1L2E2

E1L2E1 describes a pk, which is forward scattered on d1, then backward scattered on d2, forward scattered yet another time from d1, and finally hits d2 where it produces SEs (as E1L2R1 in fig. 2.41d, but with another E1 scattering at the end). Since the space between d1 and d2 is traversed with high velocities by the scattered pk, in sum the effect is still slightly early ($-0.9\,\mathrm{ns}$) despite the two extra times the distance x_{12} is crossed. The charge is equivalent to E1. Since the pk has a far higher energy than U_{12} when backscattered on d2, it either directly hits d1 or can pass the grid before it is refocused onto d1. The probability to hit d2 again after L2 (E1L2E2) thus was estimated to be virtually zero. E1L2E1 is the only triple interaction EP studied here, which actually is probable enough to be observable in principle – if it were not for the small earliness.

2.3.6.4 Late pulses

This section takes a closer look at the plethora of late pulse effects predicted by TST, that is, interactions which produce a pulse with $T_r > 0$, independent of whether it is an PP or an AP.

Only effects with $p_c \geq 5 \cdot 10^{-5}$ and noteworthy exceptions are included.
Effects starting amplification beyond d4 were not covered, since the resulting charge is too small for observation, resp. it is very probable that these effects are an always present statistical contribution to the pulse shape due to the high number of electrons which can be subject to the effect. Note the different cutoff-stage for EPs and LPs, since e.g. E3 and L4 both start multiplication on d4. LP effects can therefore be observed until one stage later than EP effects.
Furthermore, effects were capped at three interactions, due to the rapid decrease of p_c with the number of interactions, and coarsely preselected by expected probability: For triple interactions only the most dominant effects were combined, since only then the possibility for observation is given. Nevertheless, it is difficult to predict which effects are observable before actual calculations.

The distributions for late primary pulses (LPP) [132] are shown in fig. 2.46 for the TTD (LPP-TTD) and in fig. 2.47 for the CTD (LPP-CTD). Table 2.3 summarizes the TST calculations for all LP effects, *including* APs. The discussion in the text covers all LP effects as well.

One thing which immediately is apparent when comparing the EPP-TTD (fig. 2.44) with the LPP-TTD and their corresponding tables, is that far less LP

[132] Section 2.3 deals with primary pulses only, for the afterpulses resulting from the same effects see 2.5.2.2.

effects than EP effects are observable for primary pulses, although the number of analyzed LP effects is actually much larger. This is due to the inverse behavior regarding RP occurrence: If an EP produces two pulses, one of which is an RP, the RP becomes the AP and the EP still appears in the EPP-TTD. For an LP, an RP is the *primary* pulse and thus the LP is only visible in the AP distributions. Merely LP effects which produce solely one pulse or which delay all pulses can show up in the LPP-TTD. This is only the case for effects which affect the whole cascade, and due to the high $\delta 1$ this can only be the case if the incident photon or the photoelectron is affected. In addition, only effects involving photoelectrons which are produced before d1 – at k, g, m (and f for non-accelerator grids) – can produce LPPs (compare for example $\gamma 1$ and γm). This is because tk1 is much larger than $t12$: A photoelectron created at d1 would have to accumulate a lateness equivalent to $t\text{k}1 - t\gamma\text{k}1$ through additional interactions to produce an LP instead of an EP, but not many such processes are possible. The largest lateness is provided by L2, but γ1L2 is still an EP and with an additional L2 scattering the probability is not detectable anymore. For photoelectrons produced on later dynodes the skipped transit time would be even larger and in addition the resulting charge is far lower due to missed gain steps, which reduces p_c.

LP effects which can also produce earliness are plotted and tabulated in the EP section. For reasons of clarity these effects are not shown again in the LPP-distributions. The early fraction of these effects was discussed in 2.3.6.3, whereas the late fraction is studied here during the analysis of the TTD and CTD.

For the late pulses, the AP distributions are even more important than for the EP, and during lecture of this section, the DD (fig. 2.54), the AP-CTD (fig. 2.60), the CDD (fig. 2.56 and 2.57), and the FDD (fig. 2.61 and 2.62) should be consulted.

In the following, first the distributions are discussed in general, then the LP effect table is presented, and finally individual effects are analyzed.

Transit time distribution

The LPP-TTD is shown in fig. 2.46. Please refer to the EPP-TTD discussion on p. 263 for the early part of the TTD and how the plotted TST probability values were obtained.

Only effects which affect the incident photon or a photoelectron produced on k, g, or m enter the LPP-TTD. Again, the dominant effects are from single interactions, with two exceptions: γMgk which in fact has a tremendous p_c for a dual interaction effect, and L1L1 which is clearly visible due to the isolation from other effects caused by its extreme lateness.

Effects

γk is located at the peak around 0 ns and is by far the strongest effect, which is not surprising, considering the PMT is designed for this.
While the right flank of the main peak is initially still rather Gaussian, as on the left side, already around $T_r = 1$ ns notable deviations from a normal distribution become visible (best visible in fig. 2.44), which extend beyond 6 ns. The right flank resembles a twofold exponential decay with a change of slope around 3 ns.

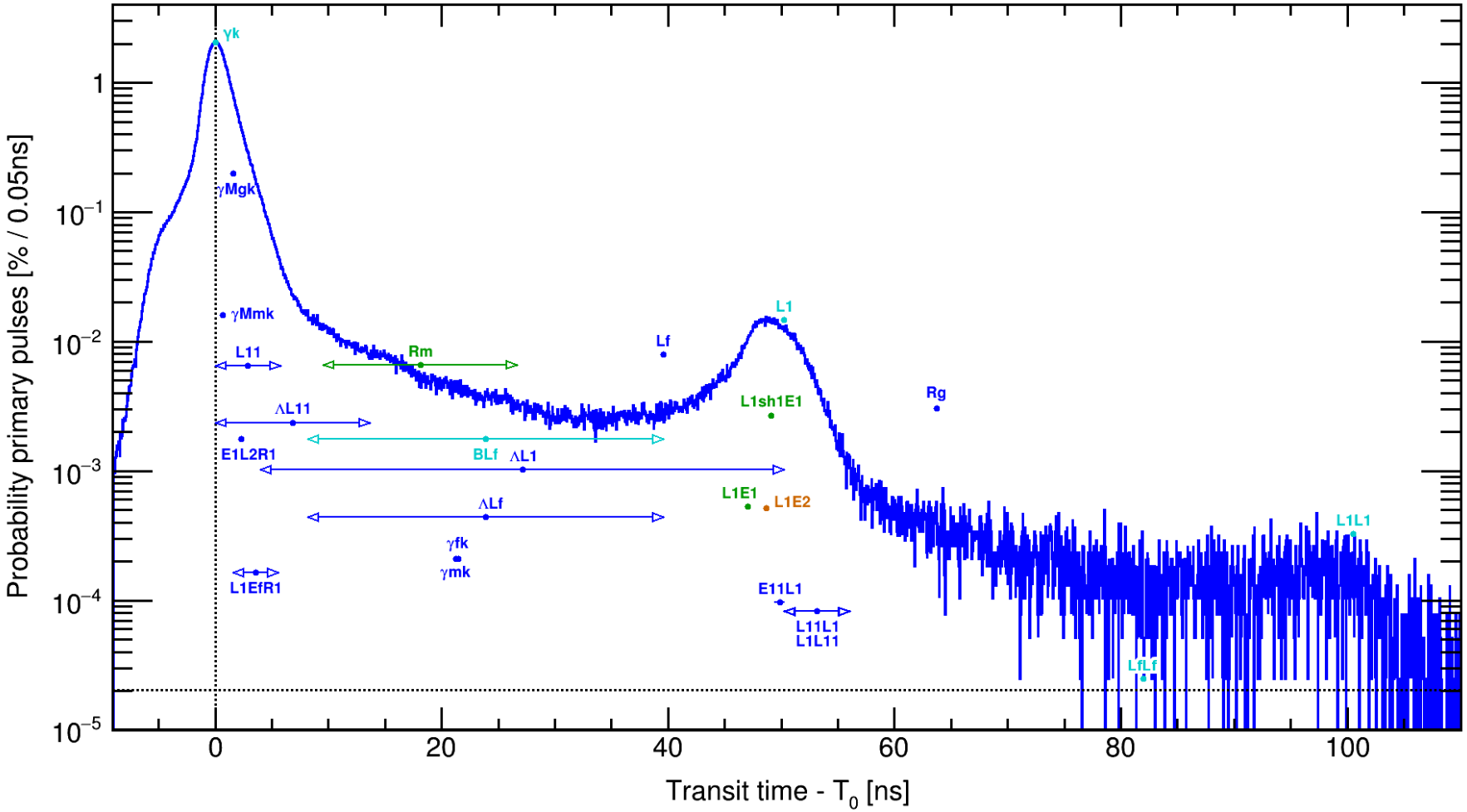


Figure 2.46: Transit time distribution (detectable probability p_c vs. relative transit time T_r) of effects causing late primary pulses (LPP-TTD). From measurement (histogram; recorded with SPAX and pulse fitting for a Hamamatsu R5912 at +1425 V, $o = 2.9\%$, $h_s = 4{,}5\%$ pe, $h_f = 1.44\%$ pe) and detectable effect probabilities p_m calculated with TST (labels, dots (peak p_m) and arrows (distributed p_m)). Detection limit (1 entry/bin) and γk time marked by dotted lines. Colors of calculated effects reflect PP charge: blue/cyan $> {}^2/_{\delta 1} = 13.3\%$, green $> {}^2/_{\delta 1 \delta 2} = 1.5\%$, orange $> {}^2/_{\delta 1 \delta 2 \delta 3} = 0.16\%$. Regular pulses (at $T_r = 0$) from late pulse (LP) effects not shown to avoid clutter (very many effects). See fig. 2.44 (EPP-TTD) to compare effects, for EP effects which can also have lateness, and a more detailed description; the plot here is the same distribution in the late pulse range. The corresponding charge-resolved distribution (LPP-CTD) is shown in fig. 2.47. In fig. 3.13d the (complete) TTD after classic analysis (no pulse fitting) of the same measurement is shown.
Only effects which cause lateness for all produced pulses enter the LPP-TTD, meaning that they have to affect the cascade while there is only one particle (photon or photoelectron). If a secondary pulse is delayed with regard to a mother pulse (mostly γk or L1), this second pulse only becomes visible in the delay distribution (DD, fig. 2.54). As this is the case for most LP effects, the LPP-TTD is sparsely populated, while the DD is filled with an abundance of effects. — Nearly all time shift effects affecting regular pulses can also modify L1 pulses in form of additional interactions, so the effect cluster around L1 mimics the main peak (only shown exemplarily for some effects). — There is a background in form of an exponential decay following the main peak due to fluorescence in the setup (see fig. 2.36 for the same PMT at identical voltage recorded with a different setup), which covers some smaller effects. — Interactions involving the grid f and the aluminized glass g seem to be systematically overestimated in probability. Apart from this, the predicted effects can well describe the distribution. Discussion see text.

On the one hand, this is caused by reflections in the region between cathode and the electron multiplier. γMgk has the largest contribution, but multiple reflections such as γMgkMkk, γMggMgk, γMgmMmk or γMgkMkgMgk will be present at higher lateness values with decreasing probabilities as well, which explains the long tail of the right flank.
On the other hand, the systematic variations of the CTS with cathode position can also explain such behavior. Without field simulations of the CTS it is difficult to distinguish the contributions from both effects. Likely both reflections and CTS variations contribute to the non-Gaussian right flank.

The region between the main peak and the L1 cluster, from 8 to about 40 ns, shows higher probability values than expected from the studied effects. This is likely due to fluorescence within the setup (SPAX) or slow secondary recombination processes in the LED (light emitting diode). The presence of such a background effect becomes apparent when looking at fig. 2.36, which was recorded for the same PMT at identical HV with a different setup (LNGS [133], see p. 457), where the exponential decay observed in fig. 2.46 (SPAX) between 7.5 and 85 ns is not present.
Instead, a flat distribution which starts to drop after about 14.5 ns becomes visible. The extent in time is in perfect agreement with Rm, but the Rm probability in this case is overestimated by a factor of 7. However, it should be taken into account that the threshold h_s used to obtain fig. 2.36 was far higher (20% pe instead of 4.5% pe), which will greatly suppress Rm. A reduction by a factor of 7 would mean that ΛL11 has comparable probabilities to Rm at lower T_r, and thus is an alternative effect which could cause the flat distribution and the subsequent drop. Presumably both effects are present. The flat distribution observed in the LNGS plot is also visible in the LPP-TTD as a bump around 15 ns on top of the exponential decay.
In the transition between the main peak right flank and Rm, the latest parts of γm and L11 likely also play a role. In LNGS data, the right flank extends until 9 ns and the slope becomes steeper around 4.5 ns, which might mark the end of reflection-based effects while subdominant effects continue to contribute a bit longer. Other EPs with lateness (ΛE11, γg, BEf, Ef, ΛEf) are predicted with too low probabilities to explain the observed features in the transition region.

In the LNGS measurement at later times ΛL1 becomes visible [134] on the left flank of L1, and shows decreasing probability with falling lateness.
Thus mainly late ΛL1 pulses are observable, likely due to kinematics: For large backscattered energies, the pk can be refocused onto d1 as it behaves similar to a pk emitted from the cathode. With falling energy, the focusing efficiency decreases and the scattered pk can be lost due to lateral momentum. At small energies, however, the scattered particle does not get far, which raises the probability to be refocused onto d1 anew. This rebound behavior is observed for BL1 in the CDD (fig. 2.56), but it is suppressed in LNGS data and does not contribute to the distribution attributed to Rm/ΛL11, since in this measurement the instrumental threshold was higher (20% pe instead of 4.5% pe) and the ΛL1 pulse charge drops with decreasing lateness. This behavior will in part contribute to the observed decrease in probability toward the main peak, since an increasing fraction of events lies below the threshold due to the charge variance.

[133] Laboratori Nazionali del Gran Sasso
[134] BLf likely has lower values as discussed subsequently.

Because of the higher threshold in LNGS data it is also expected that some BL1 pulses enter the LPP-TTD. For this effect PP and AP charges are coupled and a small PP charge might skip detection, in which case the large AP can be detected as PP, with nearly the same distribution in time as ΛL1. This ambiguity of origin is widely resolved in SPAX measurements by the small threshold; however, the setup fluorescence impedes a direct comparison of both TTDs.

γfk and γmk have very low probabilities but might be barely visible as a combined peak in LNGS data. Both effects should lie in the region of the minimum between main peak and L1 and could contribute to the excess of events over dark counts observed there.

In addition, the LNGS data (fig. 2.36) shows a bump in the ΛL1 flank around 40 ns, which is in perfect agreement with the predicted time of Lf. The probability is overestimated by a factor of at least three, which most probably comes from assuming a too high fill factor. Thus, the probabilities of BLf and ΛLf likely will be lowered by the same factor, which makes ΛL1 the dominating effect between the γk and L1 peaks.

L1 is predicted slightly later than the corresponding peak. This could be due to slightly different field geometries, the 3D movement of the backscattered particle (i.e. lateral momentum; the calculations are 1D) or small losses occurring in d1 before or after scattering (see p. 253). The probability of L1 matches the peak by definition, since the peak height was used to calculate the backward scattering direction fraction L1. The left flank of the L1 distribution can be fitted reliably with a Gaussian (red. $\chi^2 = 0.62$) between the peak and the start of notable contributions from ΛL1. The resulting $\sigma_{\mathrm{L1}} = 2.14\,\mathrm{ns}$ is far larger than $\sigma_{\gamma\mathrm{k}} = 0.799\,\mathrm{ns}$ due to transit time variations from the discussed possible causes.

In the LNGS plot (fig. 2.36), Rg is not visible at the predicted time (ca. 105 ns). This means it is either suppressed by a factor of at least about seven, or the estimated T_r is inaccurate. In fact it is very difficult to predict T_{Rg} due to the extremely low velocity of a pk which hits g. This leads to a strong dependence on U_0, the hit position on g and the potential difference U_{kg} (if different from zero), and could easily shift Rg e.g. below the L1 cluster or ΛL1 for smaller drift distances.

LfLf is neither visible in SPAX (fig. 2.46) nor in LNGS data since it is strongly suppressed by the dual interaction and the Lf p_c overestimation, beside being covered by fluorescence resp. dark counts. Its predicted time should, however, be rather accurate due to the exact prediction of Lf.

Finally, L1L1 is clearly visible as an excess over the exponential decay around 100 ns, which is followed by a steep drop at its right flank. The calculated probability is in perfect agreement with data.

Apart from the fact that the probabilities of f, g and m effect probabilities were overestimated – which can be easily matched to data by adapting their probability factors – and that the setup fluorescence covers some effects, all in all the LPP-TTD can be well described by TST calculations.

A worthwhile future endeavor would be to analytically calculate the exact $p(T)$ dependence of ΛL1, Rm and ΛL11, as these are expected to be the dominant effects in the region between the peaks. Naturally, field simulations would be even better.

Charge–time distribution

The charge–time distribution (CTD) is the charge resolved TTD, which allows to test the explanatory models put forth for the TTD. The CTD is shown in fig. 2.47. Please refer to the EPP-CTD discussion on p. 267 for the influence of thresholds, small charges and the average pulse shape. Again, a "red shift" for small charge effects is observed: For example, Rm and L1 both have similar p_c (L1 1.3%; Rm 2.2% – Rm has more events but is spread far further in time), but the bins have far more entries (→ redder on the color axis) for Rm.

Curve shape

A curved $q(T)$ behavior as for BL1 (fig. 2.60) is expected for Λ1, ΛLf, BLf, L11, and ΛL11 as well. The charge will initially remain at large values before dropping with increasing slope, which results in a concave function. For low charges the curve becomes a horizontal line. This can have two causes.
Either the SEY drops below one, in which case the charge carrier quantization sets a lower limit and only the detected probability drops with further decreasing impact energy.
Alternatively, for some effects stage skipping might occur when the electron velocity is slow enough to be directly drawn to the stage after next, since its field extends into previous interstage regions (compare fig. 2.20). For all LP effects where this is relevant, it affects photoelectrons which would normally hit d1 but are drawn to d2 instead. Since this is only expected for very small electron energies – otherwise the built-up momentum carries the particle through the deflecting fields – the charge will be similar to that of an SE emitted from d1. That means that the charge will be at Rm level. This is likely the cause of the horizontal cluster from 4 to 10 ns at 5.5% pe, which seamlessly merges into Rm. Contributions come from the end points of ΛL11, BEf and ΛEf (falling $q(T)$ dependence), and those of ΛL1, BLf and ΛLf (rising $q(T)$). Stage-skipping of BEf, ΛEf and ΛL11 could explain the sharp cutoff observed around 5 ns for charges of about 40% pe, which cannot be understood based on the SEY shape alone.

Fluorescence

The exponentially decaying background, which occurs in response to the main peak, unfortunately covers many effects, namely ΛL1, BLf, Lf, ΛLf, γmk, γfk, and still Rg and LfLf. It is therefore difficult to assess the presence of these effects.
The origin of this background might be from fluorescing materials in the setup, e.g. the reflective diffuser for which a Spectralon plate was used. The PTB (Physikalisch-Technische Bundesanstalt) reported fluorescence of PTFE [135] in [344].
The exponential decay could also stem from slow exciton recombinations in the LED used as light source.
Since this background is over two orders of magnitudes weaker than the main peak, it was not noticed before conclusion of measurements and post-analysis. For future

[135] Polytetrafluoroethylene (Teflon)

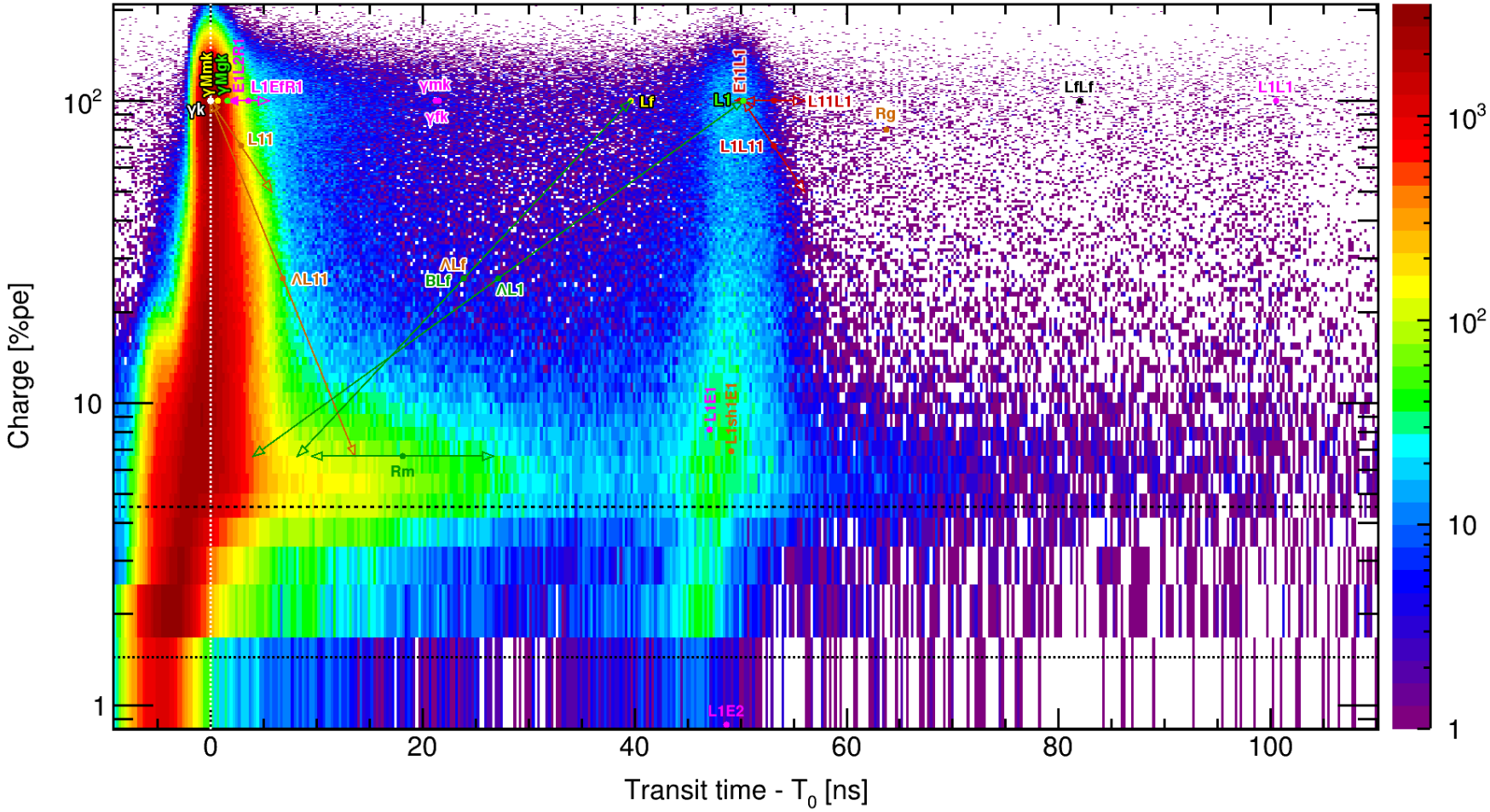


Figure 2.47: Charge–time distribution for late primary pulses (LPP-CTD) from data compared to calculated TST effects. The figure shows the charge (y-axis) and probability (z-axis) of primary pulses vs. their relative transit time (x-axis).
Measurement for a Hamamatsu R5912 (+1425 V, $g = 1.3 \cdot 10^7$) with SPAX, $o = 2.9\%$, pulse fitting, $h_s = 4.5\%$ pe, $h_f = 1.44\%$ pe, bin widths 0.25 ns and 0.02 pC. 1 pe = 2.40 pC (charge of clean γk pulses). Number of pulses per bin indicated by z-axis (colors). Discriminator threshold h_s marked by dashed line, fit threshold h_f by dotted line. γk time indicated by white dotted line; all pulses occurring later than this are by definition late pulses (LPs).
Effects calculated with TST drawn in with labels, markers and arrows/lines, where the color depicts the detectable probability p_c: $> 65\%$ white/violet (above this value $N > 1$ occurrences per PP, except for γk), $> 10\%$ cyan/blue, $> 1\%$ bright/dark green, $> 0.1\%$ yellow/orange, $> 0.01\%$ pink/magenta, $< 0.01\%$ red/black. Point-like distributions marked by dots, time and charge range of effects shown by vertical and horizontal arrows, diagonal arrows implicate the qualitative charge–time dependence of an effect.
Note the repetition of all early and late effects surrounding the RP around L1 like an echo (e.g. Rm and E1). — Λ1, ΛLf and BLf show curved behavior similar to BL1 (see fig. 2.60). This can in part explain the pulses at high charges for times between 0 and L1; however, the majority is expected to stem from setup fluorescence, which also impedes recognition of Lf, γmk, γfk, Rg and LfLf. L11 and ΛL11 will also show a concave $q(T)$ dependence due to the SEY shape. — The band around 5.5% pe extending to about 35 ns is from Rm, γm and stage-skipping ΛL11, BEf and ΛEf up to about 30 ns, followed by secondary effects on d1 which introduce additional lateness (e.g. RmL11). — For detailed discussion of plot and effects see text. — For the corresponding TTD see fig. 2.46 (LPP-TTD). Confer the EPP-CTD (fig. 2.45) for EP effects – effects already plotted there are omitted here – and more details about the measurement and display of data.

studies this "setup fluorescence" needs to be eliminated by piecewise exchange of components, for example by using $BaSo_4$ as diffuser, and by employing different LEDs – or better a ps-laser – to verify the influence of the light source.

Effects

The extra charge dimension shows that the falling flank of the main peak can not solely be explained by reflections, which should result in charges close to 100%. After about $T_r = 3\,\text{ns}$ the contributions are mostly from pulses with reduced charges of 40% pe or lower. This indicates other origins, either cathode time spread from extreme starting positions which results in reduced $\delta 1$ due to the edge effect, or other effects such as L11, ΛL11, BEf or ΛE, which cover exactly this region with concave $q(T)$ curves. Since BL1 shows a steep drop toward small charges, the curves of ΛL1, BLf and ΛLf, which are expected to have similar shapes, still yield contributions in the region of 20–30% pe from 7–9 ns. However, the expected probabilities of all these effects are one to two magnitudes lower than would be required to explain the TTD in this region. This discrepancy is too high to resolve through fine-tuning of probability factors, as this would also change the p_c of many other effects. This points to CTS with reduced charge as the most likely origin for the majority of events and also for the abrupt cutoff at 5 ns.

The band at 5.5% pe spanning the range between main peak and L1 cluster originates from Rm, γm and stage skipping for the effects discussed above up to about 27 ns. The pulses observed at later times emerge from secondary interactions on d1 following the other effects, which introduce additional lateness, e.g. RmL11 (not calculated).

The group of effects around L1 times shows a repetition of the EP and LP effects around the main peak in form of dual interactions based on L1; only a few exemplary effects are shown, but nearly all combinations are possible. The different starting position of the pk (on d1 instead of k) introduces some changes compared to γk-based effects. The spread in time is larger due to the 3D movement and possibly small inelastic scattering losses, and the L1Rm population could contain more events compared to L1 than Rm to γk, since the lateral momentum promotes re-focusing losses.
The same behavior is expected for L1L1, but this is only barely visible due to low statistics.

Lastly, L1L1 is visible as an increase of density over the complete charge range around the predicted time of 100.5 ns.

Charge distribution

When the CTD is projected onto the y-axis, the SER is obtained. By comparison of fig. 2.45 and 2.47 with the SER (fig. 2.22c) the origin of the underamplified pulses becomes clear.
Since the valley in the SER is located around 25% for this PMT, the effects which limit the peak-to-valley ratio are: CTS, B-type RPs, BE1, ΛE1, possibly the Shockley-Ramo pulse, ΛE11, ΛL11, BEf, ΛEf, BLf, ΛLf, and ΛL1. In a classic analysis without pulse shape analysis and with moderate threshold, the Shockley-Ramo pulse is

unresolvable. In addition, B-type EPs will be present as pileup on the regular pulse, and the resulting charge sums will also be situated in the valley region.

As a next step it would be beneficial to calculate the spe SER predicted by TST and compare it with data. This could be accomplished by approximating the charge distributions of effects with scaled Poisson functions. For effects with a charge range, however, this is no valid model and the expected function of the probability vs. charge could be folded with a scaled Poisson whose λ depends on the current charge. From the resulting "itemized" SER the dominant TSEs contributing to the underamplified region could be identified, which would allow to determine the best phenomenological fit functions for the undersized pulses.

Branching and total gain

A near even energy branching for an L-effect can actually be favorable with regard to the total charge, since the primary creates almost as many SEs on its second hit as a full deposit because of the concave shape of SEY(E). Thus the total charge is larger than for a regular full energy deposit on this stage. If the backward backscattering could be exploited purposefully, this would allow to improve the gain – e.g. by using a material with an even higher effective atomic number Z_{eff}, or by a layout which is designed to focus L-scattered primaries back onto their scattering stage. This, however, comes at the cost of timing and likely would also increase the charge resolution, since branching is a statistical process and does not always occur.

Comparison to literature

The CTD in fig. 4 shown in [301] shows a large number of the effects visible in the LPP-TTD resp. EPP-TTD for a different PMT, the Hamamatsu R7081 (10"), since a low threshold of 10% pe was used in these measurements as well. Their data shows γ1, E1, the events around 50% pe which occur a few nanoseconds before γk, a similar falling flank of the main peak with an abrupt charge drop, Rm, ΛL1 or BL1 (if some PPs are not recognized), L1 and secondary effects below L1. The independent measurement for a different device confirms the findings presented here.

Table of effects

In table 2.3, the results of TST for all LP effects are presented. Please confer p. 273 for the explanations on how to read the table, and general comments.

Note the large number of LP effects compared to the few effects which can cause PP lateness, shown in fig. 2.47. A great many photonic effects from luminescence (of any type) are found, which dominate the delay distribution after about 12 ns delay (fig. 2.54). Intradynode forward scattering in many cases only introduces lateness and no phase space producing earliness is found (sx1E11, sx3E33). Dual and higher order interactions only play a minor part due to the rapid decrease of probability, with the exception of photon reflections.

In the following, individual effects of note are discussed, using the same order as for the EP effects and as used in table 2.3.6.4.

Table 2.3: Effects causing late pulses (PP or AP) derived with TST. Values and conventions as in table 2.2. Continued on next pages. Discussion see text.

Effect	**Topology**	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
γ**k**	γPk→pR1	1	100	99.12	99.73	1	[(6.7,) 100.0]	0	0							
γ**Mgk**	γMgk→γPk→pR1	$\mathcal{T}\frac{A_g}{A_k}R_g\cdot\gamma tMgk$	9.7	9.6	9.7	1	100.0	$2t\gamma kg_\uparrow$	1.6							
γ**Mmk**	γMmk→γPk→pR1	$\mathcal{T}\frac{A_m}{A_k}R_m\cdot\gamma tMmk$	0.78	0.78	0.78	1	100.0	$2t\gamma km$	0.7							
L1	pL1→pR1	$\frac{b1}{u1}e1\cdot L1\cdot u1_{\varepsilon1}$	1.3	1.2	1.3	$\frac{\delta1_{\varepsilon1}}{\delta1}$	100.0	$2t1k_{\varepsilon1}$	50.3							
Λ**L1**	pΛL1→pR1	$\frac{b1}{u1}i1\cdot\Lambda1\cdot L\Lambda1\cdot u1_{\Lambda1}$	1.0	1.0	1.0	$\frac{[1,\delta1_{\varepsilon1}]}{\delta1}$	[6,7, 100.0]	$2[t1k_{\Lambda1}, t1k_{\varepsilon1}]$	[4.0, 50.3]							
BL1	pBL1→a1R2 ↘ pR1	$\frac{b1}{u1}i1\cdot B1\cdot LB1\cdot u1_{s1}$	4.2	4.1	4.1	$\frac{[1,\delta_s1]}{\delta1}$	[6.7, 100.0]	0	0	$\frac{1}{\delta1}\left[\max(1, \delta1_{s1}), \delta_s1_{\varepsilon1}\right]$	[6.7, 100.0]	[106.7, 106.7]	[6.2, 93.8]	$2[t1k_{s1}, t_s1k_{\varepsilon1}]$	[2.6, 45.3]	[2.6, 45.3]
L2	pR1→r1R2 ↘ s1L2→s1R2	$1-(1-b2\cdot e2\cdot L2\cdot u2_{\varepsilon2})^{\delta1}$	20 (0.23)	20	19	$1-\frac{1}{\delta1}$	93.3	0	0	$\frac{1}{\delta1}\frac{\delta2_{\varepsilon2}}{\delta2}$	6.7	100.0	6.7	$2t21_{\varepsilon2}$	8.7	8.7
Λ**L2**	pR1→r1R2 ↘ s1ΛL2→s1R2	$1-(1-b2\cdot i2\cdot\Lambda2\cdot L\Lambda2\cdot u2_{s2})^{\delta1}$	15 (0.16)	15	14	$1-\frac{1}{\delta1}$	93.3	0	0	$\frac{1}{\delta1}\frac{[1,\delta2_{\varepsilon2}]}{\delta2}$	[0.77, 6.7]	[94.1, 100.0]	[0.81, 6.7]	$2[t21_{s2}, t21_{\varepsilon2}]$	[1.0, 8.7]	[1.0, 8.7]
BL2	pR1→r1R2 →a2jR3 ↘ s1BL2 ↗ →s1R2	$1-(1-b2\cdot i2\cdot B2\cdot LB2\cdot u2_{s2})^{\delta1}$	48 (0.63)	47	45	$1-\frac{1}{\delta1}+\frac{[0,\delta_s2]}{\delta1\delta2}$	[93.3, 99.9]	0	0	$\frac{1}{\delta1}\frac{[1,\delta2_{\varepsilon2}]}{\delta2}$	[0.77, 6.7]	[100.0, 100.6]	[0.76, 6.7]	$2[t21_{s2}, t21_{\varepsilon2}]$	[1.0, 8.7]	[1.0, 8.7]
L3	pR1→a1R2→r2R3 ↘ s2L3→s2R3	$1-(1-b3\cdot e3\cdot L3\cdot u3_{\varepsilon3})^{\delta1\delta2}$	81 (1.7)	81	33	$1-\frac{1}{\delta1\delta2}$	99.2	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\varepsilon3}}{\delta3}$	0.77	100.0	0.77	$2t32_{\varepsilon3}$	5.4	5.4
Λ**L3**	a1R2→r2R3 ↘ s2ΛL3→s2R3	$1-(1-b3\cdot i3\cdot\Lambda3\cdot L\Lambda3\cdot u3_{s3})^{\delta1\delta2}$	70 (1.2)	70	12	$1-\frac{1}{\delta1\delta2}$	99.2	0	0	$\frac{1}{\delta1\delta2}\frac{[1,\delta3_{\varepsilon3}]}{\delta3}$	[0.08, 0.77]	[99.3, 100.0]	[0.08, 0.77]	$2[t32_{s3}, t32_{\varepsilon3}]$	[0.6, 5.4]	[0.6, 5.4]
BL3	a1R2→r2R3 →a3jR4 ↘ s2BL3 ↗ →s2R3	$1-(1-b3\cdot i3\cdot B3\cdot LB3\cdot u3_{s3})^{\delta1\delta2}$	99 (4.9)	99	98	$1-\frac{1}{\delta1\delta2}+\frac{[0,\delta_s3]}{\delta1\delta2\delta3}$	[99.2, 100.0]	0	0	$\frac{1}{\delta1\delta2}\frac{[1,\delta3_{\varepsilon3}]}{\delta3}$	[0.08, 0.77]	[100.0, 100.1]	[0.08, 0.77]	$2[t32_{s3}, t32_{\varepsilon3}]$	[0.6, 5.4]	[0.6, 5.4]
L4	a2R3→r3R4 ↘ s3L4→s3R4	$1-(1-b4\cdot e4\cdot L4\cdot u4_{\varepsilon4})^{\delta1\delta2\delta3}$	100 (14.7)	100	9.2	$1-\frac{1}{\delta1\delta2\delta3}$	99.9	0	0	$\frac{1}{\delta1\delta2\delta3}\frac{\delta4_{\varepsilon4}}{\delta4}$	0.08	100.0	0.08	$2t43_{\varepsilon4}$	7.1	7.1

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
ΛL4	a2R3→r3R4 ↘ s3ΛL4→s3R4	$1-(1-b4\cdot i4\cdot\Lambda4\cdot L\Lambda4\cdot u4_{s4})^{\delta1\delta2\delta3}$	100 (10.0)	100	$2.9\cdot10^{-5}$	$1-\frac{1}{\delta1\delta2\delta3}$	99.9	0	0	$\frac{1}{\delta1\delta2\delta3}\cdot\frac{[1,\delta4_{\varepsilon4}]}{\delta4}$	[0.01, 0.08]	[99.9, 100.0]	[0.01, 0.08]	$2[t43_{s4}, t43_{\varepsilon4}]$	[1.0, 7.1]	[1.0, 7.1]
BL4	a2R3→r3R4 →a4jR5 ↘ ↗ s3BL4→s3R4	$1-(1-b4\cdot i4\cdot B4\cdot LB4\cdot u4_{s4})^{\delta1\delta2\delta3}$	100 (39.9)	100	100	$1-\frac{1}{\delta1\delta2\delta3}+\frac{[0,\delta_s4]}{\delta1\delta2\delta3\delta4}$	[99.9, 100.0]	0	0	$\frac{1}{\delta1\delta2\delta3}\cdot\frac{[1,\delta4_{\varepsilon4}]}{\delta4}$	[0.01, 0.08]	[100.0, 100.0]	[0.01, 0.08]	$2[t43_{s4}, t43_{\varepsilon4}]$	[1.0, 7.1]	[1.0, 7.1]
L11	pL11→pR1	$\frac{b1}{u1}e1\cdot L11\cdot u1_{\varepsilon1}$	0.77	0.76	0.77	$\frac{\delta1_{\varepsilon1}}{\delta1}[r1,1]$	[50, 100.0]	$[0, t11_{\uparrow}+t12_{\uparrow}-t12]$	[0, 5.8]							
ΛL11	pΛL11→pR1	$\frac{b1}{u1}i1\cdot\Lambda1\cdot L\Lambda11\cdot u1_{\Lambda1}$	0.63	0.62	0.63	$\frac{[1,\delta1_{\varepsilon1}]}{\delta1}$	[6.7, 100.0]	$\left[0,[t11_{\uparrow},t11_{\Lambda1\uparrow}]+t12_{\uparrow}-t12\right]$	[0, 13.7]							
BL11	pBL11→a1R2 ↘ pR1	$\frac{b1}{u1}i1\cdot B1\cdot LB11\cdot u1_{B1}$	2.5	2.5	2.5	$\frac{[1,\delta_s1]}{\delta1}$	[6.7, 100.0]	0	0	$\frac{[1,\delta_s1_{\varepsilon1}]}{\delta1}$	[6.7, 100.0]	[106.7, 106.7]	[6.2, 93.8]	$\left[0,[t11_{\uparrow},t11_{B1\uparrow}]+t12_{\uparrow}-t12\right]$	[0, 13.7]	[0, 13.7]
L22	pR1→r1R2 ↘ s1L22→s1R2	$1-(1-b2\cdot e2\cdot L22\cdot u2_{\varepsilon2})^{\delta1}$	2.8	2.8	2.6	$1-\frac{1}{\delta1}$	93.3	0	0	$\frac{1}{\delta1}\frac{\delta2_{\varepsilon2}}{\delta2}[r2,1]$	[3.3, 6.7]	[96.7, 100.0]	[3.4, 6.7]	$[0,t22_{\uparrow}+t23_{\uparrow}-t23]$	[0, 3.2]	[0, 3.2]
ΛL22	pR1→r1R2 ↘ s1ΛL22→s1R2	$1-(1-b2\cdot i2\cdot\Lambda2\cdot L\Lambda22\cdot u2_{\Lambda2})^{\delta1}$	1.9	1.9	1.8	$1-\frac{1}{\delta1}$	93.3	0	0	$\frac{1}{\delta1}\frac{[1,\delta2_{\varepsilon2}]}{\delta2}$	[0.77, 6.7]	[94.1, 100.0]	[0.81, 6.7]	$\left[0,[t22_{\uparrow},t22_{\Lambda2\uparrow}]+t23_{\uparrow}-t23\right]$	[0, 6.3]	[0, 6.3]
BL22	pR1→r1R2 →a2jR3 ↘ ↗ s1BL22→s1R2	$1-(1-b2\cdot i2\cdot B2\cdot LB22\cdot u2_{B2})^{\delta1}$	7.5 (0.078)	7.5	7.1	$1-\frac{1}{\delta1}+\frac{[0,\delta_s2]}{\delta1\delta2}$	[93.3, 99.9]	0	0	$\frac{1}{\delta1}\frac{[1,\delta2_{\varepsilon2}]}{\delta2}$	[0.77, 6.7]	[100.0, 100.6]	[0.76, 6.7]	$\left[0,[t22_{\uparrow},t22_{B2\uparrow}]+t23_{\uparrow}-t23\right]$	[0, 6.3]	[0, 6.3]
L33	pR1→a1R2→r2R3 ↙ s2L33→s2R3	$1-(1-b3\cdot e3\cdot L33\cdot u3_{\varepsilon3})^{\delta1\delta2}$	43 (0.56)	42	3.4	$1-\frac{1}{\delta1\delta2}$	99.2	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\varepsilon3}}{\delta3}\cdot[r3,1]$	[0.38, 0.77]	[99.6, 100.0]	[0.38, 0.77]	$[0,t33_{\uparrow}+t34_{\uparrow}-t34]$	[0, 3.1]	[0, 3.1]
ΛL33	pR1→a1R2→r2R3 ↙ s2ΛL33→s2R3	$1-(1-b3\cdot i3\cdot\Lambda3\cdot L\Lambda33\cdot u3_{\Lambda3})^{\delta1\delta2}$	33 (0.40)	33	2.6	$1-\frac{1}{\delta1\delta2}$	99.2	0	0	$\frac{1}{\delta1\delta2}\frac{[1,\delta3_{\varepsilon3}]}{\delta3}$	[0.08, 0.77]	[99.3, 100.0]	[0.08, 0.77]	$\left[0,[t33_{\uparrow},t33_{\Lambda3\uparrow}]+t34_{\uparrow}-t34\right]$	[0, 5.3]	[0, 5.3]
BL33	a1R2→r2R3 →a3jR4 ↘ ↗ s2BL33→s2R3	$1-(1-b3\cdot i3\cdot B3\cdot LB33\cdot u3_{B3})^{\delta1\delta2}$	80 (1.6)	80	29	$1-\frac{1}{\delta1\delta2}+\frac{[0,\delta_s3]}{\delta1\delta2\delta3}$	[99.2, 100.0]	0	0	$\frac{1}{\delta1\delta2}\frac{[1,\delta3_{\varepsilon3}]}{\delta3}$	[0.08, 0.77]	[100.0, 100.1]	[0.08, 0.77]	$\left[0,[t33_{\uparrow},t33_{B3\uparrow}]+t34_{\uparrow}-t34\right]$	[0, 5.3]	[0, 5.3]

Effect	**Topology**	$\boldsymbol{p_r}$	$\boldsymbol{p_r(N)}$	p_s	p_f	$\boldsymbol{q_{PP}}$		$\boldsymbol{T_{PP}}$		$\boldsymbol{q_{AP}}$		$\boldsymbol{\Sigma_q}$	$\boldsymbol{\eta_q}$	$\boldsymbol{T_{AP}}$		$\boldsymbol{d}$
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
s1L1	pR1→r1R2 ↘ s1L1→s1R1	$1-(1-\mathrm{c1}\cdot\mathrm{L1_{s1}}\cdot\mathrm{u1_{s1}}\cdot\mathrm{c1_{s1}})^{\delta 1}$	2.4 (0.025)	2.4	2.4	$\frac{[\delta 1-1,\delta_{\mathrm{sh}}1]}{\delta 1}$	[93.3, 100.0]	0	0	$\frac{1}{\delta 1}\max\left(1,[\delta 1_{\mathrm{s1}},\delta 1_{\mathrm{sh1}}]\right)$	[6.7, 14]	[100.0, 114]	[6.7, 12]	$2[t1\mathrm{k_{s1}},t1\mathrm{k_{sh1}}]$	[2.6, 5.9]	[2.6, 5.9]
sa1L1	pR1→r1R2 ↘ sa1L1→sa1R1	$1-(1-\mathrm{s1_a}\cdot\mathrm{L1_{sa1}}\cdot\mathrm{u1_{sa1}}\cdot\mathrm{c1_{sa1}})^{\delta 1}$	0.53	0.52	0.52	$\frac{\delta_{\mathrm{sa}}1}{\delta 1}$	100.0	0	0	$\frac{\max(1,\delta 1_{\mathrm{sa1}})}{\delta 1}$	20	120	17	$2t1\mathrm{k_{sa1}}$	7.1	7.1
s2L2	pR1→a1R2→r2R3 ↘ s2L2→s2R2	$1-(1-\mathrm{c2}\cdot\mathrm{L2_{s2}}\cdot\mathrm{u2_{s2}}\cdot\mathrm{c2_{s2}})^{\delta 1\delta 2}$	48 (0.65)	47	18	$1-\frac{[1,\delta 2-\delta_{\mathrm{sh}}2]}{\delta 1\delta 2}$	[99.2, 99.6]	0	0	$\frac{1}{\delta 1\delta 2}\max\left(1,[\delta 2_{\mathrm{s2}},\delta 2_{\mathrm{sh2}}]\right)$	[0.77, 1.2]	[100.0, 100.8]	[0.77, 1.2]	$2[t21_{\mathrm{s2}},t21_{\mathrm{sh2}}]$	[1.0, 2.3]	[1.0, 2.3]
sa2L2	a1R2→r2R3 ↘ sa2L2→sa2R2	$1-(1-\mathrm{s2_a}\cdot\mathrm{L2_{sa2}}\cdot\mathrm{u2_{sa2}}\cdot\mathrm{c2_{sa2}})^{\delta 1\delta 2}$	4.2 (0.043)	4.2	2.6	$1-\frac{1}{\delta 1}+\frac{\delta_{\mathrm{sa}}2}{\delta 1\delta 2}$	99.6	0	0	$\frac{1}{\delta 1\delta 2}\max(1,\delta 2_{\mathrm{sa2}})$	1.7	101.3	1.7	$2t21_{\mathrm{sa2}}$	2.8	2.8
s3L3	a2R3→r3R4 ↘ s3L3→s3R3	$1-(1-\mathrm{c3}\cdot\mathrm{L3_{s3}}\cdot\mathrm{u3_{s3}}\cdot\mathrm{c3_{s3}})^{\delta 1\delta 2\delta 3}$	100 (9.2)	100	8.9	$1-\frac{[1,\delta 3-\delta_{\mathrm{sh}}3]}{\delta 1\delta 2\delta 3}$	[99.9, 100.0]	0	0	$\frac{1}{\delta 1\delta 2\delta 3}\cdot\max\left(1,[\delta 3_{\mathrm{s3}},\delta 3_{\mathrm{sh3}}]\right)$	[0.08, 0.12]	[100.0, 100.1]	[0.08, 0.12]	$2[t32_{\mathrm{s3}},t32_{\mathrm{sh3}}]$	[0.6, 1.3]	[0.6, 1.3]
sa3L3	a2R3→r3R4 ↘ sa3L3→sa3R3	$1-(1-\mathrm{s3_a}\cdot\mathrm{L3_{sa3}}\cdot\mathrm{u3_{sa3}}\cdot\mathrm{c3_{sa3}})^{\delta 1\delta 2\delta 3}$	64 (1.0)	63	0	$1-\frac{1}{\delta 1\delta 2}+\frac{\delta_{\mathrm{sa}}3}{\delta 1\delta 2\delta 3}$	100.0	0	0	$\frac{1}{\delta 1\delta 2\delta 3}\cdot\max(1,\delta 3_{\mathrm{sa3}})$	0.18	100.1	0.18	$2t32_{\mathrm{sa3}}$	1.5	1.5
s1E11	pR1→r1R2 ↘ s1E11→s1R1	$1-(1-\mathrm{c1}\cdot\mathrm{E11_{s1}}\cdot\mathrm{u1_{s1}}\cdot\mathrm{c1_{s1}})^{\delta 1}$	0.98 (0.010)	0.97	0.96	$\frac{[\delta 1-1,\delta_{\mathrm{sh}}1]}{\delta 1}$	[93.3, 100.0]	0	0	$\frac{1}{\delta 1}\max\left(1,[\delta 1_{\mathrm{s1}},\delta 1_{\mathrm{sh1}}]\cdot[\mathrm{r1},1]\right)$	[6.7, 14]	[100.0, 114]	[6.7, 12]	$\left[0,[t11_{\mathrm{s1}\downarrow},t11_{\mathrm{sh1}\downarrow}]+t12_{\downarrow}-t12\right]$	[0, 0.5]	[0, 0.5]
sa1E11	pR1→r1R2 ↘ sa1E11→sa1R1	$1-(1-\mathrm{s1_a}\cdot\mathrm{E11_{sa1}}\cdot\mathrm{u1_{sa1}}\cdot\mathrm{c1_{sa1}})^{\delta 1}$	0.20	0.20	0.20	$\frac{\delta_{\mathrm{sa}}1}{\delta 1}$	100.0	0	0	$\frac{1}{\delta 1}\max\left(1,\delta 1_{\mathrm{sa1}}[\mathrm{r1},1]\right)$	[10, 20]	[110, 120]	[9.2, 17]	$[0,t11_{\mathrm{sa1}\downarrow}+t12_{\downarrow}-t12]$	[0, 0.0]	[0, 0.0]
s1L11	pR1→r1R2 ↘ s1L11→s1R1	$1-(1-\mathrm{c1}\cdot\mathrm{L11_{s1}}\cdot\mathrm{u1_{s1}}\cdot\mathrm{c1_{s1}})^{\delta 1}$	4.8 (0.049)	4.8	4.7	$\frac{[\delta 1-1,\delta_{\mathrm{sh}}1]}{\delta 1}$	[93.3, 100.0]	0	0	$\frac{1}{\delta 1}\max\left(1,[\delta 1_{\mathrm{s1}}\cdot\mathrm{r1},\delta 1_{\mathrm{sh1}}]\right)$	[6.7, 14]	[100.0, 114]	[6.7, 12]	$\left[0,[t11_{\mathrm{s1}\uparrow},t11_{\mathrm{sh1}\uparrow}]+t12_{\uparrow}-t12\right]$	[0, 5.8]	[0, 5.8]
sa1L11	pR1→r1R2 ↘ sa1L11→sa1R1	$1-(1-\mathrm{s1_a}\cdot\mathrm{L11_{sa1}}\cdot\mathrm{u1_{sa1}}\cdot\mathrm{c1_{sa1}})^{\delta 1}$	0.40	0.39	0.39	$\frac{\delta_{\mathrm{sa}}1}{\delta 1}$	100.0	0	0	$\frac{1}{\delta 1}\max\left(1,\delta 1_{\mathrm{sa1}}[\mathrm{r1},1]\right)$	[10, 20]	[110, 120]	[9.2, 17]	$[0,t11_{\mathrm{sa1}\uparrow}+t12_{\uparrow}-t12]$	[0, 5.4]	[0, 5.4]
s2L22	a1R2→r2R3 ↘ s2L22→s2R2	$1-(1-\mathrm{c2}\cdot\mathrm{L22_{s2}}\cdot\mathrm{u2_{s2}}\cdot\mathrm{c2_{s2}})^{\delta 1\delta 2}$	62 (0.97)	62	23	$1-\frac{[1,\delta 2-\delta_{\mathrm{sh}}2]}{\delta 1\delta 2}$	[99.2, 99.6]	0	0	$\frac{1}{\delta 1\delta 2}\max\left(1,[\delta 2_{\mathrm{s2}}\cdot\mathrm{r2},\delta 2_{\mathrm{sh2}}]\right)$	[0.77, 1.2]	[100.0, 100.8]	[0.77, 1.2]	$\left[0,[t22_{\mathrm{s2}\uparrow},t22_{\mathrm{sh2}\uparrow}]+t23_{\uparrow}-t23\right]$	[0, 6.3]	[0, 6.3]

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
sa2L22	a1R2→r2R3 ↘ sa2L22→sa2R2	$1-(1-\mathrm{s2_a}\cdot\mathrm{L22_{sa2}}\cdot\mathrm{u2_{sa2}}\cdot\mathrm{c2_{sa2}})^{\delta1\delta2}$	4.2 (0.043)	4.2	2.6	$1-\frac{1}{\delta1}+\frac{\delta_{\mathrm{sa}}2}{\delta1\delta2}$	99.6	0	0	$\frac{1}{\delta1\delta2}\max\left(1,\delta2_{\mathrm{sa2}}[\mathrm{r2},1]\right)$	[0.84, 1.7]	[100.5, 101.3]	[0.83, 1.7]	$[0,t22_{\mathrm{sa2}\uparrow}+t23_{\uparrow}-t23]$	[0, 5.1]	[0, 5.1]
s3E33	a2R3→r3R4 ↘ s3E33→s3R3	$1-(1-\mathrm{c3}\cdot\mathrm{E33_{s3}}\cdot\mathrm{u3_{s3}}\cdot\mathrm{c3_{s3}})^{\delta1\delta2\delta3}$	95 (3.1)	95	0	$1-\frac{[1,\delta3-\delta_{\mathrm{sh}}3]}{\delta1\delta2\delta3}$	[99.9, 100.0]	0	0	$\frac{1}{\delta1\delta2\delta3}\cdot\max\left(1,[\delta3_{\mathrm{s3}}\cdot\mathrm{r3},\delta3_{\mathrm{sh3}}]\right)$	[0.08, 0.12]	[100.0, 100.1]	[0.08, 0.12]	$\left[0,[t33_{\mathrm{s3}\downarrow},t33_{\mathrm{sh3}\downarrow}]+t34_{\downarrow}-t34\right]$	[0, 0.7]	[0, 0.7]
sa3E33	a2R3→r3R4 ↘ sa3E33→sa3R3	$1-(1-\mathrm{s3_a}\cdot\mathrm{E33_{sa3}}\cdot\mathrm{u3_{sa3}}\cdot\mathrm{c3_{sa3}})^{\delta1\delta2\delta3}$	18 (0.20)	18	0	$1-\frac{1}{\delta1\delta2}+\frac{\delta_{\mathrm{sa}}3}{\delta1\delta2\delta3}$	100.0	0	0	$\frac{1}{\delta1\delta2\delta3}\cdot\max\left(1,\delta3_{\mathrm{sa3}}[\mathrm{r3},1]\right)$	[0.09, 0.18]	[100.1, 100.1]	[0.09, 0.18]	$[0,t33_{\mathrm{sa3}\downarrow}+t34_{\downarrow}-t34]$	[0, 0.2]	[0, 0.2]
s3L33	a2R3→r3R4 ↘ s3L33→s3R3	$1-(1-\mathrm{c3}\cdot\mathrm{L33_{s3}}\cdot\mathrm{u3_{s3}}\cdot\mathrm{c3_{s3}})^{\delta1\delta2\delta3}$	100 (12.3)	100	60	$1-\frac{[1,\delta3-\delta_{\mathrm{sh}}3]}{\delta1\delta2\delta3}$	[99.9, 100.0]	0	0	$\frac{1}{\delta1\delta2\delta3}\cdot\max\left(1,[\delta3_{\mathrm{s3}}\cdot\mathrm{r3},\delta3_{\mathrm{sh3}}]\right)$	[0.08, 0.12]	[100.0, 100.1]	[0.08, 0.12]	$\left[0,[t33_{\mathrm{s3}\uparrow},t33_{\mathrm{sh3}\uparrow}]+t34_{\uparrow}-t34\right]$	[0, 5.3]	[0, 5.3]
sa3L33	a2R3→r3R4 ↘ sa3L33→sa3R3	$1-(1-\mathrm{s3_a}\cdot\mathrm{L33_{sa3}}\cdot\mathrm{u3_{sa3}}\cdot\mathrm{c3_{sa3}})^{\delta1\delta2\delta3}$	56 (0.81)	55	0	$1-\frac{1}{\delta1\delta2}+\frac{\delta_{\mathrm{sa}}3}{\delta1\delta2\delta3}$	100.0	0	0	$\frac{1}{\delta1\delta2\delta3}\cdot\max\left(1,\delta3_{\mathrm{sa3}}[\mathrm{r3},1]\right)$	[0.09, 0.18]	[100.1, 100.1]	[0.09, 0.18]	$[0,t33_{\mathrm{sa3}\uparrow}+t34_{\uparrow}-t34]$	[0, 4.5]	[0, 4.5]
Rg	pRg→agR1	$\frac{\mathrm{CE_{kg}}}{\mathrm{CE_{k1}}}\mathrm{bg}\cdot\mathrm{CE_{g1}}\frac{\mathrm{u1_{sg}}}{\mathrm{u1}}$	0.15	0.15	0.15	$\max(1,\delta\mathrm{g})\frac{\delta1_{\mathrm{sg}}}{\delta1}\cdot\mathrm{r1_g}$	80	$t\mathrm{kg}+t\mathrm{g1}-t\mathrm{k1}$	63.7							
Rm	pRm→amR1	$\frac{\mathrm{CE_{km}}}{\mathrm{CE_{k1}}}\mathrm{um}\cdot\mathrm{cm}\cdot\mathrm{CE_{m1}}\frac{\mathrm{u1_{sm}}}{\mathrm{u1}}$	3.0	2.2	2.8	$\delta\mathrm{m}\frac{\delta1_{\mathrm{sm}}}{\delta1}$	6.7	$t\mathrm{km}+[t\mathrm{m1}\downarrow,t\mathrm{m1}\uparrow]-t\mathrm{k1}$	[9.5, 26.7]							
Rf	pRf→afTf	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\mathrm{uf}\cdot\mathrm{cf}\cdot\mathrm{CE_{f1}}\cdot\frac{\mathrm{u1_{sf}}}{\mathrm{u1}}$	0	0	0	-		-								
Lf	pLf→pR1	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\frac{\mathrm{bf}}{\mathrm{u1}}\mathrm{ef}\cdot\mathrm{Lf}\cdot\mathrm{u1}_{\varepsilon\mathrm{f}}$	0.39	0.38	0.39	$\frac{\delta1_{\varepsilon\mathrm{f}}}{\delta1}$	100.0	$t\mathrm{kf}+2t\mathrm{fk}_{\varepsilon\mathrm{f}}+t\mathrm{f1}_{\varepsilon\mathrm{f}}-t\mathrm{k1}$	39.6							
ΛLf	pΛLf→pR1	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\frac{\mathrm{bf}}{\mathrm{u1}}\mathrm{if}\cdot\Lambda\mathrm{f}\cdot\mathrm{L}\Lambda\mathrm{f}\cdot\mathrm{u1}_{\Lambda\mathrm{f}}$	0.32	0.32	0.32	$\frac{[1,\delta1_{\varepsilon\mathrm{f}}]}{\delta1}$	[6.7, 100.0]	$t\mathrm{kf}-t\mathrm{k1}+[2t\mathrm{fk}_{\Lambda\mathrm{f}}+t\mathrm{f1}_{\Lambda\mathrm{f}},2t\mathrm{fk}_{\varepsilon\mathrm{f}}+t\mathrm{f1}_{\varepsilon\mathrm{f}}]$	[8.2, 39.6]							

Effect	**Topology**	$\boldsymbol{p_r}$	$\boldsymbol{p_r(N)}$	p_s	p_f	$\boldsymbol{q_{PP}}$		$\boldsymbol{T_{PP}}$		$\boldsymbol{q_{AP}}$		$\boldsymbol{\Sigma_q}$	$\boldsymbol{\eta_q}$	$\boldsymbol{T_{AP}}$		$\boldsymbol{d}$
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
BLf	pBLf→afTf ↘ pR1	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\frac{\mathrm{bf}}{\mathrm{u1}}\mathrm{if}\cdot\mathrm{Bf}\cdot\mathrm{LBf}$ $\cdot\mathrm{u1_{Bf}}$	1.3	1.3	1.3	$\frac{[1,\delta 1_{\varepsilon f}]}{\delta 1}$	[6.7, 100.0]	$t\mathrm{kf}-t\mathrm{k1}$ $+[2t\mathrm{fk_{Bf}}+t\mathrm{f1_{Bf}},$ $2t\mathrm{fk}_{\varepsilon\mathrm{f}}+t\mathrm{f1}_{\varepsilon\mathrm{f}}]$	[8.2, 39.6]							
γgk	pRγg→agR1 ↘ γPk→pR1	$\mathrm{CE_{kg}}\cdot\mathrm{ug}\cdot\gamma\mathrm{cg}\cdot\gamma\mathrm{eg}$ $\cdot\gamma\mathrm{tgk}\cdot\mathrm{QE_k}$	0	0	0	$\max(1,\delta\mathrm{g})\frac{\delta 1_{\mathrm{sg}}}{\delta 1}$ $\cdot\mathrm{r1_g}$	80	$t\mathrm{kg}+t\mathrm{g1}-t\mathrm{k1}$	63.7	1	100.0	180	44	$t\mathrm{kg}$ $+[t\gamma\mathrm{gk},t\gamma\mathrm{gk}_{\uparrow}]$	[63.9, 64.5]	[0.1, 0.8]
γg1	pRγg→agR1 ↘ γP1→pR2	$\frac{\mathrm{CE_{kg}}}{\mathrm{CE_{k1}}}\frac{\mathrm{ug}}{\mathrm{u1}}\gamma\mathrm{cg}\cdot\gamma\mathrm{eg}$ $\cdot\gamma\mathrm{tg1}\cdot\mathrm{QE_1}\cdot\mathrm{CE_{12}}$ $\cdot\mathrm{u2_{p1}}$	0	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	$t\mathrm{kg}+t\gamma\mathrm{g1}-t\mathrm{k1}$ $+t12_{\mathrm{p1}}-t12$	40.0	$\max(1,\delta_{\mathrm{g}})$ $\cdot\frac{\delta 1_{\mathrm{sg}}}{\delta 1}\cdot\mathrm{r1_g}$	80	87	92.4	$t\mathrm{kg}+t\mathrm{g1}-t\mathrm{k1}$	63.7	23.8
γmk	pRγm→smTm ↘ γPk→pR1	$\mathrm{CE_{km}}\cdot\mathrm{um}\cdot\gamma\mathrm{cm}$ $\cdot\gamma\mathrm{em}\cdot\gamma\mathrm{tmk}\cdot\mathrm{QE_k}$	0.010	0.010	0.010	1	100.0	$t\mathrm{km}+t\gamma\mathrm{mk}$	21.3							
γsk	pRγs→asT3 ↘ γPk→pR1	$\mathrm{CE_{ks}}\cdot\mathrm{us}\cdot\gamma\mathrm{cs}\cdot\gamma\mathrm{es}$ $\cdot\gamma\mathrm{tsk}\cdot\mathrm{QE_k}$	$1.1\cdot10^{-9}$	$1.1\cdot10^{-9}$	$1.1\cdot10^{-9}$	1	100.0	$t\mathrm{ks}+t\gamma\mathrm{sk}$	25.5							
γfk	pRγf→afTf ↘ γPk→pR1	$\mathrm{CE_{kf}}\cdot\mathrm{uf}\cdot\gamma\mathrm{cf}\cdot\gamma\mathrm{ef}$ $\cdot\gamma\mathrm{tfk}\cdot\mathrm{QE_k}$	0.010	0.010	0.010	1	100.0	$t\mathrm{kf}+t\gamma\mathrm{fk}$	21.5							
γ1k	pRγ1→a1R2 ↘ γPk→pR1	$\gamma\mathrm{c1}\cdot\gamma\mathrm{e1}\cdot\gamma\mathrm{t1k}\cdot\mathrm{QE_k}$ $\cdot\mathrm{CE_{k1}}\cdot\mathrm{u1}$	0.046	0.045	0.045	1	100.0	0	0	1	100.0	200	50	$t\gamma\mathrm{1k}+t\mathrm{k1}$	25.5	25.5
γ11	pRγ1→a1R2 ↘ γP1→pR2	$\gamma\mathrm{c1}\cdot\gamma\mathrm{e1}\cdot\gamma\mathrm{t11}\cdot\mathrm{QE_1}$ $\cdot\mathrm{CE_{12}}\cdot\mathrm{u2_{p1}}$	0.017	0.017	0.016	1	100.0	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	106.6	6.2	$2[t\gamma 12,t\gamma 1k]$ $+t12_{\mathrm{p1}}-t12$	[1.1 1.7]	[1.1 1.7]
γ1f	pRγ1→a1R2 ↘ γPf→afTf	$\gamma\mathrm{c1}\cdot\gamma\mathrm{e1}\cdot\gamma\mathrm{t1f}\cdot\mathrm{QE_f}$ $\cdot\mathrm{CE_{f1}}\cdot\mathrm{u1_{pf}}$	0	0	0	-		-								
pRγ3k	pRγ3→a3R4 ↘ γPk→pR1	$\mathrm{CE_{k3}}\cdot\mathrm{u3_p}\cdot\gamma\mathrm{c3}\cdot\gamma\mathrm{e3}$ $\cdot\gamma\mathrm{t3k}\cdot\mathrm{QE_k}$	$1.1\cdot10^{-6}$	$1.1\cdot10^{-6}$	$1.1\cdot10^{-7}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}}}{\delta 3}$	0.80	$t\mathrm{k3}-t\mathrm{k1}-t12-t23$	−6.2	1	100.0	100.8	99.2	$t\mathrm{k3}+t\gamma\mathrm{3k}$	26.5	32.7
pRγ31	pRγ3→a3R4 ↘ γP1→pR2	$\frac{\mathrm{CE_{k3}}}{\mathrm{CE_{k1}}}\frac{\mathrm{u3_p}}{\mathrm{u1}}\gamma\mathrm{c3}\cdot\gamma\mathrm{e3}$ $\cdot\gamma\mathrm{t31}\cdot\mathrm{QE_1}\cdot\mathrm{CE_{12}}$ $\cdot\mathrm{u2_{p1}}$	$1.3\cdot10^{-6}$	$9.6\cdot10^{-7}$	$1.3\cdot10^{-7}$	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p}}}{\delta 3}$	0.80	$t\mathrm{k3}-t\mathrm{k1}-t12-t23$	−6.2	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	7.4	89	$t\mathrm{k3}+t\gamma\mathrm{31}-t\mathrm{k1}$ $+t12_{\mathrm{p1}}-t12$	1.9	8.1

Effect	**Topology**	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
γ2k	pR1→a1Rγ2→a2R3 ↘ γPk→pR1	$1-(1-\gamma c2\cdot\gamma e2\cdot\gamma t2k\cdot \mathrm{QE_k}\cdot\mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta 1}$	0.27	0.27	0.27	1	100.0	0	0	1	100.0	200	50	$t\gamma 2\mathrm{k}+t\mathrm{k1}+t12$	29.9	29.9
γ21	pR1→a1Rγ2→a2R3 ↘ γP1→pR2	$1-(1-\gamma c2\cdot\gamma e2\cdot\gamma t21\cdot \mathrm{QE_1}\cdot\mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta 1}$	0.26	0.26	0.25	1	100.0	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	106.6	6.2	$t\gamma 21+t12_{\mathrm{p1}}$	5.3	5.3
γ22	pR1→a1Rγ2→a2R3 ↘ γP2→pR3	$1-(1-\gamma c2\cdot\gamma e2\cdot\gamma t22\cdot \mathrm{QE_2}\cdot\mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta 1}$	$4.2\cdot10^{-5}$	$4.2\cdot10^{-5}$	$3.1\cdot10^{-6}$	1	100.0	0	0	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p2}}}{\delta 3}$	0.8	100.8	0.8	$2t\gamma 23+t23_{\mathrm{p2}}-t23$	0.6	0.6
γ3k	a1R2→a2Rγ3→a3R4 ↘ γPk→pR1	$1-(1-\gamma c3\cdot\gamma e3\cdot\gamma t3k\cdot \mathrm{QE_k}\cdot\mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta 1\delta 2}$	0.48	0.48	0.48	1	100.0	0	0	1	100.0	200	50	$t\gamma 3\mathrm{k}+t\mathrm{k1}+t12+t23$	32.7	32.7
γ31	a1R2→a2Rγ3→a3R4 ↘ γP1→pR2	$1-(1-\gamma c3\cdot\gamma e3\cdot\gamma t31\cdot \mathrm{QE_1}\cdot\mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta 1\delta 2}$	0.58	0.58	0.55	1	100.0	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	106.6	6.2	$t\gamma 31+t12_{\mathrm{p1}}+t23$	8.1	8.1
γ32	a1R2→a2Rγ3→a3R4 ↘ γP2→pR3	$1-(1-\gamma c3\cdot\gamma e3\cdot\gamma t32\cdot \mathrm{QE_2}\cdot\mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta 1\delta 2}$	1.1	1.1	0.080	1	100.0	0	0	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p2}}}{\delta 3}$	0.8	100.8	0.8	$t\gamma 32+t23_{\mathrm{p2}}$	3.3	3.3
γ4k	a2R3→a3Rγ4→a4R5 ↘ γPk→pR1	$1-(1-\gamma c4\cdot\gamma e4\cdot\gamma t4k\cdot \mathrm{QE_k}\cdot\mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta 1\delta 2\delta 3}$	1.2	1.1	1.2	1	100.0	0	0	1	100.0	200	50	$t\gamma 4\mathrm{k}+t\mathrm{k1}+t12+t23+t34$	36.3	36.3
γ41	a2R3→a3Rγ4→a4R5 ↘ γP1→pR2	$1-(1-\gamma c4\cdot\gamma e4\cdot\gamma t41\cdot \mathrm{QE_1}\cdot\mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta 1\delta 2\delta 3}$	3.6 (0.037)	3.6	3.4	1	100.0	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2}$	6.6	106.6	6.2	$t\gamma 41+t12_{\mathrm{p1}}+t23+t34$	11.7	11.7
γ42	a2R3→a3Rγ4→a4R5 ↘ γP2→pR3	$1-(1-\gamma c4\cdot\gamma e4\cdot\gamma t42\cdot \mathrm{QE_2}\cdot\mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta 1\delta 2\delta 3}$	6.7 (0.069)	6.6	0.49	1	100.0	0	0	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{p2}}}{\delta 3}$	0.8	100.8	0.8	$t\gamma 42+t23_{\mathrm{p2}}+t34$	6.9	6.9

Effect	**Topology**	$\boldsymbol{p_r}$	$\boldsymbol{p_r(N)}$	p_s	p_f	$\boldsymbol{q_{PP}}$		$\boldsymbol{T_{PP}}$		$\boldsymbol{q_{AP}}$		$\boldsymbol{\Sigma_q}$	$\boldsymbol{\eta_q}$	$\boldsymbol{T_{AP}}$		$\boldsymbol{d}$
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
γ**5k**	a3R4→a4Rγ5→a5R6 ↘ γPk→pR1	$1-(1-\gamma c5\cdot\gamma e5\cdot\gamma t5k\cdot \mathrm{QE_k}\cdot \mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta1\delta2\delta3\delta4}$	0.20	0.20	0.20	1	100.0	0	0	1	100.0	200	50	$t\gamma 5\mathrm{k}+t\mathrm{k}1+t12+\ldots+t45$	39.7	39.7
γ**51**	a3R4→a4Rγ5→a5R6 ↘ γP1→pR2	$1-(1-\gamma c5\cdot\gamma e5\cdot\gamma t51\cdot \mathrm{QE_1}\cdot \mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta1\delta2\delta3\delta4}$	1.1	1.1	1.1	1	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{\mathrm{p1}}}{\delta2}$	6.6	106.6	6.2	$t\gamma 51+t12_{\mathrm{p1}}+t23+t34+t45$	15.1	15.1
γ**52**	a3R4→a4Rγ5→a5R6 ↘ γP2→pR3	$1-(1-\gamma c5\cdot\gamma e5\cdot\gamma t52\cdot \mathrm{QE_2}\cdot \mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta1\delta2\delta3\delta4}$	3.6	3.5	0.27	1	100.0	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\mathrm{p2}}}{\delta3}$	0.8	100.8	0.8	$t\gamma 52+t23_{\mathrm{p2}}+t34+t45$	10.4	10.4
γ**6k**	a4R5→a5Rγ6→a6R7 ↘ γPk→pR1	$1-(1-\gamma c6\cdot\gamma e6\cdot\gamma t6k\cdot \mathrm{QE_k}\cdot \mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta1\delta2\ldots\delta5}$	0.80	0.79	0.80	1	100.0	0	0	1	100.0	200	50	$t\gamma 6\mathrm{k}+t\mathrm{k}1+t12+\ldots+t56$	43.7	43.7
γ**61**	a4R5→a5Rγ6→a6R7 ↘ γP1→pR2	$1-(1-\gamma c6\cdot\gamma e6\cdot\gamma t61\cdot \mathrm{QE_1}\cdot \mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta1\delta2\ldots\delta5}$	4.4 (0.045)	4.3	4.1	1	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{\mathrm{p1}}}{\delta2}$	6.6	106.6	6.2	$t\gamma 61+t12_{\mathrm{p1}}+t23+\ldots+t56$	19.2	19.2
γ**62**	a4R5→a5Rγ6→a6R7 ↘ γP2→pR3	$1-(1-\gamma c6\cdot\gamma e6\cdot\gamma t62\cdot \mathrm{QE_2}\cdot \mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta1\delta2\ldots\delta5}$	14 (0.15)	13	1.0	1	100.0	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\mathrm{p2}}}{\delta3}$	0.8	100.8	0.8	$t\gamma 62+t23_{\mathrm{p2}}+t34+t45+t56$	14.4	14.4
γ**7k**	a5R6→a6Rγ7→a7R8 ↘ γPk→pR1	$1-(1-\gamma c7\cdot\gamma e7\cdot\gamma t7k\cdot \mathrm{QE_k}\cdot \mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta1\delta2\ldots\delta6}$	0.14	0.14	0.14	1	100.0	0	0	1	100.0	200	50	$t\gamma 7\mathrm{k}+t\mathrm{k}1+t12+\ldots+t67$	47.8	47.8
γ**71**	a5R6→a6Rγ7→a7R8 ↘ γP1→pR2	$1-(1-\gamma c7\cdot\gamma e7\cdot\gamma t71\cdot \mathrm{QE_1}\cdot \mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta1\delta2\ldots\delta6}$	0.78 (0.0079)	0.78	0.74	1	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{\mathrm{p1}}}{\delta2}$	6.6	106.6	6.2	$t\gamma 71+t12_{\mathrm{p1}}+t23+\ldots+t67$	23.2	23.2
γ**72**	a5R6→a6Rγ7→a7R8 ↘ γP2→pR3	$1-(1-\gamma c7\cdot\gamma e7\cdot\gamma t72\cdot \mathrm{QE_2}\cdot \mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta1\delta2\ldots\delta6}$	2.5 (0.026)	2.5	0.19	1	100.0	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\mathrm{p2}}}{\delta3}$	0.8	100.8	0.8	$t\gamma 72+t23_{\mathrm{p2}}+t34+\ldots+t67$	18.4	18.4

Effect	**Topology**	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
γ**8k**	a6R7→a7Rγ8→a8R9 ↘ γPk→pR1	$1-(1-\gamma c8\cdot\gamma e8\cdot\gamma t8k\cdot \mathrm{QE_k}\cdot\mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta1\delta2\ldots\delta7}$	0.21	0.21	0.21	1	100.0	0	0	1	100.0	200	50	$t\gamma 8\mathrm{k}+t\mathrm{k}1+t12+\ldots+t78$	51.8	51.8
γ**81**	a6R7→a7Rγ8→a8R9 ↘ γP1→pR2	$1-(1-\gamma c8\cdot\gamma e8\cdot\gamma t81\cdot \mathrm{QE_1}\cdot\mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta1\delta2\ldots\delta7}$	1.2	1.1	1.1	1	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{\mathrm{p1}}}{\delta2}$	6.6	106.6	6.2	$t\gamma 81+t12_{\mathrm{p1}}+t23+\ldots+t78$	27.2	27.2
γ**82**	a6R7→a7Rγ8→a8R9 ↘ γP2→pR3	$1-(1-\gamma c8\cdot\gamma e8\cdot\gamma t82\cdot \mathrm{QE_2}\cdot\mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta1\delta2\ldots\delta7}$	3.7 (0.038)	3.7	0.27	1	100.0	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\mathrm{p2}}}{\delta3}$	0.8	100.8	0.8	$t\gamma 82+t23_{\mathrm{p2}}+t34+\ldots+t78$	22.4	22.4
γ**9k**	a7R8→a8Rγ9→a9R10 ↘ γPk→pR1	$1-(1-\gamma c9\cdot\gamma e9\cdot\gamma t9k\cdot \mathrm{QE_k}\cdot\mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta1\delta2\ldots\delta8}$	0.057	0.057	0.057	1	100.0	0	0	1	100.0	200	50	$t\gamma 9\mathrm{k}+t\mathrm{k}1+t12+\ldots+t89$	55.8	55.8
γ**91**	a7R8→a8Rγ9→a9R10 ↘ γP1→pR2	$1-(1-\gamma c9\cdot\gamma e9\cdot\gamma t91\cdot \mathrm{QE_1}\cdot\mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta1\delta2\ldots\delta8}$	0.32	0.31	0.30	1	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{\mathrm{p1}}}{\delta2}$	6.6	106.6	6.2	$t\gamma 91+t12_{\mathrm{p1}}+t23+\ldots+t89$	31.2	31.2
γ**92**	a7R8→a8Rγ9→a9R10 ↘ γP2→pR3	$1-(1-\gamma c9\cdot\gamma e9\cdot\gamma t92\cdot \mathrm{QE_2}\cdot\mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta1\delta2\ldots\delta8}$	1.0	1.0	0.076	1	100.0	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\mathrm{p2}}}{\delta3}$	0.8	100.8	0.8	$t\gamma 92+t23_{\mathrm{p2}}+t34+\ldots+t89$	26.5	26.5
γ**10k**	a8R9→a9Rγ10→a10Ra ↘ γPk→pR1	$1-(1-\gamma c10\cdot\gamma e10\cdot\gamma t10k\cdot \mathrm{QE_k}\cdot\mathrm{CE_{k1}}\cdot \mathrm{u1})^{\delta1\delta2\ldots\delta9}$	0.13	0.13	0.13	1	100.0	0	0	1	100.0	200	50	$t\gamma 10\mathrm{k}+t\mathrm{k}1+t12+\ldots+t910$	59.8	59.8
γ**10,1**	a8R9→a9Rγ10→a10Ra ↘ γP1→pR2	$1-(1-\gamma c10\cdot\gamma e10\cdot\gamma t101\cdot \mathrm{QE_1}\cdot\mathrm{CE_{12}}\cdot \mathrm{u2_{p1}})^{\delta1\delta2\ldots\delta9}$	0.69 (0.0070)	0.69	0.66	1	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{\mathrm{p1}}}{\delta2}$	6.6	106.6	6.2	$t\gamma 101+t12_{\mathrm{p1}}+t23+\ldots+t910$	35.3	35.3
γ**10,2**	a8R9→a9Rγ10→a10Ra ↘ γP2→pR3	$1-(1-\gamma c10\cdot\gamma e10\cdot\gamma t102\cdot \mathrm{QE_2}\cdot\mathrm{CE_{23}}\cdot \mathrm{u3_{p2}})^{\delta1\delta2\ldots\delta9}$	2.2 (0.023)	2.2	0.17	1	100.0	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\mathrm{p2}}}{\delta3}$	0.8	100.8	0.8	$t\gamma 102+t23_{\mathrm{p2}}+t34+\ldots+t910$	30.5	30.5

Effect	**Topology**	$\boldsymbol{p_r}$	$\boldsymbol{p_r(N)}$	p_s	p_f	$\boldsymbol{q_{PP}}$		$\boldsymbol{T_{PP}}$		$\boldsymbol{q_{AP}}$		$\boldsymbol{\Sigma_q}$	$\boldsymbol{\eta_q}$	$\boldsymbol{T_{AP}}$		$\boldsymbol{d}$
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
γak	a9R10→a10Rγa γPk↙→pR1	$1-(1-\gamma\text{ca}\cdot\gamma\text{ea}\cdot\gamma\text{tak}\cdot\text{QE}_\text{k}\cdot\text{CE}_\text{k1}\cdot\text{u1})^{\delta1\delta2\ldots\delta10}$	0.21	0.20	0.20	1	100.0	0	0	1	100.0	200	50	$t\gamma\text{ak}+t\text{k1}+t12+\ldots+t10\text{a}$	60.6	60.6
γa1	a9R10→a10Rγa γP1↙→pR2	$1-(1-\gamma\text{ca}\cdot\gamma\text{ea}\cdot\gamma\text{ta1}\cdot\text{QE}_1\cdot\text{CE}_{12}\cdot\text{u2}_{\text{p1}})^{\delta1\delta2\ldots\delta10}$	1.0	1.0	0.99	1	100.0	0	0	$\frac{1}{\delta1}\frac{\delta2_{\text{p1}}}{\delta2}$	6.6	106.6	6.2	$t\gamma\text{a1}+t12_{\text{p1}}+t23+\ldots+t10\text{a}$	36.0	36.0
γa2	a9R10→a10Rγa γP2↙→pR3	$1-(1-\gamma\text{ca}\cdot\gamma\text{ea}\cdot\gamma\text{ta2}\cdot\text{QE}_2\cdot\text{CE}_{23}\cdot\text{u3}_{\text{p2}})^{\delta1\delta2\ldots\delta10}$	3.4	3.3	0.25	1	100.0	0	0	$\frac{1}{\delta1\delta2}\frac{\delta3_{\text{p2}}}{\delta3}$	0.8	100.8	0.8	$t\gamma\text{a2}+t23_{\text{p2}}+t34+\ldots+t10\text{a}$	31.3	31.3
L1E1	pL1→pE1→pR2	$\frac{\text{b1}}{\text{u1}}\text{e1}\cdot\text{L1}\cdot\text{b1}_{\varepsilon1}\cdot\text{e1}_{\varepsilon1}\cdot\text{E1}_{\varepsilon1}\cdot\text{u2}_{\text{p}\varepsilon1^2}$	0.032	0.026	0.031	$\frac{1}{\delta1}\frac{\delta2_{\text{p}\varepsilon1^2}}{\delta2}$	8.1	$2t1\text{k}_{\varepsilon1}+t12_{\text{p}\varepsilon1^2}-t12$	47.1							
L1L1	pL1→pL1→pR1	$\frac{\text{b1}}{\text{u1}}\text{e1}\cdot\text{L1}\cdot\text{b1}_{\varepsilon1}\cdot\text{e1}_{\varepsilon1}\cdot\text{L1}_{\varepsilon1}\cdot\text{u1}_{\varepsilon1^2}$	0.016	0.016	0.016	$\frac{\delta1_{\varepsilon1^2}}{\delta1}$	100.0	$2t1\text{k}_{\varepsilon1}+2t1\text{k}_{\varepsilon1^2}$	100.5							
E1L2R1	pE1→pL2→pR1	$\frac{\text{b1}}{\text{u1}}\text{e1}\cdot\text{E1}\cdot\text{b2}_{\text{p}\varepsilon1}\cdot\text{e2}_{\text{p}\varepsilon1}\cdot\text{L2}_{\text{p}\varepsilon1}\cdot\text{A1}_{\varepsilon1\varepsilon2}\cdot\text{u1}_{\varepsilon1\varepsilon2}$	0.087	0.086	0.086	$\frac{\delta1_{\varepsilon1\varepsilon2}}{\delta1}$	100.0	$t12_{\text{p}\varepsilon1}+t21_{\text{p}\varepsilon1\varepsilon2}$	2.3							
L1E2	pL1→pR1→r1R2 ↘s1E2→s1R3	$\frac{\text{b1}}{\text{u1}}\text{e1}\cdot\text{L1}\cdot\text{u1}_{\varepsilon1}\cdot\left(1-(1-\text{b2}\cdot\text{e2}\cdot\text{E2}\cdot\text{u3}_{\text{s1}\varepsilon2})^{\delta1_{\varepsilon1}}\right)$	0.17 (0.0019)	0.17	0.025	$\frac{1}{\delta1\delta2}\frac{\delta3_{\text{s1}\varepsilon2}}{\delta3}$	0.86	$2t1\text{k}_{\varepsilon1}+t23_{\text{s1}\varepsilon2}-t23$	48.7	$\frac{\delta1_{\varepsilon1}-1}{\delta1}$	93.3	94.2	99.1	$2t1\text{k}_{\varepsilon1}$	50.3	1.6
L1L2	pL1→pR1→r1R2 ↘s1L2→s1R2	$\frac{\text{b1}}{\text{u1}}\text{e1}\cdot\text{L1}\cdot\text{u1}_{\varepsilon1}\cdot\left(1-(1-\text{b2}\cdot\text{e2}\cdot\text{L2}\cdot\text{u2}_{\varepsilon2})^{\delta1_{\varepsilon1}}\right)$	0.26 (0.0028)	0.25	0.24	$\frac{\delta1_{\varepsilon1}-1}{\delta1}$	93.3	$2t1\text{k}_{\varepsilon1}$	50.3	$\frac{1}{\delta1}\frac{\delta2_{\varepsilon2}}{\delta2}$	6.7	100.0	6.7	$2t1\text{k}_{\varepsilon1}+2t21_{\varepsilon2}$	59.0	8.7
L1E3	pL1→pR1→a1R2→r2R3 s2E3↙→s2R4	$\frac{\text{b1}}{\text{u1}}\text{e1}\cdot\text{L1}\cdot\text{u1}_{\varepsilon1}\cdot\left(1-(1-\text{b3}\cdot\text{e3}\cdot\text{E3}\cdot\text{u4}_{\text{s2}\varepsilon3})^{\delta1_{\varepsilon1}\delta2}\right)$	1.1 (0.026)	1.1	0	$\frac{1}{\delta1\delta2\delta3}\frac{\delta4_{\text{s2}\varepsilon3}}{\delta4}$	0.11	$2t1\text{k}_{\varepsilon1}+t34_{\text{s2}\varepsilon3}-t34$	48.0	$\frac{\delta1_{\varepsilon1}}{\delta1}-\frac{1}{\delta1\delta2}$	99.2	99.3	99.9	$2t1\text{k}_{\varepsilon1}$	50.3	2.3

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
L1L3	pL1→pR1→a1R2→r2R3 s2L3→s2R3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L1}\cdot\mathrm{u1}_{\varepsilon 1}$ $\cdot\left(1-(1-\mathrm{b3}\cdot\mathrm{e3}\cdot\mathrm{L3}\right.$ $\left.\cdot\,\mathrm{u3}_{\varepsilon 3})^{\delta 1_{\varepsilon 1}\delta 2}\right)$	1.0 (0.021)	1.0	0.082	$\frac{\delta 1_{\varepsilon 1}}{\delta 1}-\frac{1}{\delta 1\delta 2}$	99.2	$2t1\mathrm{k}_{\varepsilon 1}$	50.3	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\varepsilon 3}}{\delta 3}$	0.77	100.0	0.77	$2t1\mathrm{k}_{\varepsilon 1}$ $+2t32_{\varepsilon 3}$	55.7	5.4
E2L3R2	pR1→r1R2 s1E2→s1L3→s1R2	$1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{E2}$ $\cdot\mathrm{b3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{e3}_{\mathrm{s1}\varepsilon 2}$ $\cdot\mathrm{L3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{A2}_{\varepsilon 2\varepsilon 3}$ $\cdot\mathrm{u2}_{\varepsilon 2\varepsilon 3})^{\delta 1}$	0.37	0.36	0.35	$1-\frac{1}{\delta 1}$	93.3	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\varepsilon 2\varepsilon 3}}{\delta 2}$	6.7	100.0	6.7	$t23_{\mathrm{s1}\varepsilon 2}$ $+t32_{\mathrm{s1}\varepsilon 2\varepsilon 3}$	2.3	2.3
E2L3R3	pR1→r1R2 s1E2→s1L3→s1R3	$1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{E2}$ $\cdot\mathrm{b3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{e3}_{\mathrm{s1}\varepsilon 2}$ $\cdot\mathrm{L3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{A3}_{\mathrm{s1}\varepsilon 2\varepsilon 3}$ $\cdot\mathrm{u3}_{\mathrm{s1}\varepsilon 2\varepsilon 3})^{\delta 1}$	0	0	0	$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{s1}\varepsilon 2\varepsilon 3}}{\delta 3}$	0.86	$t23_{\mathrm{s1}\varepsilon 2}$ $+2t32_{\mathrm{s1}\varepsilon 2\varepsilon 3}-t23$	0.7							
E11L1	pE11→pL1→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E11}\cdot\mathrm{b1}_{\varepsilon 1}$ $\cdot\mathrm{e1}_{\varepsilon 1}\cdot\mathrm{L1}_{\varepsilon 1}\cdot\mathrm{u1}_{\varepsilon 1^2}$	$4.8\cdot10^{-5}$	$4.8\cdot10^{-5}$	$4.8\cdot10^{-5}$	$\frac{\delta 1_{\varepsilon 1^2}}{\delta 1}$	100.0	$[t11_{\downarrow}+t12_{\downarrow}-t12,0]$ $+2t1\mathrm{k}_{\varepsilon 1^2}$	[49.4, 50.3]							
L11L1	pL11→pL1→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L11}\cdot\mathrm{b1}_{\varepsilon 1}$ $\cdot\mathrm{e1}_{\varepsilon 1}\cdot\mathrm{L1}_{\varepsilon 1}\cdot\mathrm{u1}_{\varepsilon 1^2}$	$9.7\cdot10^{-5}$	$9.6\cdot10^{-5}$	$9.6\cdot10^{-5}$	$\frac{\delta 1_{\varepsilon 1^2}}{\delta 1}$	100.0	$[0,t11_{\uparrow}+t12_{\uparrow}-t12]$ $+2t1\mathrm{k}_{\varepsilon 1^2}$	[50.3, 56.0]							
L1L11	pL1→pL11→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L1}\cdot\mathrm{b1}_{\varepsilon 1}\cdot\mathrm{e1}_{\varepsilon 1}$ $\cdot\mathrm{L11}_{\varepsilon 1}\cdot\mathrm{u1}_{\varepsilon 1^2}$	$9.7\cdot10^{-5}$	$9.6\cdot10^{-5}$	$9.6\cdot10^{-5}$	$\frac{\delta 1_{\varepsilon 1^2}}{\delta 1}[\mathrm{r1},1]$	[50, 100.0]	$2t1\mathrm{k}_{\varepsilon 1}+[0,t11_{\varepsilon 1\uparrow}$ $+t12_{\uparrow}-t12]$	[50.3, 56.0]							
L1sh1E1	pL1→pR1→r1R2 sh1E1→sh1R2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L1}\cdot\mathrm{u1}_{\varepsilon 1}$ $\cdot\left(1-(1-\mathrm{s1_h}\cdot\mathrm{E1}_{\mathrm{sh1}}\right.$ $\left.\cdot\,\mathrm{u2}_{\mathrm{sh1}})^{\delta 1_{\varepsilon 1}}\right)$	0.14	0.13	0.13	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{sh1}}}{\delta 2}$	6.9	$2t1\mathrm{k}_{\varepsilon 1}+t12_{\mathrm{sh1}}$ $-t12$	49.1	$\frac{\delta_{\mathrm{sh}}1_{\varepsilon 1}}{\delta 1}$	100.0	106.9	93.5	$2t1\mathrm{k}_{\varepsilon 1}$	50.3	1.1
sh1E1-L2R1	pR1→r1R2 sh1E1→sh1L2→sh1R1	$1-(1-\mathrm{s1_h}\cdot\mathrm{E1}_{\mathrm{sh1}}$ $\cdot\mathrm{b2}_{\mathrm{sh1}}\cdot\mathrm{e2}_{\mathrm{sh1}}\cdot\mathrm{L2}_{\mathrm{sh1}}$ $\cdot\mathrm{A1}_{\mathrm{sh1}\varepsilon 2}\cdot\mathrm{u1}_{\mathrm{sh1}\varepsilon 2}$ $\cdot\mathrm{c1}_{\mathrm{sh1}\varepsilon 2})^{\delta 1}$	0.093	0.092	0.091	$\frac{\delta_{\mathrm{sh}}1}{\delta 1}$	100.0	0	0	$\frac{1}{\delta 1}\max(1,$ $\delta 1_{\mathrm{sh1}\varepsilon 2})$	14	114	12	$t12_{\mathrm{sh1}}$ $+t21_{\mathrm{sh1}\varepsilon 2}$	6.5	6.5
L2sh2E2	pR1→r1R2 s1L2→s1R2→r2R3 sh2E2→sh2R3	$\left(1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{L2}\right.$ $\left.\cdot\,\mathrm{u2}_{\varepsilon 2})^{\delta 1}\right)$ $\cdot\left(1-(1-\mathrm{s2_h}\cdot\mathrm{E2}_{\mathrm{sh2}}\right.$ $\left.\cdot\,\mathrm{u3}_{\mathrm{sh2}})^{\delta 2_{\varepsilon 2}}\right)$	1.2 (0.014)	1.2	0.10	$1-\frac{1}{\delta 1}$	93.3	0	0	$\frac{1}{\delta 1}\frac{\delta_{\mathrm{sh}}2_{\varepsilon 2}}{\delta 2}$ $\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\mathrm{sh2}}}{\delta 3}$	6.3 0.78	99.6 94.1	6.3 0.83	$2t21_{\varepsilon 2}$ $2t21_{\varepsilon 2}$ $+t23_{\mathrm{sh2}}-t23$	8.7 8.0	8.7 8.0

Effect	Topology	p_r	$p_r(N)$	p_s	p_f	q_{PP}		T_{PP}		q_{AP}		Σ_q	η_q	T_{AP}		d
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
sh2E2-L3R2	pR1→a1R2→r2R3 sh2E2→sh2L3→sh2R2	$1-(1-\mathrm{s2_h}\cdot\mathrm{E2_{sh2}}\cdot\mathrm{b3_{sh2}}\cdot\mathrm{e3_{sh2}}\cdot\mathrm{L3_{sh2}}\cdot\mathrm{A2_{sh2\varepsilon3}}\cdot\mathrm{u2_{sh2\varepsilon3}}\cdot\mathrm{c2_{sh2\varepsilon3}})^{\delta1\delta2}$	0.88	0.87	0.32	$1-\frac{1}{\delta1}+\frac{\delta_{sh}2}{\delta1\delta2}$	99.6	0	0	$\frac{1}{\delta1\delta2}\max(1,\ \delta2_{sh2\varepsilon3})$	1.2	100.8	1.2	$t23_{sh2}+t32_{sh2\varepsilon3}$	4.1	4.1
L3sh3E3	pR1→a1R2→r2R3 s2L3→s2R3→r3R4 sh3E3→sh3R4	$\left(1-(1-\mathrm{b3}\cdot\mathrm{e3}\cdot\mathrm{L3}\cdot\mathrm{u3_{\varepsilon3}})^{\delta1\delta2}\right)\cdot\left(1-(1-\mathrm{s3_h}\cdot\mathrm{E3_{sh3}}\cdot\mathrm{u4_{sh3}})^{\delta3_{\varepsilon3}}\right)$	5.2 (0.11)	5.1	0	$1-\frac{1}{\delta1\delta2}$	99.2	0	0	$\frac{1}{\delta1\delta2}\frac{\delta_{sh}3_{\varepsilon3}}{\delta3}$	0.73	100.0	0.73	$2t32_{\varepsilon3}$	5.4	5.4
										$\frac{1}{\delta1\delta2\delta3}\frac{\delta4_{sh3}}{\delta4}$	0.08	99.3	0.09	$2t32_{\varepsilon3}+t34_{sh3}-t34$	4.5	4.5
s1E11L1	pR1→r1R2 s1E11→s1L1→s1R1	$1-(1-\mathrm{c1}\cdot\mathrm{E11_{s1}}\cdot\mathrm{b1_{s1}}\cdot\mathrm{e1_{s1}}\cdot\mathrm{L1_{s1}}\cdot\mathrm{u1_{s1\varepsilon1}}\cdot\mathrm{c1_{s1\varepsilon1}})^{\delta1}$	$6.0\cdot10^{-7}$	$6.0\cdot10^{-7}$	$5.9\cdot10^{-7}$	$\frac{[\delta1-1,\delta_{sh}1]}{\delta1}$	[93.3, 100.0]	0	0	$\frac{1}{\delta1}\max\left(1,\ [\delta1_{s1\varepsilon1},\ \delta1_{sh1\varepsilon1}]\right)$	[6.7 14]	[100.0, 114]	[6.7, 12]	$[0, t11_{s(h)1\downarrow}+t12_{s(h)1\varepsilon1\downarrow}-t12]+2t1\mathrm{k_{s(h)1\varepsilon1}}$	[2.6, 5.9]	[2.6, 5.9]
s1L11L1	pR1→r1R2 s1L11→s1L1→s1R1	$1-(1-\mathrm{c1}\cdot\mathrm{L11_{s1}}\cdot\mathrm{b1_{s1}}\cdot\mathrm{e1_{s1}}\cdot\mathrm{L1_{s1}}\cdot\mathrm{u1_{s1\varepsilon1}}\cdot\mathrm{c1_{s1\varepsilon1}})^{\delta1}$	$3.0\cdot10^{-6}$	$3.0\cdot10^{-6}$	$2.9\cdot10^{-6}$	$\frac{[\delta1-1,\delta_{sh}1]}{\delta1}$	[93.3, 100.0]	0	0	$\frac{1}{\delta1}\max\left(1,\ [\delta1_{s1\varepsilon1},\ \delta1_{sh1\varepsilon1}]\right)$	[6.7 14]	[100.0, 114]	[6.7, 12]	$[0, t11_{s(h)1\uparrow}+t12_{s(h)1\varepsilon1\uparrow}-t12]+2t1\mathrm{k_{s(h)1\varepsilon1}}$	[2.6, 9.3]	[2.6, 9.3]
LfLf	pLf→pLf→pR1	$\frac{\mathrm{CE_{kf}}}{\mathrm{CE_{k1}}}\frac{\mathrm{bf}}{\mathrm{u1}}\mathrm{ef}\cdot\mathrm{Lff}\cdot\mathrm{bf_{\varepsilon f}}\cdot\mathrm{ef_{\varepsilon f}}\cdot\mathrm{Lf_{\varepsilon f}}\cdot\mathrm{u1_{\varepsilon f^2}}$	$1.2\cdot10^{-5}$	$1.2\cdot10^{-5}$	$1.2\cdot10^{-5}$	$\frac{\delta1_{\varepsilon f2}}{\delta1}$	100.0	$t\mathrm{kf}+2t\mathrm{fk_{\varepsilon f}}+2t\mathrm{fk_{\varepsilon f^2}}+t\mathrm{f1_{\varepsilon f^2}}-t\mathrm{k1}$	82.0							
L1EfR1	pL1→pEf→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L1fo}\cdot\mathrm{bf_{\varepsilon1}}\cdot\mathrm{ef_{\varepsilon1}}\cdot\mathrm{Ef_{\varepsilon1Lo}}\cdot\mathrm{u1_{\varepsilon1\varepsilon f}}$	0.014	0.014	0.014	$\frac{\delta1_{\varepsilon1\varepsilon f}}{\delta1}$	100.0	$t1\mathrm{f_{\varepsilon1}}-t12+[t\mathrm{f1_{\varepsilon1\varepsilon f\uparrow}}+t12_{\downarrow},\ t\mathrm{f1_{\varepsilon1\varepsilon f\downarrow}}+t12_{\uparrow}]$	[1.6, 5.6]							
L1L1L1	pL1→pL1→pL1→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L1}\cdot\mathrm{b1_{\varepsilon1}}\cdot\mathrm{e1_{\varepsilon1}}\cdot\mathrm{L1_{\varepsilon1}}\cdot\mathrm{b1_{\varepsilon1^2}}\cdot\mathrm{e1_{\varepsilon1^2}}\cdot\mathrm{L1_{\varepsilon1^2}}\cdot\mathrm{u1_{\varepsilon1^3}}$	$2.0\cdot10^{-6}$	$2.0\cdot10^{-6}$	$2.0\cdot10^{-6}$	$\frac{\delta1_{\varepsilon1^3}}{\delta1}$	100.0	$2t1\mathrm{k_{\varepsilon1}}+2t1\mathrm{k_{\varepsilon1^2}}+2t1\mathrm{k_{\varepsilon1^3}}$	150.9							

Effect	**Topology**	$\boldsymbol{p_r}$	$\boldsymbol{p_r(N)}$	p_s	p_f	$\boldsymbol{q_{PP}}$		$\boldsymbol{T_{PP}}$		$\boldsymbol{q_{AP}}$		$\boldsymbol{\Sigma_q}$	$\boldsymbol{\eta_q}$	$\boldsymbol{T_{AP}}$		$\boldsymbol{d}$
		Formula	[% (1)]	[%]	[%]	Formula	[% pe]	Formula	[ns]	Formula	[% pe]	[% pe]	[% pe]	Formula	[ns]	[ns]
E1L2-L2R1	pE1→pL2→pL2→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{L2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{A2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{L2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{A1}_{\varepsilon 1\varepsilon 2^2}\cdot\mathrm{u1}_{\varepsilon 1\varepsilon 2^2}$	0	0	0	$\frac{\delta 1_{\varepsilon 1\varepsilon 2^2}}{\delta 1}$	100.0	$t12_{\mathrm{p}\varepsilon 1}+2t21_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $+t21_{\mathrm{p}\varepsilon 1\varepsilon 2^2}$	4.6							
E1L2-L2R2	pE1→pL2→pL2→pR2	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{L2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{A2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{L2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{A2}_{\mathrm{p}\varepsilon 1\varepsilon 2^2}\cdot\mathrm{u2}_{\mathrm{p}\varepsilon 1\varepsilon 2^2}$	0	0	0	$\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p}\varepsilon 1\varepsilon 2^2}}{\delta 2}$	8.1	$t12_{\mathrm{p}\varepsilon 1}+2t21_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $+2t21_{\mathrm{p}\varepsilon 1\varepsilon 2^2}-t12$	1.4							
L1E2E3	pL1→pR1→r1R2 s1E2→s1E3→s1R4	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L1}\cdot\mathrm{u1}_{\varepsilon 1}$ $\cdot\Big(1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{E2}$ $\cdot\mathrm{b3}_{\mathrm{s1}\varepsilon 2}\cdot\mathrm{e3}_{\mathrm{s1}\varepsilon 2}$ $\cdot\mathrm{E3}_{\mathrm{s1}\varepsilon 2}$ $\cdot\mathrm{u4}_{\mathrm{s1}\varepsilon 2\varepsilon 3})^{\delta 1_{\varepsilon 1}}\Big)$	$5.2\cdot10^{-5}$	$5.2\cdot10^{-5}$	0	$\frac{1}{\delta 1\delta 2\delta 3}\frac{\delta 4_{\mathrm{s1}\varepsilon 2\varepsilon 3}}{\delta 4}$	0.11	$2t1\mathrm{k}_{\varepsilon 1}+t23_{\mathrm{s1}\varepsilon 2}$ $-t23+t34_{\mathrm{s1}\varepsilon 2\varepsilon 3}$ $-t34$	46.1	$\frac{\delta 1_{\varepsilon 1}-1}{\delta 1}$	93.3	93.4	99.9	$2t1\mathrm{k}_{\varepsilon 1}$	50.3	4.1
L1L2L3	pL1→pR1→r1R2 s1L2→s1R2→r2R3 s2L3→s2R3	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{L1}\cdot\mathrm{u1}_{\varepsilon 1}$ $\cdot\Big(1-(1-\mathrm{b2}\cdot\mathrm{e2}\cdot\mathrm{L2}$ $\cdot\mathrm{u2}_{\varepsilon 2})^{\delta 1_{\varepsilon 1}}\Big)$ $\cdot\Big(1-(1-\mathrm{b3}\cdot\mathrm{e3}\cdot\mathrm{L3}$ $\cdot\mathrm{u3}_{\varepsilon 3})^{\delta 1_{\varepsilon 1}\delta 2_{\varepsilon 2}}\Big)$	0.21 (0.0047)	0.21	0.017	$\frac{\delta 1_{\varepsilon 1}-1}{\delta 1}$	93.3	$2t1\mathrm{k}_{\varepsilon 1}$	50.3	$\frac{1}{\delta 1}\frac{\delta 2_{\varepsilon 2}}{\delta 2}-\frac{1}{\delta 1\delta 2}$	5.9	99.2	5.9	$2t1\mathrm{k}_{\varepsilon 1}$ $+2t21_{\varepsilon 2}$	59.0	8.7
										$\frac{1}{\delta 1\delta 2}\frac{\delta 3_{\varepsilon 3}}{\delta 3}$	0.77	94.1	0.81	$2t1\mathrm{k}_{\varepsilon 1}$ $+2t21_{\varepsilon 2}$ $+2t32_{\varepsilon 3}$	64.4	14.1
E1L22-L2R1	pE1→pL22→pL2→pR1	$\frac{\mathrm{b1}}{\mathrm{u1}}\mathrm{e1}\cdot\mathrm{E1}\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1}\cdot\mathrm{L22}_{\mathrm{p}\varepsilon 1}$ $\cdot\mathrm{b2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{e2}_{\mathrm{p}\varepsilon 1\varepsilon 2}$ $\cdot\mathrm{L2}_{\mathrm{p}\varepsilon 1\varepsilon 2}\cdot\mathrm{A1}_{\varepsilon 1\varepsilon 2^2}$ $\cdot\mathrm{u1}_{\varepsilon 1\varepsilon 2^2}$	$3.6\cdot10^{-6}$	$3.5\cdot10^{-6}$	$3.6\cdot10^{-6}$	$\frac{\delta 1_{\varepsilon 1\varepsilon 2^2}}{\delta 1}$	100.0	$t12_{\mathrm{p}\varepsilon 1}$ $+[t21_{\mathrm{p}\varepsilon 1\varepsilon 2^2},$ $t22_{\mathrm{p}\varepsilon 1\uparrow}$ $+t21_{\mathrm{p}\varepsilon 1\varepsilon 2^2\downarrow}$	[2.3, 2.7]							

Incident photon

For effects with photoelectron production outside of the cathode, merely γg and γm (see 2.3.6.3) can cause lateness. Only then the sum of photon and photoelectron transit times can be larger than for γk events due to distance or low field gradients. All other such effects can cause only earliness.
However, if the photon is reflected once or more often from reflective surfaces back onto the cathode, this introduces small delays due to the photon transit times inside the PMT, e.g. from the single reflection effects γMmk and γMgk. Since γg and γm already yield only small contributions to the late part of the TTD, additional reflections for these effects were not considered.
Reflections can occur outside the PMT as well and introduce external time shift effects (not from interactions inside the photosensor). For example, in [275], a reflection on the far end of the light source fiber was suspected to introduce an artificial peak between L1 and the main peak. If a reflective light concentrator is placed around the PMT, this can broaden the main peak significantly. This is due to greatly different photon path lengths inside the concentrator between direct hits, single reflections and spiral paths (see 5.3.1 and p. 618).

γMgk

Due to the large area and favorable shape of the aluminized glass, γMgk (fig. 2.39c) with $p_c = 9.7\%$ is the strongest reflection-based effect *and* LP effect. Therefore multiple reflections between g and k with higher lateness will also be observable with notable probabilities, for example γMggMgk (the reflected ray passes the dynode mount at its side, which is not well explainable with 2D schematics), γMgkMkgMgk, and so on. These higher-order reflections play a large role in the trailing flank of the main peak, where they contributing pulse with regular charges (see fig. 2.47), but as discussed cannot solely explain the falling flank.
γMgk and its secondary effects will strongly depend on the position and angle of incidence. The prominence of these effects heightens the importance of surface scans for adequate sensor characterization.

γMmk

γMmk is shown in fig. 2.39c, has smaller p_c and lateness than γMgk due to the surface and geometry of the deck, and also shows a strong angular and positional dependence.
[170; p. 74] observed a 10% drop in response when a light spot, which was moving over the deck, disappeared in the aperture to d1. The effect was stronger for red light, which could be due to the lower cathode absorbance at these wavelengths and the reflectance spectrum of the dynode mount. This measurements shows the magnitude of the influence of reflected photons on the overall response. Since the measurements here were for integral illumination, the strength of the effect is expected to be smaller, which was considered in the calculation of p_r.
Due to the far lower probability and the closeness to $T_r = 0$, secondary reflective effects, while expected, will have less importance.

Since γMgk, γMmk and their higher-order effects yield consistent contributions to the TTD for the same illumination, they can be regarded as part of the CTS; the electronic CTS is no less position- and angle-dependent.

Interdynode scattering

d1-based

L1

For L1 – in the literature often dubbed "late pulses" – like for E1 the pk is elastically backscattered on d1, but towards the cathode [264, 275, 276, 301, 336, 345, 346] [170; pp. 435–439]. Due to a curved trajectory or small energy losses in d1, the scattered pk does not quite reach the cathode and the electric fields focus it back onto d1 (fig. 2.40d), where it produces a pulse with normal charge and a lateness of about $2\,t\mathrm{k1}$ which results from the extra loop ($2\,t1\mathrm{k}_{\varepsilon 1} = 50.25\,\mathrm{ns}$, $2\,t\mathrm{k1} = 50.21\,\mathrm{ns}$). The effect was first observed by [333] according to [170].

Wright [170; pp. 435–439] states that scattering occurs on d1 (L1) or metal structures near it (Lm). The latter was not studied here, but should have similar time and charge to L1. Only the probability factors will be different, so L1 and Lm can be subsumed in an "effective" L1 type.
Ianni et al. [336] attributed the L1 peak to Lf instead and based this on the fact that U_{k1} was stabilized by Zener diodes for the studied ET 9351 whereas the peak shifted with HV, and the claim that d1 were designed for 90° forward scattering. The latter is not true, since electrons are scattered with a Lambertian angular distribution [170] and the fields between d1 and d2 are too small to strongly affect pk. The former argument is invalid, since the grid is an accelerator at d3 potential in the 9351 with the used VD [80], which *does* change with HV and alters the field gradients between k and f and thus $t\mathrm{k1}$. In a previous publication [275] the same authors attributed the dominant peak to L1 and identified Lf as a bump on the left flank formed by ΛL1, as was observed here.

The probability of L1 depends on the backscatter coefficient $\mathrm{b1} = 30.7\%$, but is far lower than this number due to the small elastic scattering fraction $\mathrm{e1} = 25\%$, the low backward scattered fraction $\mathrm{L1} = 16.4\%$ and the exclusion of secondary effects through $\mathrm{u1}_{\varepsilon 1} = 69.3\%$. The product of these factors p_r is therefore only 1.3%, which is less than for E1 (2.5%). The ME is connected to the direction fraction L1, since it returns backscattered pk to d1, which otherwise would be lost.

As previously discussed, the value of the L1 peak time was calculated to lie slightly after the measured peak. The predicted $t1\mathrm{k}_{\varepsilon 1}$ was based on the measured $t\mathrm{k1}$ of γ1, so the measured value is *smaller*. The peak time with 48.8 ns actually is *larger* than twice the earliness of γ1 (−23.8 ns), which is due to the other transit time terms in $T_{\gamma 1}$. The latter was also reported by [301] for Hamamatsu R7081 PMTs.
In addition, in the TTD the L1 peak is far broader than for γk. Secondary reflections such as γMgk-pRk-L1 etc. contribute to the L1 cluster just as for the main peak, exactly as the CTS and electronic secondary effects. The broadening therefore must stem from other effects.
If the pk backscattering loop covers a large fraction of the k–d1 volume, this introduces T variations as for the CTS, but *doubly*, namely for the way to k and back to d1, and *in addition* to the CTS of the first crossing from k to d1.
Lateral momentum could smear out the transit times due to curved paths – the calculations are one-dimensional and cannot account for 3D trajectories. This explanation was put forth by [301]. The lateral momentum can also carry the pk to regions with stronger or lower field gradients and thus influence transit time.

Finally, the shorter observed transit time could be from small energy losses in d1 before or after elastic scattering; a reduction by 2.89 eV produces matching times and could result from several electron–phonon interactions, which are very frequent. If a pk is scattered with Rutherford scattering losses only, the scattered energy would suffice to reach k, which likely eliminates the cascade since the backscattering probability at small energies is very low. On the other hand, the analytic calculation only incompletely represent the fields, which could introduce deviations between the *measured* tk1 and $t1\mathrm{k}_{\varepsilon 1}$, and the lateral momentum can prevent a cathode hit for energies above E_{k1}. If the pk in fact is subject to phonon losses in d1, the energy loss would vary with the number of phonon interactions and the transmitted energy, which would result in a continuum of scattering energies and varying transit times. In this case, the right flank of L1 would be from interactions with no or nearly no losses and the peak would represent the most likely losses. Since the right flank of L1 is broader than that of γk, this cannot be the sole reason.
Summarizing, the large broadening of the L1 peak likely is from "backscattering time spread" and lateral momentum, but small losses in d1 cannot be excluded as an additional source without field simulations. The mismatch between predicted and observed time could be explained by these losses, but lateral momentum or model inaccuracies can cause it as well.

Regarding L1 charge, elastic scattering losses were found to be negligible, and $\delta 1$ would even be unaltered by a loss of 3 eV since the SEY is flat in this region. This is in agreement with the observed charge and also the findings of [336] for the ET 9351. The charge distribution for the L1 cluster in [301], however, showed a lowered L1 peak charge of 86% pe for the R7081. One explanation might be that the fields of this series favor L1 trajectories which have reduced $\delta 1$ gains due to edge effect.
For L1 thus there is almost no threshold dependence, similar to γk, since the charges are virtually identical.

L1 can be reliably fitted with a Gaussian on the left flank (red. $\chi^2 = 0.61$), resulting in a p_c of 1.26%. The complete L1 peak contains about 1.43% of pulses, which with regard to the above discussion is a more reliable estimate of the L1 probability.

[346] studied the dependence of L1 on pk emission position for Hamamatsu R12860 PMTs (20") and observed strong variations. The probability is heightened directly above d1, reduced in the dynode mirror axis (as in the cross section in fig. 2.39a) for pk coming from the d2 side and from the d1 side near the cathode brink, and slightly lower towards the brinks on the axis perpendicular to this. The reductions could be from the dynode mount blocking the escape vectors towards the cathode and the inclination of the d1 surface.

In the literature, the measured probability of L1 varies between about 2 and 3%, the lateness ranges between 48 and 110 ns [276, 336, 346] (see also 3.1).

ΛL1

ΛL1 is the same as L1, but with inelastic scattering without SE production on d1. The backscattered pk flies a shorter loop towards the cathode due to its smaller initial energy (in fig. 2.40d as L1 but with smaller loop) and the produced pulse thus

has a shorter lateness than L1 and is smaller. Since the covered distance towards k is larger for greater emission energy, the lateness increases with energy and thus charge, and a rising $q(T)$ tendency results.
This effect was studied by [264, 275, 301] with measured probabilities between 2 and 3.8%. Here $p_r = 1.0\%$ is predicted. The numbers from literature, however, include all effects between γk and L1 peaks, which are treated separately here.

BL1

BL1 is identical to ΛL1, but in this case *with* SE production on d1 before or after scattering (fig. 2.40d). This initiates a second pulse with reduced charge but regular time, which precedes the pulse created by the backward scattered pk on its return to d1. As for ΛL1 the loop is smaller than for L1 due to the reduced scattered energy.
This effect was discussed by [184; p. 77] [190, 288, 347] [170; pp. 435–439] [276].

As for ΛL1, a rising $q(T)$ dependence is found, which for BL1 was calculated explicitly and shown in the AP-CTD, CDD and FDD (see pp. 243 and 244).
While in the CDD in fig. 2.56 the obtained BL1 $q(d)$ curve can not be unambiguously distinguished from BL11 at very low delays, in the FDD (fig. 2.62) this ambiguity is resolved and it is clearly visible that BL1 occurs preferably for large delays above 29 ns and for small delays below about 20 ns, and in between only with reduced probability, if at all. This is also visible in the CDD between 10 and 16 ns at high charges.
This behavior was attributed to an energy-dependent efficiency of the focusing optics: The late APs have scattering energies similar to L1 and are effectively guided back to d1, for the small energies at short delays the pk has less lateral momentum and does not stray too far from d1, so it has less chances to get lost (except on the attracting grid, which has a small cross-section, however). In between, the lateral momentum is large enough to stimulate losses.
While it would be expected that B-scattered pk have reduced emission probabilities from d1 when they penetrate deeper into the dynode, this seems to not be the case, since low-energy BL1 are observed as well and the distribution of the inelastically backscattered primaries in fig. 2.12a is more or less flat. This probability reduction for lower emission energies, however, *is* the case for larger impact energies in fig. 2.12b, which leads to the conclusion that for PMT-typical impact energies the penetration depths of backscattered primaries are small compared to the interaction volume in the dynode, in agreement with fig. 2.13a. The low-delay BL1 cannot solely stem from high-energetic SEs, since sa1L1 only reaches a delay of 7 ns (s1L1 and sa1L1 curves should resemble BL1). Virtually no SEs are expected for even higher energies, while BL1 is visible up to 20 ns. Thus the BL1 $p(d)$ behavior results from the kinematics of backscattered primaries.
The observation of low-energy backscattered pk which are refocused onto L1 is in tension with [232], according to whom these should be focused efficiently onto d2. It could, however, be that this only applies to BE1 and BE11, and not the backward scattered variants BL1 and BL11. In the FDD at high delays an increase of BL1 probability with delay can be seen, at low delays unfortunately it is difficult to assess the $p_c(d)$ dependence against the background of other effects.

[170; pp. 435–439] reports BL1 with 10–40 ns delay in ET 9351 PMTs (8”). [276] studied 255 R5912-HQE with thresholds of 10% pe and observes BL1 in a range from

10 to 60 ns, where the charge fraction rises with delay. The peak is found around charge fractions of 0.85 at delays of 45 ns, earlier than the L1 peak at 58 ns. A typical PMT had a BL1 probability of $2.7 \pm 0.01\%$. For the R5912 analyzed here, TST predicts a p_c of 4.1%, charge fractions between 6.2 and 93.8%, and delays from 2.6 to 45.3 ns, which are in agreement with the measurements here and reported in [276]; the HQE and SQE variants should not have different BL1 values.

d2-based

L2, ΛL2, and BL2 are equivalent to the effects on d1 (fig. 2.42d), but far more likely due to the large number of s1 electrons, which can contribute ($p_c = 14$–45%). All effects produce one RP with slightly reduced charge and a small charge AP from backscattering.
The lateness is far smaller than for d1-effects due to the shorter distance and higher field gradients. The maximum delay thus might be estimated somewhat too large, which would match with the cluster found at 6 ns delay and around 7% pe.
The $p(T)$ dependence should favor medium and large emission energies due to kinematics, since low-energy s1 electrons can be caught by d3. For this reason a decrease of p_c towards smaller delays is expected for BL2 and ΛL2. The same argument applies to the d3- and d4-based effects.

d3-based

L3, ΛL3, and BL3 produce too small AP charges to be directly visible in the CDD. Nevertheless, their charge distribution extends above the threshold and their tremendous probability ($p_c = 12$–98%) in fact makes BL3 the strongest predicted AP effect and L3 the third-strongest. However, since they have such large p_r that $N > 1$ (1.2–4.9), they are present most of the time and contribute to the falling flank of the standard pulse shape. Thus they will be subtracted together with the RP in the pulse search routine of PEST, which should diminish detection probability. They *can* be detectable if one pulse more or less than on average is observed, which still very frequently is the case.

d4-based

L4, ΛL4, and BL4 with N ranging from 10.0 to 39.9 are already part of the average pulse shape – this is one of the reasons for the trailing flank after the peak (see 2.4). Thus subtraction of the RP effectively eliminates these effects, while the sum pulse of all identical interactions otherwise would make BL4 the most probable AP effect observable here (see fig. 2.54). Individual pulses or the variation from one pulse more or less cannot be resolved due to the minute charge.

Intradynode scattering

d1-based

L11

An L11 pulse occurs if a pk is scattered off d1 at a shallow angle and onto the same dynode at a point closer to the cathode, where it produces SEs (fig. 2.40f). The increased distance to d2 which the s1 electrons have to cross, together with the intradynode transit time leads to lateness, whose extent depends on emission angle

resp. second hit position. Due to the edge effect also a charge reduction compared to the regular charge follows. Together a falling $q(T)$ tendency results.

ΛL11

The losses in d1 reduce the emission energy, which prolongs the lateness and further reduces the charge down to one s1, which is required to produce a pulse. The $q(T)$ tendency is falling for both variables emission angle and emission energy. ΛL11 contributes to the Rm band at 5.5% pe due to dynode skipping and charge quantization.

BL11

BL11 (fig. 2.40f) shows similar behavior to ΛL11, but for its AP, since an RP is produced. From the same argumentation as for BE11 (see p. 286) would follow that the probability should drop for large delays, but the observation of low-energy BL1 suggests that this might not be the case.

d2-based

L22, ΛL22, and BL22 resemble their d1-counterparts, but with shorter lateness due to the smaller dimensions of d2. The charge is still large enough to be mostly observable and due to their considerable p_c of 1.8–7.1% they contribute to the most densely populated cluster in the AP distributions.

d3-based

L33, ΛL33, and BL33 are even more frequently observed (2.6–29%) despite charge suppression, since their charge tail still extends above the threshold. Due to similar delays as the d2-effects, this group also contributes to the cluster around 4 ns and 2% pe.

Scattered stage skipping

Backward scattered electrons can pass the scattering stage on their return path and directly hit the next dynode for a fraction of emission angles (effects not calculated). If for example a pk is emitted away from d2 with very shallow angle, it can happen that it hits d2 (L12) or even d3 (L13) when focused back (similar to p3 or the γ2k trajectory in fig. 2.43b). This stage skipping leads to a reduction in charge due to missed amplification steps and shifts the pulse earlier, just as for an E-effect. Thus only L12, L13, L23 are relevant, for later hit stages the charges are too low for observation.
For L12 the time should be a bit earlier than for L1E1 (47.1 ns), since the path is more direct. The charge would correspond to L1E1 (8.1% pe). It is therefore almost indistinguishable from L1E1 except for the slightly shorter lateness.
L13 has a time as for L1 (50.3 ns) plus the earliness of p3 (−6.2 ns), resulting in about 44.1 ns, and a charge like p3 (0.80% pe), which makes it undetectable since only one pulse is produced.
Finally, L23 will be slightly earlier than L2 (8.7 ns) plus E2 (−1.6 ns) with a charge as E2 (0.86%), which is detectable with reduced probability, since also an RP is present.

Secondary electron interdynode emission

d1-based – s1L1, sh1L1, sa1L1

It is questionable whether s1L1 (fig. 2.42a) can occur in this form considering that the fields from f and d2 strongly accelerate s1 electrons away from d1. Instead, this could lead to stage skipping (s1L1R2) with a short lateness due to the extra path – impossible to calculate without knowledge of the exact fields – and charge $1/\delta 1$. In the other case, if an s1 is emitted too close to the grid it could be captured by it and disappears (s1Tf). The potential difference U_{f1} likely is chosen that small (28 V) to avoid that the grid removes too many s1 electrons from the cascade. While a reduced U_{f1} thus could increase $\delta 1$, this would deteriorate CE and TTS since the grid acts as an accelerator. s1L11 and s1L1R2 are both similar in charge and time to s1L1(R1), which makes it difficult to distinguish from data which effect occurs. While s1L11 has a different $q(T)$ tendency, the transition to s1L1 would be fluent. s1L1 was first postulated in [335] (in fact also for the R5912), but the described behavior (double pulses with AP delays of 8–60 ns) originated mostly from BL1 and γAP instead of s1L1.

sh1L1 and sa1L1 are less susceptible to stage skipping due to their larger emission energies. Here the attractive potential of f could be more critical and reduce the probability of these effects.

d2-based – s2L2, sh2L2, sa2L2

s2L2 (including sh2L2) with $p_c = 18\%$ is expected very frequently. However, looking at the CDD (fig. 2.56), the region above s2L2 (which lies around 1.7 ns delay and 1% pe charge) is mostly empty, possibly since the fit routine cannot resolve small pulses that close to an RP. sh2L2 comes closer to detection, while sa2L2 has detectable charges and matching events.

d3-based – (s3L3), (sh3L3), {sa3L3}

s3L3 (including sh3L3) shows the typical combination of probability reduction through dropping charges and an extremely high number of trials, which results in a still high $p_c = 8.9\%$ despite being far below thresholds. Its $N = 9.2$ makes this a statistical part of the pulse shape and unobservable. sa3L3 occurs too seldom for detection.

Secondary electron intradynode emission

sx1E11 (s1E11 in fig. 2.42b) and sx3E33 can only produce late pulses, since the distance reduction to the next dynode is not large enough to compensate the extra intradynode transit time. The maximum lateness ranges from 0.0 (sa1E11; near-perfect compensation for all paths) to 0.7 ns (s3L33), so these pulses cannot be resolved against the natural time spread of SEs of the RP cascade. Furthermore, with $N = 3.1$ s3E33 already is a part of the pulse shape.
sxnLnn (s1L11 in fig. 2.42b) shows a falling $q(T)$ tendency compared to a rising one for sxnLn. For the intradynode effects, a smaller emission energy increases the transit time in the near-static field near the dynode surface, and a longer traversed distance leads to a charge reduction through edge effect for longer transit times. For interdynode L-emission on the other hand, for a smaller emission energy the sxn can run against the adverse fields only for a shorter distance and time before returning. For all sxnLnn effects the lateness starts at zero and extends to a maximum value

between 4.5 and 6.3 ns. s3L33 has a mean number of occurrences of $N = 12.3$ and is a fixed contribution to each pulse. The probability of the other effects shows the typical opposed influence of charge reduction and availability of primary particles.

Secondary electron stage skipping

Secondary electrons emitted in L-direction can also be subject to stage skipping (not calculated). The lateness from the backward loop and the earliness from stage skipping compete and only computation can show whether in sum an EP or LP results. For one-stage skipping and snLn(n+1), no earliness can occur, since the regular SEs in the cascade ($T_r = 0$) cannot be overtaken. For two-stage skipping or shn and san, which have more energy, matters are more complicated though. Only sx1L12, sx1L13, and sx2L23 produce measurable charges and shall be quickly analyzed.

s1L12 has a charge comparable to s1E2, which is part of the RP cascade ($\frac{1}{\delta 1} = 6.7\,\%$pe). The time is roughly equal to the lateness of s1L1 (2.6 ns), since $t12$ is unaltered.
sa1L12 has a charge like sa1E1 (7.1% pe), and a lateness shorter than 5.7 ns, resulting from sa1L1 (7.1 ns) plus sa1E1 (−1.4 ns).
sh1L12 will lie between the both.

The charge of s2L23 equals $^1/_{\delta 1 \delta 2}$ (0.77% pe), and its time will roughly correspond to s2L2 (1.0 ns).
sa2L23 should match the charge of sa2E2 (0.79% pe) and should be a bit earlier than $T_r = 2.0$ ns from sa2L2 (2.8 ns) plus sa2E2 (−0.8 ns) due to the more direct path.
Again, sh2L23 will have values between the both effects.

s1L13 has the same charge as E2 (0.86%), and arrives earlier than the sum of 1.0 ns of s1L1 (2.6 ns) and E2 (−1.6 ns) because of a more direct route to d3 than E2. This might come close to $T_r = 0$ and a fraction of events could actually be EPs.
sh1L13 has a charge similar to sh1E1E2 (0.86% pe) and a slightly smaller T_r than sh1L1 (5.9 ns) plus E2 (−1.6 ns).
sa1L13 will be somewhat earlier than the sum of sa1L1 (7.1 ns) and E2 (−1.6 ns) and its charge will be somewhat larger than sh1L13.
So of all observable L-type SE stage skipping effects only a fraction of s1L13 might produce early pulses, in all other cases a late pulse will result.

Aluminized glass

Rg

For Rg (fig. 2.39f) a pk is emitted from the cathode with large lateral momentum close to the aluminized glass. Due to the small field gradients in this region the pk can hit g. Since the impact energy is very low (only U_0pk, on average 1.1 eV), the SEY actually will consist of nearly only backscattered primaries. Therefore the electron emission probability from g was approximated by the backscatter coefficient bg. SE production was neglected since only electrons in the valence band, which is nearly empty, could be possibly knocked out with such low impact energies.
The lateness strongly depends on the exact fields and U_0 and thus likely is smeared in time and not discernible in the CTD or TTD, since the effect is not very strong ($p_c = 0.15\%$).

Deck

Rm

Rm (fig. 2.39f) is observed in both SPAX and LNGS data and was overestimated in probability (see p. 297), most likely due to too high δm, CE_{km} or CE_{m1}. A correlation between p_c and T similar to γm (see p. 266) seems likely. The area of the deck increases with distance to the aperture to d1, so the probability is expected to rise with the lateness; the influence of CE_{km}, and CE_{m1}, however, can modify this behavior. If the Rm is assumed to decrease for short lateness, the additional events between 9 and 17 ns must be from ΛL11, BEf, ΛEf, BLf and ΛLf.

Grid

[Rf]

In the R5912, no sf can escape the grid, so Rf (fig. 2.39f) is unobservable.
For a non-accelerator-type grid, this effect is possible, but charge and time strongly depend on U_{f1}. In general, this will introduce a lateness, since the sf have to be accelerated towards d1. The larger U_{f1} the higher the charge will be, since effectively another gain stage is added, which could even increase the total charge to above 100% pe.

Lf

Lf (fig. 2.40b) is not discernible against the fluorescence background in the TTD and CTD (figs. 2.46 and 2.47). However, in LNGS data (fig. 2.36) a perfect match in time is observed, albeit the probability is estimated too high by a factor of three, likely from a too high fill factor FF_f or direction fraction Lf.
[275] attributed the same bump in the slope of ΛL1 to Lf. [276] mentions Lf as a possible effect.

BLf

For BLf (fig. 2.40b) no SE can escape the grid, so the pulse from the backscattered pk becomes the PP, which is the only case where a B-type pulse enters the PP distributions (see figs. 2.46 and 2.47).

Luminescence-based

[γgk] and [γg1]

If $U_g = U_k$, the pk impact energy is U_0pk, which is not sufficient for production of a photon with 3 eV or more, so the probability of γgk and γg1 (fig. 2.43e) is zero. Even if U_g were larger and photons could be produced, the p_f would still be undetectably low due to other probability factors.

γmk

The effect γmk requires that only a photon and no (successfully amplified) sm is produced on pk impact on the deck. For the assumed $\delta m \approx 0.5$ (which possibly is too high, see pp. 297 and 323) about 39% of pk which hit the deck produce no sm. If an sm is produced, a Rγmk results (not calculated), which is a combination of Rm (T_r = 9.5–26.7 ns) and γmk (21.3 ns). The photonic pulse is the AP, if the Rm pulse

is earlier, and the PP, if Rm is very late (see fig. 2.46). The delay of the γAP lies between 0 and 11.8 ns, the delay of the Rm-AP ranges from 0 to 5.4 ns. This results in two curves in the AP plots, both with low probabilities, however.

{γsk}

γsk (fig. 2.16 shows Rs) can only occur if there is a light shield which prevents light emission from reaching the gap between d1 and m and hitting k via external paths (see fig. 2.43c; here the cylinder around the dynodes acts as an effective light shield). Whether an additional light shield as in fig. 2.16 is installed in the R5912 is not known to the author, but the expected probability would be far too low for observation anyway.

γfk

γfk (fig. 2.43d) produces no SEs on the grid, so the pulse started by the photon on the cathode becomes the PP.
Unfortunately, neither in SPAX nor LNGS data the effect can be clearly verified or excluded due to nearby larger effects, the dark count floor and too low statistics.
However, this effect was observed in [336] for the ET 9351 (8") thanks to an enormous amount of statistics (2200 PMTs). γfk was attributed to a peak at 20.2 ns lateness, which had a standard deviation of 1.07 ns – in agreement with the main peak – and a probability of $5.8 \cdot 10^{-4}$. For the R5912 (also 8") TST predicts a $p_c = 1.0 \cdot 10^{-4}$ at 21.5 ns. The probability could be estimated too low for several factors, and thus is in agreement with data within errors. The transit times match as well as can be expected between different series. This corroborates the existence of γfk.

[γ1f]

Also γ1f is unobservable, since no sf can escape. For non-accelerator grids, it would create a small pulse with short lateness.

γnk, γn1, γn2

At all dynodes the impact of electrons can create photons as a side-effect. This can be from various types of luminescence (bremsstrahlung, cathodoluminescence, photoluminescence; see 2.6.1). The photons can in principle hit all stages, but only hits on the cathode or the first two dynodes result in charges large enough for detection and are studied further. The emission of such photons from d1 is shown in fig. 2.43a, from d2 in fig. 2.43b and from d10 in fig. 2.43c. For photons from d10, the cathode can be hit either due to reflections inside the dynode chain on the dynodes (internal paths) or outside the dynode mount on the glass and other reflective structures (external paths). The box-and-line design actually greatly facilitates internal paths through the arrangement of the dynodes and the gap in between; in fact, d2 is visible from d10, so a γ10k can reach k with as few as two reflections (on d2 and d1) if it is emitted within the right solid angle.

The effects studied here are for the most frequent case, when a photon is created in an RP, but photonic afterpulses can occur for nearly all TSEs – whenever there is

enough energy to allow photon creation (about 3 eV) and continuation of the mother pulse.

The charge for the γAP which results from γnk is always the same, namely 100% pe – the same as for an incident photon. For γn1 it is always $\frac{1}{\delta 1}\frac{\delta 2_{\mathrm{p1}}}{\delta 2} = 6.6\%$ pe, and for γn2 the charge is $\frac{1}{\delta 1 \delta 2}\frac{\delta 3_{\mathrm{p2}}}{\delta 3} = 0.76\%$ pe.

For bremsstrahlung sometimes larger energies than for a photon in the visible range can occur, since bremsstrahlung can carry away up to the total impact energy. Nevertheless, only one photoelectron can be created *directly*. In the photoeffect the photon is absorbed, which makes creation of further photoelectrons impossible. Alternatively, Compton scattering and subsequent photoeffect could take place, but for the energies relevant here the cross-section for Compton scattering is very small, and the photosensitive surfaces are very thin, so the generation of two photoelectrons through this effect is negligible.
However, this does *not* prevent the single produced photoelectron, which can have high energies, from scattering on other electrons in the material, which can receive sufficient energy to enter the vacuum. These scattered electron thus constitute effective *additional* photoelectrons. The emission of 2 and even 3 photoelectrons was observed for VUV photons by [192, 193]. For electron energies found in PMTs, the occurring bremsstrahlung has its peak in the VUV range (10–200 nm) or even in the region of soft X-rays ($\approx$0.2–10 nm). In addition, bremsstrahlung was found to be the dominant light emission process in PMTs (apart from unfavorable selection of the dynode mount material), see 2.6.1.
From this follows that for bremsstrahlung multiple-pe emission in fact is possible and expected. The γAP effects can thus have charges corresponding to multiples of their regular charge. This is observed in the CDD (fig. 2.56) for γ4k to γ6k and γ51 to γ71. However, this may also be caused by multiple photon production in the cascade, and both causes are difficult to distinguish.

For each target stage k, d1 and d2 a chain of γAP is observed in the delay axis (see DD and CDD in figs. 2.54 and 2.56). Later photon origin stages lead to greater delays, mainly due to the time required for the RP cascade to reach them. The photon transit time also increases for later stages but is far smaller ($t\gamma$10k = 0.73 ns vs. tk10 = 59.1 ns). For γnk the delay chain ranges from 25.5 (γ1k) to 60.6 ns (γak), and extends even beyond L1 (50.3 ns) as was already observed by [264] for anode glow (γAP from the last dynodes). For a fixed stage of photon origin, the delay becomes shorter if the stage, which is hit, is later (i.e. closer to the origin). This is due to the shorter transit time difference between cascades. While γ3k has a delay of 32.7 ns, it is only 8.1 ns for γ31, and 3.3 ns for γ32. The chains for γn1 and γn2 are thus shifted forward against γnk resp. γn1; for γn1 it ranges from 1.1 ns (γ11) to 36.0 ns (γa1), for γn2 from 0.6 ns (γ22) to 31.3 ns (γa2). This behavior will continue for γni with i $\geq$ 3, but these pulses are too small to be observed. They, however, will contribute to the average pulse shape. For these effects high occurrences N are expected due to the high number of electrons at later stages and the higher transmission probability between origin and target stage, since fewer stages have to be traversed.

Looking at the DD, in the delay range between about 11 and 30 ns the distribution is dominated by γn1 effects, afterwards by γnk. After about 10 ns delay no other effects except BL1 and BL11 have notable contributions. The comparably large

delays explain why γAP and BL1, but almost no earlier AP effects were reported in literature (see 2.5.2.2 and p. 318).
The probability factors for photon transport were coarsely tuned to align with the observed curve. Due to the favorable geometry and the large amount of SEs present in the cascade at later stages, the observed probabilities are large and in fact initially *rise* up to n ≈ 5 and only slowly decrease for later stages.
For γ10n and γan, external paths were included in the calculations, which increases their probability versus γ9n. External paths should also occur for earlier stages through gaps in the dynode mount and will raise their probabilities accordingly. Since the probability is slightly underestimated for most points this would move the predicted values even closer to data.
For γan, the traversed distance to k/d1/d2 actually is shorter than for γ10n, and the same number of reflections is required, since the anode lies before d10. Because more electrons are in the cascade at the anode than at d10, therefore γan has higher probability than γ10n.
The oscillating probabilities for even/uneven starting stages results from dynode geometry. Looking for example at fig. 2.43d, for even dynodes (d4, d6, ...) d2 is directly visible which allows paths with low numbers of reflections and raises p_c, while for uneven dynodes d2 is occluded. Judging from data, the assumed geometry apparently is not quite correct, since no strong oscillation is observed, γ51 shows a higher probability than γ41 and γ61, and the probability of γ5k is equal to γ4k and γ6k.
For γnk p_c starts at 0.045% (γk1), reaches 1.1% for γ4k and drops again to 0.057% for γ9k. γn1 reaches a predicted peak probability of 4.1% for γ61, but for γ51 even higher values are observed. The calculated peak for γn2 is at n = 6 with $p_c = 1.0\%$, which results from the large $p_r = 14\%$ seeing as the charge is already below h_f.
In total, for all γnk effects a p_c of 3.7% is expected, for γn1 13.2% and for γn2 2.8%. The measured total γAP rate is even a bit higher (see p. 380), since the predicted probabilities are slightly underestimated on average. A total detectable probability of more than 19.7% makes the γAPs an effect group, which has a substantial impact on the charge of an event in a detector, and which cannot safely be ignored.

γ11

For γ11 (fig. 2.43a) reflections on k and d2 were considered; the inside of the dynode mount should have lower reflectance resp. transport factors and was neglected. Since the cathode is oriented towards d1 and its origin of sphere lies on d1, in principle up to all photons reflected on the cathode could hit d1. The orientation of d2 is also favorable for reflections back to d1. The photon can also hit d1 directly without being reflected on another structure, which was not included in the calculations and will increase p_c (currently 0.016%), but it still would be covered by many stronger effects. Direct hits can actually even cause earliness as for E11.

γ22

γ22 (fig. 2.43a) has far smaller solid angles for self-reflection than γ11, since d1 actually reflects photons away from d2 and towards k, and the orientation of d3 tends to reflect towards d4. This shows in the small $p_c = 3.1 \cdot 10^{-6}$, which is too low for observation. Again, direct hits on d2 were not included and could raise p_c, but a multitude of more dominant effects would still impede detection.

{pRγ3k}, {pRγ31}

These secondary effects occurring after a p3 are examples of an effect other than γk which causes γAPs. However, they have too small probabilities for observation due to the low p_c of p3.

$\gamma_p$gk

Phosphorescence from external photons in the glass would also be considered an LP and in principle be a part of the TTD. However, emission times can extend to beyond 100 µs and the effect is weak (see 2.6.1 and 2.5.2.4). If a dark count occurs between illumination and phosphorescence, it becomes indistinguishable whether the phosphorescence pulse is from a delayed PP or an AP.

Dual interactions

Interdynode scattering

L1L1

If after L1 the pk is elastically scattered backward another time due to the high b1, an L1L1 results with the double lateness of L1. The probability of 0.016% still allows detection in the LPP-TTD (fig. 2.46) and LPP-CTD (fig. 2.47).
L1L1 was observed by the author for the Hamamatsu R6594 as well, where it was far stronger due to a higher L1 probability.

E1L2R1, E2L3R2, [E2L3R3]

Shown in fig. 2.41d, E1L2R1 happens if a pk is elastically forward scattered off d1 onto d2, on which it is scattered back onto d1, where it produces SEs. The additional transit time of 2.3 ns makes this an LP, but otherwise the charge is as for an RP.
E2L3R2 is the same effect occurring one dynode later for an s1 instead of a pk, and produces an AP with a delay of 2.3 ns.
E2L3R3 is assumed to not be possible, since the large s1 energy makes it very unlikely that d2 can be circumvent on the return path, which is required for a return to d3.

Intradynode scattering

E11L1 (fig. 2.41b), L11L1 (fig. 2.41c) and L1L11 can shift L1 to slightly earlier or moderately later times.

Grid

L1EfR1

This effect (fig. 2.41e) is similar to E1L2R1, but the scattering back to d1 occurs on the grid, which due to its small fill factor reduces the probability ($p_c = 0.014\%$ vs. 0.086% for E1L2R1). Depending on which position on d1 is hit on the return path, the lateness varies between 1.6 and 5.6 ns. In the CTD these events contribute to the trailing flank at regular charges after γk.

2.3.6.5 Applicability of time shift theory

Looking back at the thorough discussion of the time, charge and probability of the effects constituting the TTD and CTD in the present section, and anticipating the one to follow for the DD, CDD and FDD of the EAPs (2.5.2.2) and iAPs (2.5.2.3) as well as for the voltage dependence and applicability to other PMT series (3.2.2), the following conclusion can be drawn: Despite a thorough test with data from multiple PMT types at different voltages and using 8 distinct distributions types to cover all aspects of the PMT response as completely as possible, it was not possible to falsify time shift theory within the error margins expected from the used approximations. The theory is thus recommended for general use.
Field simulations should further improve the accuracy of predictions and allow to better assess the applicability of the theory, respectively help to advance it.

The following properties of the theory suggest its correctness:

- **Surjective:** TST provides a sufficient description of observations; all clusters in all distributions have explanations (TTD, CTD; DD, CDD and FDD for EAPs and iAPs).
  While it was necessary to extend the theory multiple times based on observations (induction) by including additional effect groups, the core concepts remained unchanged. The theory also predicted effects which then were identified in data upon closer inspection (deduction).
  In particular, the calculated BL1 charge–delay and charge fraction–delay curves agree well with the CDD and FDD, while the prediction of BL1 appearance in the TTD for low thresholds is in fact often visible in literature (see below). Furthermore, the repetition of effects which are centered on γk around L1 in both theory and data indicates that the underlying processes are understood.

- **Serial (left-total):** All interactions were combined in all reasonable combinations, but no effects were predicted where no clusters are observed in data (within the resolving power of the fit routine).

- **Few input parameters:** A good description of data is achieved although only very few parameters were extracted from the measurement, and despite using analytic transit time calculations instead of field simulations. Most parameter values were taken from literature, calculated, or estimated, and all input parameters were fixed ab initio (except two values relevant for γAPs only).

- **Effects verified in literature:** Many effects resulting from TST were already described and measured independently in literature (see p. 209).

- **Agreement with independent data:** Distributions which contain events at the positions predicted by TST are also reported in literature for the TTD [160, 185, 190, 228, 264, 275, 290, 301, 306, 336, 347–352], CTD [160, 301, 347, 350], EAP DD [185, 190], EAP CDD [347], EAP FDD [276], iAP DD [228, 290, 292, 336, 351, 352] and iAP CDD [276, 350]. This shows that the measurements performed here with SPAX and PEST are not flawed and do not create artificial clusters through not well understood setup properties.

2.3.7 Dependences

Due to the central role of timing for PMT performance, the transit time, TTS, and TSEs are influenced by a large number of factors.

Mean number of photoelectrons per light pulse

As discussed on p. 169, photosensors such as PMTs and SiPMs in general observe Poisson statistics regarding the dependence of the TTS on the mean number of detected photons λ. This means that the TTS scales with $\frac{1}{\sqrt{\lambda}}$ [324, 353, 354]. That the time resolution decreases for more simultaneously detected photons is intuitively clear. All produced pulses are sampled independently from the TTD [136], so the sum pulse and the time when it exceeds the CFD threshold – which is used to determine T (see p. 155) – are increasingly averaged out with growing λ.
The TTS can, however, not become arbitrarily small in this manner, since the time resolution of the measuring setup imposes a limit [354]. For SiPMs measured with low thresholds, dark counts cause fake triggers, which also sets a lower limit on the TTS [354].

Wavelength

The TTS decreases with growing wavelength [170; p. 371]. This is likely due to the lower mean emission energy U_0, which reduces the maximum lateral component of the initial velocity, resulting in a smaller spot size. However, this only applies to monochromatic light. Illumination with a broader spectrum will *increase* the TTS due to the larger variation of U_0. The same argument applies for a wider spectral response of the sensor, assuming a broad-spectrum source. Thus a HQE cathode can *worsen* the TTS due to its enhanced sensitivity at large wavelengths [353].
The average transit time should decrease with smaller wavelength, since the mean U_0 is larger, which has a notable influence on the initial pk drift in the small extraction fields and thus decreases tk1.
Finally, the QE spectrum will be different from the cathode for other photosensitive materials inside the device. This applies to the dynodes and aluminum, but also to the grid and deck, which are unintentionally coated with cathode material and thus in a less controlled manner, which alters material properties (the relevant factors are summarized on p. 64).

Photocathode

The TTS depends on the shape of the cathode (CTS) and its material, which affects the initial momentum and angular distribution of the pk electrons (spot size and U_0 distribution). NEA materials will further introduce a small photoemission lag (see p. 156).

Collection efficiency

TTS and CE have a negative correlation, since the electron-optics can be optimized only to one of both. This was already addressed on p. 80.

[136] Assuming operation within the linear range.

Dynodes

The TTS also depends on the dynode types, the number of dynodes, and their material.

Linear-focused, mesh, and metal-channel dynodes have very low time resolutions [169; pp. 45–46, 50]. For the linear-focused design this is due to the transit time compensation described on p. 171, the mesh and metal channel layouts are proximity-focused and benefit from the short distances between dynodes. The circular focused layout is basically a linear focused arrangement, which is coiled up for compactness, and thus also results in a good TTS [170; p. 232]. Box-and-line dynodes – a hybrid between linear-focused and box-and-grid – offer a good compromise between good timing and high CE [169; p. 46]. The worst TTS results from venetian blind and box-and-grid layouts, since the extraction fields are low and no transit time balancing occurs [169; p. 50] [170; pp. 226–229, 234–235]. For venetian blind PMTs, pk electrons frequently skip the first dynode if the electron trajectory is parallel to the slats of d1, which results in p2-type undersized early pulses [170; p. 227].
A small TTS usually is coupled to a short total transit time T_0.

For the same total voltage V_t, a higher number of dynodes decreases the interstage voltages and thus raises the overall transit time T_0 and TTS. Even for equal interdynode voltages, an additional stage will increase T_0 due to the extra interstage transit time, and slightly increase the TTS.

Regarding material, for BeCu a higher interstage voltage can be applied than for Cs_3Sb, which lowers TTS and T_0. On the other hand, the SEY is smaller, so the interdynode jitter is larger due to statistical transit time fluctuations of the SEs.
For BeCu, Z_{eff} (6.7 for BeO at the surface [170; p. 217]) is far lower than for Cs_3Sb (54), which will result in greatly reduced backscattering and thus EP and LP probabilities, despite the higher impact energies.

High voltage and voltage divider

A higher voltage leads to shorter total transit times, reduced TTS and smaller earliness, lateness and delay for TSEs. It also increases the charge of effects due to higher impact gains, and their probabilities through kinematics and the number of particles. The extent of this influence on each of the mentioned properties depends on the voltage distribution and whether interstage voltages are stabilized.
A front-tapered voltage divider decreases tk1 and its systematic variations, which are the main source of the TTS (see (2.150)).

Pulse shape

As already discussed on p. 131, a short rise time or pulse width typically indicates a small TTS [170; p. 352], since a larger total transit time broadens the pulse shape (see 2.4).

Magnetic field

External magnetic fields such as the terrestrial field deflect the pk electrons from their designed trajectories, which reduces CE (for some regions on the cathode, which

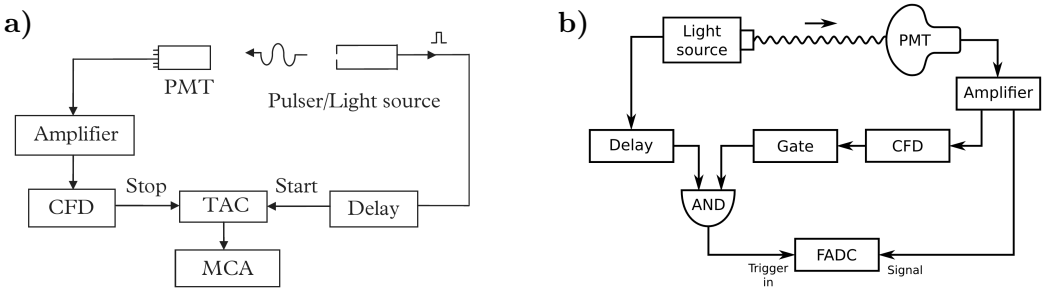


Figure 2.48: Time response measurement methods:
a) Start–stop method, from [170; p. 366].
b) FADC acquisition. Description see text.

normally have low CE, it can *increase* it) and influences transit times through the modified path and the pk filtering.
In the interdynode trajectories the field gradients are higher and the distances are shorter, so the magnetic field has less influence, but will play a small role in CE_{ij} and $t$ij as well.

2.3.8 Measurement

The start–stop method (fig. 2.48a) is the classic way to obtain the transit time distribution and transit time spread [170; p. 366] (confer p. 154). The trigger out signal from a light source is delayed and starts a time-to-amplitude converter (TAC). The PMT signal is amplified and fed to a constant fraction discriminator (CFD), which stops the TAC. The resulting time difference is the transit time of the PMT pulse plus an offset from cables and the processing time in the electronics of both branches. The TAC output can be used to fill a multichannel analyzer (MCA) to obtain the TTD, and its offset can be adjusted with the delay line. To reduce dead time, the start and stop channels can be switched, which requires additional delays. Alternatively, the trigger output can be replaced by a PMT which is illuminated with high occupancy by the same light source, which renders the jitter of the triggering PMT negligible. Its signal then is amplified and fed to a CFD, whose output starts the TAC. The measured PMT itself is illuminated with single photons (spe), e.g. through filters or a plane glass mirror.
[323] describes a start–stop TTS measurement for a multi-pe signal, which is made feasible by adding a charge analog-to-digital converter (ADC) to also obtain the pulse charge. The TTS for n detected photons is obtained through charge cuts. To improve results, overlaps of the n pe charges are eliminated by solving an equation system which results from the Poisson distribution of the number of coincident photons. From this the TTS of spe and n pe pulses can be calculated.

With a flash analog-to-digital converter (FADC), a coincidence unit is required if the acquisition of empty waveforms shall be avoided (fig. 2.48b). So as to preserve the start signal of the light source, its trigger-out signal is delayed until after the PMT signal was detected by a CFD and a gate was opened. Trigger and gate are connected to a coincidence (logic AND), which triggers acquisition. A second output

of the amplifier provides the PMT signal to the FADC input. The internal delay of the FADC is used to postpone acquisition until the light source trigger-out signal arrives at the trigger-in input. The stop time can be obtained from the PMT pulse in the recorded waveform, either with online or offline analysis (see p. 126).

Light sources

As discussed on p. 155, the jitter of the start signal depends on the light source. Thus the time spread of photon emission – and arrival – should be as small as possible, and the trigger out signal of the light source should not introduce additional jitter. This is usually achieved either through ps diode lasers or LEDs driven by a fast pulser.
The emission wavelength should be chosen close to the desired application of the PMT, usually around 400 nm. As stated on p. 92, for a clean spe measurement all light sources need to be attenuated so that the occupancy o amounts to only a few percent. This can be done either through neutral density filters (NDFs), or through distance if the light source has a sufficiently large viewing angle, as for LEDs or a laser in combination with a diffuser. Some devices also allow to tune intensity.

Picosecond diode lasers

Picosecond diode lasers have proven to be excellent light sources with regard to source jitter and normally can easily be coupled into a fiber, which allows to illuminate the photosensor at specific points on the active area and with controlled incident angles. Attaching the fiber end to a positioning frame allows to measure the uniformity of a sensor with regard to all properties (see below). The coherent, collimated light beam of lasers, however, requires the use of a diffuser and large distances to the sensor in order to homogeneously illuminate the photodetector. In addition, such a laser is costly. Currently available devices with wavelengths close to the PMT QE maximum, fiber coupling, and tunable intensity, which are externally triggerable, as of today cost around 10–13 k€. However, their low jitter of a few dozen ps (including trigger output jitter) allows to measure even photosensors with extremely low TTS such as SiPMs and LAPPDs.

Light emitting diode

Using an LED with a fast pulser is far more economical (especially if a self-built pulser is used) and usually has sufficiently low jitter to perform timing measurements for large area PMTs, if a fast LED with short decay times is chosen – which can be challenging to find. For good LED and drivers, jitters of several hundred ps can be achieved [355]. Kapustinsky's pulser is one example of an affordable, yet fast LED circuit [356, 357]. Other designs incorporate avalanche transistors [170; pp. 367, 572], or reflections on the end of a short cable (cable pulser) which helps to draw the bias voltage back to zero more quickly. Fast commercial LED–driver combinations are also available.
Temperature stabilization of LED circuits is advisable, as otherwise the light output (and thus o) will vary with time due to the strong temperature coefficient of the LED [170; p. 572].

Cherenkov light source [137]

Another alternative to a laser is to embed a low activity ^{90}Sr source in a transparent medium such as acrylic glass. The daughter nuclide ^{90}Y emits a 2.28 MeV β^- which will produce Cherenkov light in the acrylic. This Cherenkov converter is usually cylindrical, with a fast photosensor (fast PMT or SiPM) coupled to one end, which provides the start signal with negligible jitter due to the large number of detected photons. The measured PMT faces the side of the cylinder, into which a notch is cut to facilitate photon extraction, at a distance which reduces the occupancy sufficiently. The advantages of this light source are its intrinsically small jitter, simplicity, minimal cost and long-term stability. Drawbacks are long acquisition times and the non-Poissonian enhanced multi-pe tail, which results when the PMT is hit by the Cherenkov cone and which heightens the pollution from 2 pe events [190, 306].

Integral measurements

Measuring the average PMT response in a single measurement requires integral illumination.
While LEDs can have sufficiently large viewing angles to cover the whole cathode, the intensity quickly drops with angle unless the PMT is placed a large distance away. Better homogeneity is reached by using a diffuser.
For lasers a diffuser or beam shaper has to be used.

Diffusers

Diffusers can be reflective or transmissive. For an ideal diffuser a Lambertian $\cos(\theta)$ intensity distribution results. Therefore even for a perfect diffuser a sufficiently large distance between diffuser and a large PMT would have to be kept to achieve a set maximum intensity variation between center and brink of the photocathode.

Sandblasted glass is a very simple transmissive diffuser. Since there is only one scattering surface, the diffuser is still comparably inhomogeneous, however. This poses problems with a laser since it forms large laser speckles.

PTFE (Teflon) can be used as a diffuse reflector, but also is not very Lambertian. Spectralon on the other hand – sintered PTFE spheres with µm diameters – shows near perfect diffuse reflection for all wavelengths and is used as a white standard and in integrating spheres [358, 359] [170; pp. 81–82]. Nevertheless, a minimum thickness of about 8 mm is required for complete reflection, which in combination with the multiple scattering of photons inside the material will lead to small delays between entry and exit, which deteriorate timing. Furthermore, Spectralon was found to fluoresce [344].
Due to the large penetration depth, Spectralon could also be used as a transmissive diffuser with smaller thickness, which also would decrease the extent of the internal delay. First promising tests were carried out in the course of this work in collaboration with Hanna Kellermann and Jill Kaindl.

$BaSO_4$ is a reflective diffuser, also used as white standard [359] and can be applied as a paint to surfaces. TiO_2, while also very reflective in the visible range,

[137] Content based on [170; p. 368] [190]

has strong absorption in the near-UV [172; p. 6-3] and thus is less suited for PMT measurements.

The best Lambertian emission is reached with an integrating sphere; the larger the diameter, the better the homogeneity. However, this ruins timing and the integrating sphere can also fluoresce, if Spectralon is used for the walls.

Under no circumstances should paper be used as a diffuser for timing-relevant measurements despite its matt white surface, since fluorescing additives emitting in the blue are admixed to make the otherwise yellowish substance appear more white.

Uniformity measurements

If the fiber output or the LED is attached to a rotating frame or x-y stage, the uniformity of photosensor properties can be measured (QE, pulse efficiency, gain, TTS, TSEs, ...) [161, 185, 228, 346, 360–366].
It is reasonable to use a motorized positioning instead of a manual one for both ease of measurement, as the process then can be automated, and systematic error, since for manual repositioning the HV has to be turned off and on and the sensor is invariably illuminated with room light for a short duration:

- When turning the HV on again, some settling time is required before original behavior is restored, likely due to surface charges on the dynode mounts building up, and such an on-off cycle can cause small jumps in gain and other properties.

- The illumination of the photocathode causes a rise in dark count rate due to the excitation of long-lived states, which changes the background rate and can cause severe systematic errors. Because of the fast changes in DCR shortly after illumination, the experimentalist is well-advised to wait for at least a couple of minutes before undertaking measurements and even then should measure the DCR before and after the actual measurement to be able to interpolate and subtract the background during data taking. This reduces the systematic error, but cannot eliminate it.

Manual repositioning is time-consuming and introduces irreducible systematic errors, so for repeated measurements it is more economical to invest the work and cost of constructing an automatized setup.

Stability of light source and electronics

Variations in temperature and voltage can influence occupancy, threshold and charge, among others. It is recommended to stabilize the temperature of light source and electronics through Peltier elements or climate control and to let the electronics thermalize by operation for e.g. an hour before measurement. Thermalization is especially important for FADCs, since they have to be recalibrated after temperature changes. Placing the light source in an enclosure such as the dark box further helps to isolate it from day–night temperature fluctuations, but then it is even more important that thermal equilibrium is reached before measurement. In addition, ultra-stable power supplies should be used for light source and photosensor.

Noise

A common problem when the driver of a fast light source is placed inside the dark box is that electromagnetic waves from the pulse generation (e.g. the avalanche transistor discharge) are being picked up by the PMT like an antenna, likely in large part by the aluminum backside of the bulb. In this case it is necessary to shield the light source.
Large PMTs have sufficient noise pickup that when running a Fourier analysis on the output of the switched off PMT all surrounding radio sources become clearly visible, including WLAN, radio stations, non-public radio and whatever lurks in the laboratory next door. The root mean square (RMS) of the waveform baseline, and thus the minimal achievable instrumental threshold, greatly benefits from shielding the PMT electromagnetically.

Discriminators and pulse search

At some point of the measurement it is unavoidable to determine the presence and time of PMT pulses with a discriminator – even if the PMT output waveform is recorded for every light pulse and analyzed offline.

When looking for pulses, one wants to set the threshold close to the baseline to include as many underamplified pulses as possible in order to minimize PDE loss or to be able to study TSEs. This requires to use a low-threshold discriminator or to amplify the signal beforehand.

On the other hand, the inaccuracy of the threshold-crossing time must be minimized. Three effects play a role here [367].
For one, noise fluctuations on top of the signal can shift the detection time earlier or later, which introduces a jitter to the crossing time.
For a simple leading-edge discriminator (LED) the so-called time walk occurs, a systematic shift of the threshold-crossing time with pulse height: Since PMT pulse heights vary according to the pulse height distribution (PHD), a larger pulse crosses a fixed threshold earlier than a smaller pulse. Naturally, the resulting time walk lies in the order of the rise time of the PMT pulse shape. Since the time walk decreases for lower thresholds, this would require to even further lower the limit, but then one is at risk of running into noise.
Third, thermal drift can affect the discriminator circuit and thus alter the crossing-time, as already discussed on p. 334.

The solution is to use a constant-fraction discriminator (CFD), which eliminates time walk by triggering on a fixed fraction of the peak height for pulses which exceed a set threshold [170; p. 324].
To minimize noise jitter, the constant fraction (CF) trigger point should be set to the steepest point on the rising flank [170; pp. 356–358, 361]. Noise fluctuations on the baseline with standard deviation σ_n can shift the voltage values and cause the threshold to be passed earlier or later than the crossing time t without noise. If the rising flank of the pulse is approximated around t by a linear function with slope $dV(t)/dt$, the shift δt in threshold time from a voltage shift δV results to

$$\delta t \approx \frac{\delta V}{dV(t)/dt}$$

Substituting δV by σ_n yields the threshold time jitter from noise σ_t:

$$\sigma_t \approx \frac{\sigma_n}{dV(t)/dt} \tag{2.188}$$

Thus the noise jitter of the trigger time is minimized, if the threshold is set at the steepest point of the flank. Eq. (2.188) is an approximation, however, since the exact δt depends on higher order derivatives, too.

The time uncertainty of the PMT pulse can be further minimized by waveform acquisition and offline pulse search. In the latter a software CFD should be used, where the CF time is found by interpolation between datapoints. For an offline analysis, smoothing algorithms can be applied to minimize noise jitter, the steepest point can be extracted with high accuracy from an average pulse shape, and pileup as well as oscillations in the pulse shape can be recognized and eliminated.

For leading edge discriminators, the time walk can be reduced by a time–amplitude correction in offline analysis [323]. This requires to measure both time and charge of the pulse, either with a time-to-digital converter (TDC) and a charge ADC, or with an FADC. In the analysis the charge is binned and plotted against the mean time of the pulses contained in each charge bin. The resulting curve shows the time-walk and can be fitted by e.g. a polynomial, which allows to subsequently correct the time-walk event by event using the fit function.
While this allows to reduce the time walk, the correction does neither account for time shifts and charge reduction from TSEs (e.g. E1) nor for pileup from 2 pe pulses. This introduces systematic errors, since for small charges EPs shift the LED time forward and cause the time-walk correction to be too small. The CTD itself cannot be used to correct this effect because it is distorted by time-walk. Thus a CFD is clearly preferable.

So far, the stop signal, i.e. the PMT pulse was discussed. The start signal from the light source trigger in general is a logic pulse far larger than σ_n, so its noise jitter and time walk play a minor role and even a leading-edge discriminator can be used.

Transit time offset

In order to obtain absolute transit time values, the offset introduced by the setup has to be determined. The possible contributions are:

1. **Trigger transit time:** If a pulser is used to trigger the light source, the transit time of the trigger signal to the light source driver has to be measured.
2. **Photon emission delay:** The delay between trigger-in and light pulse emission in the light source can be retrieved from its data sheet.
3. **Photon flight time:** The photon flight time to the sensor can be calculated with knowledge of the optics in the setup.
4. **Signal transit time:** The time after sensor output until the electronics triggers data acquisition has to be measured.
5. **FADC delays:** For waveform acquisition the adjusted internal delay of the FADC has to be subtracted.

The trigger and signal transit times can be elegantly measured by connecting the trigger cable with the signal cable and sending a pulser signal through them. The same electronics as in the measurement is used, but without the photosensor and light source. A pulser mimics the PMT pulse shape as closely as possible, e.g. through a single square wave pulse with adjusted rise and fall times and a height corresponding to the mean pulse height. The cable which provides the trigger signal to the light source and the PMT output cable are connected by a cable of known transit time. With this arrangement a measurement is carried out with sufficient statistics to determine both transit time offset and jitter of the electronics by using the same data analysis procedure as for the signal. The transit time of the connection cable can be measured with a pulser and oscilloscope by placing a T-piece on the oscilloscope input with the pulser signal on one end and the measured cable with open or terminated end on the other. The time difference between pulser signal and reflected signal is the double transit time of the cable.

Setup jitter

The statistic and systematic variations of the abovementioned sources of offset introduce jitter. Since effects with probabilities as low as $\approx 10^{-3}\%$ can be observed in the TTD, even small contributions play a role and must be eliminated.

Regarding light source jitter, long trailing flanks in the emission time should be avoided. This occurs for slow LEDs, but also to some extent in ps-lasers with delays in the nanosecond regime. These emission time lags can play a role for sensitive TTD measurements with low DCR.

Optical jitter originates from variations in photon flight time. The optics can introduce delays for a fraction of photons due to sensor dimensions, fluorescence, reflections inside the optics and scattered light. Just as for emission delays, this can cause a trailing flank after each peak in the TTD. The distance between light source and sensor should be sufficiently large to reduce flight time differences from cathode brink to center. Using a reflective diffuser can double the available distance. To eliminate reflections from the NDFs placed before the light source, they should be tilted slightly towards the beam. As a measure against scattered light, the dark box surfaces should be covered with anti-reflective coating or matt black fabric such as molton. No materials which is hit by photons should fluoresce.

Each part of the electronics also introduces jitter, which is why as few components as possible should be used on the branch triggering acquisition. The summed electronics jitter can be measured together with the transit time offset as described there.

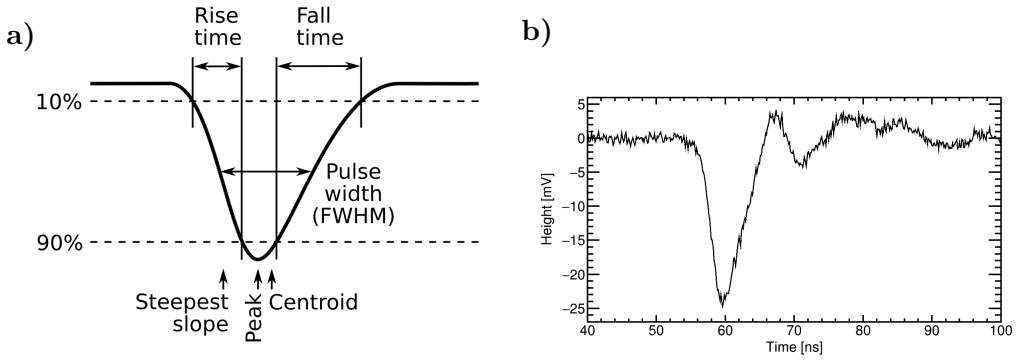


Figure 2.49: Pulse shape: a) Pulse shape parameters.
b) Typical γk pulse shape for a Hamamatsu R5912 with E7694-01 voltage divider at $g = 1.3 \cdot 10^7$, time axis corresponding to transit time. Noise and oscillations following the pulse are clearly visible.

2.4 Pulse shape

The pulse shape, also called single-electron time response [170; p. 352], describes the time distribution of the electrons in the cascade, which arrive at the anode, and produce a voltage spike at the output which can be measured by electronics. The schematic pulse shape of a photomultiplier (PMT) is shown in fig. 2.49a with its most important characteristics.

The rise time t_r is typically defined as the the duration during which the pulse increases from 10 to 90% peak height on the rising flank before the peak [169; pp. 51–52] [170; p. 369]. As visible in fig. 2.49a and the example of a real pulse in fig. 2.49b, the onset of the pulse as well as the region around the peak are quite flat. This makes height fractions closer to the baseline or the peak more susceptible to noise fluctuations, and raises the uncertainty of t_r. For particularly strong background or flat pulse shapes, sometimes t_r is also defined between 20 and 80%. Note that PMT pulses usually have negative amplitude, which does not affect the convention to call the region, which precedes the peak "rising" flank since the absolute pulse amplitude increases.
The rise time is the most crucial pulse shape parameter for timing applications [170; p. 369]. There is a weak correlation between t_r and the transit time spread (TTS); PMTs with short t_r tend to also have short TTS, but there are exceptions and measuring the pulse shape is no substitute for a transit time distribution (TTD) measurement.

The fall time t_f is the time which the pulse requires to subside from 90 to 10% peak height on the falling flank after the peak [169; pp. 51–52] [170; p. 369]. The fall time is typically 2–3 times as long as the rise time. The reason for this are interdynode transit time fluctuations of the secondary electrons (SEs) due to emission direction and energy, combined with small time shifting effects which statistically contribute to the falling flank, as discussed e.g. on p. 319.

The pulse width t_w is normally defined as the full width at half maximum (FWHM) of the peak and usually 1.5 to 2 times as long as the rise time [169, 170, 267].

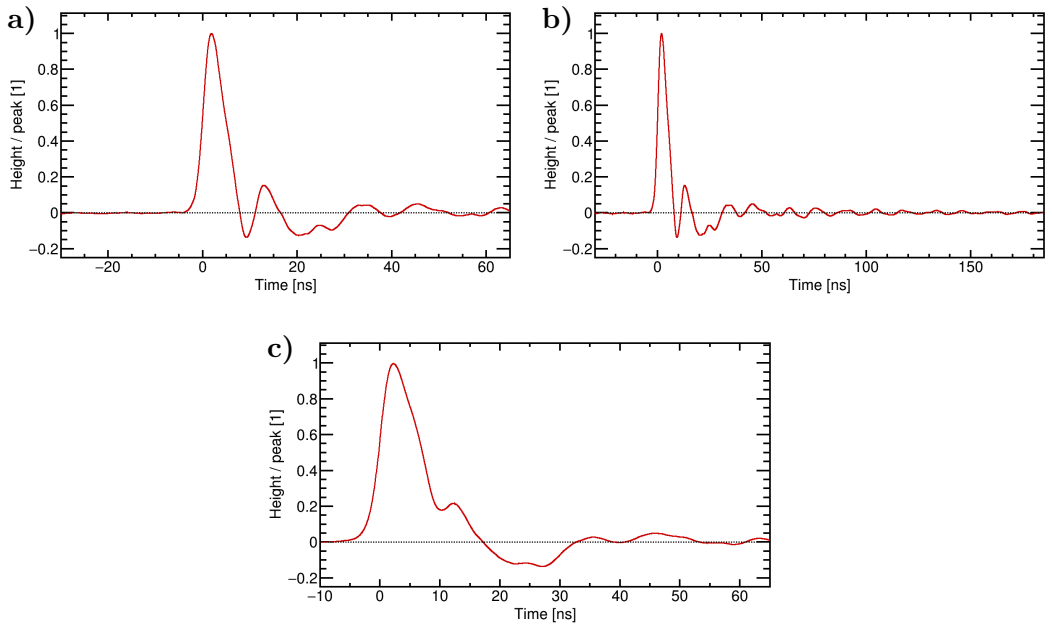


Figure 2.50: Average pulse shape (APS):
a) + b) APS for an R5912 (E7694-01 voltage divider, $g = 1.3 \cdot 10^7$) for waveforms containing only a primary pulse (PP), which was ascertained through stringent cuts. The origin of the time axis is set to the time of the steepest slope on the rising flank, which is at 57% height and 1.85 ns before the peak, and is used to determine the pulse time. The voltage divider produces severe oscillations, which complicate the reliable detection of afterpulses and remain at the 1% level up to over 180 ns after the peak.
c) The same APS without prior elimination of pileup and afterpulses (APs). The first overshoot disappears completely and the shape is greatly altered through the appearance of time shift effects (TSEs).

Statistic pulse width fluctuations are very small and practically negligible [267]. The relation between the FWHM and σ of a Gaussian distribution was given in (2.68). Sometimes also the full width at tenth of maximum (FWTM), which connects the start of the rise time with the end of the fall time, is encountered in literature [170; p. 354]. The FWTM gives a good idea of the time scale of the pulse and shall be termed pulse length t_l here. For a Gaussian shape with standard deviation σ, the FWTM follows from

$$\mathrm{FWTM} = 2\sqrt{2\ln(10)}\,\sigma \approx 4.29\,\sigma \tag{2.189}$$

Average pulse shape

The effect of noise on individual pulses can be eliminated by averaging over many pulses, which yields an average pulse shape (APS), see fig. 2.50.
A rough APS can be obtained with an oscilloscope or an flash analog-to-digital converter (FADC) by summing up all acquired pulses in a persistent plot (fig. 2.51). This already gives an idea of the pulse shape as well as the dominant time shift effects (TSEs) including early afterpulses (EAPs), and is helpful to determine the acquisition window.

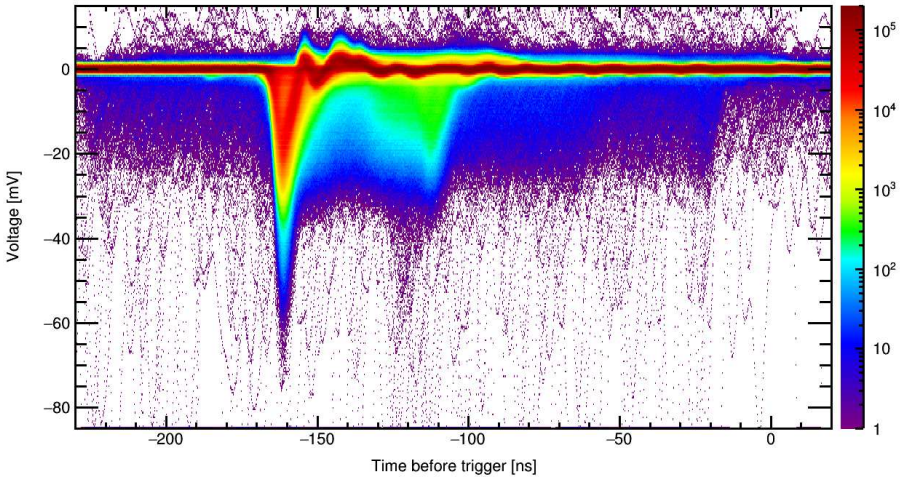


Figure 2.51: Persistent plot adding up $5 \cdot 10^6$ recorded waveforms in time–charge bins (0.25 ns, 0.4 mV), for R5912 at $g = 1.3 \cdot 10^7$. The γk pulse shape with its oscillations, and the effects known from the charge–time distribution (CTD) (figs. 2.45 and 2.47) and charge–delay distribution (CDD) (fig. 2.56) are visible. The mean height of γk is 24.28 mV (= 1 pe). The entries around 2 pe heights near −120 ns are from γ4k to γ6k (the most frequent γnk effects), which is also visible in the CDD around $d = 40$ ns and 2 pe charge. Dark count pulses with small to multi-pe charges fill the complete time range. The peak at −20 ns is from the end of the acquisition gate and artificial. Arbitrary time offset.

Since the persistent plot contains a lot of overlapping information, it is better to determine the APS through recording and analysis of waveforms, however.

For pulse summation, the height (charge) of each found pulse must be normalized to peak height (total charge), including a polarity correction for negative pulses, and a suitable alignment point must be chosen. As discussed on pp. 154 and 335, a constant-fraction threshold on the point of the steepest rising flank introduces the least uncertainties. Due to the quantized time values, the exact threshold time must be determined by interpolation, e.g. linearly or through Akima splines [368]. Since the average pulse shape must already be known to be able to determine the steepest point t_s, calculating the APS is an iterative process, where the first t_s must be estimated (e.g. at 50% or from a persistent plot).

In addition, the baseline offset must be subtracted. The quality of the APS is greatly improved, if this is done individually for each waveform, since slow oscillations can shift this value. This is achieved by fitting a region without pulses (normally before the PMT response at the waveform start) with a constant or a function which describes slow background variations, and subtracting the fit function from the complete waveform.

Smoothing helps to reduce noise and considerably improves APS quality, since the alignment point is far more accurate, and the APS converges faster with the number of recorded waveforms. A moving-average filter would reduce pulse height and should be avoided. Using a Savitzky-Golay filter with suitable polynomial order and number of datapoints (e.g. order 4 and 11 datapoints for a waveform with 8 gigasamples/s) yields superior results; however, it does not completely preserve the pulse charge.

Finally, correlated background should be subtracted. This can be done by recording empty waveforms, where no photons were detected by the sensor, averaging them using the above measures, and subtracting the average empty waveform from all waveforms.

Best results were obtained for the order background subtraction, smoothing, and baseline subtraction before pulse search.

In order to obtain a clean PMT APS for γk without influence from TSEs through EAPs or pulse shape distortions, rigorous cuts on pulse shape parameters and charge, as well as pulse fitting are necessary. Pulse shape cuts alone are not sufficient and the APS quality improves notably by fitting pulses with the APS from the previous iteration. Again, since the APS is needed in order to obtain a clean APS, this can only be done iteratively.
The merit of this tedious procedure is the PMT pulse response without TSEs, which allows to fit all (possibly superimposed) pulses with good accuracy, although many of them have a reduced first amplification δ. Only this made it possible to resolve the TSEs and develop time shift theory (TST).

Mathematical model

Based on [170; pp. 351–358], the pulse shape can be approximated in the following manner.
For a single photon (spe) pulse shape, only one pk strikes d1, so the SE emission from d1 is a delta function in time (neglecting SE emission times). The pk jitter from e.g. cathode time spread (CTS) plays no role, because this shifts the complete pulse in time but causes no change to the pulse shape.
For the following dynodes, however, the SE initial energies and emission angles together with TSEs cause varying transit times to the next dynode, whose distribution can be approximated phenomenologically with an exponential decay according to Wright [170].
From this follows that the electron arrival times on the anode after passing n dynodes is the convolution of n exponential functions. For simplicity, identical exponential distributions e^{-at} shall be assumed, where t is the time and a is a constant. The distribution $y(t)$ of electron transit times from one stage to the next then reads

$$y(t) = e^{-at}\, H(t) \tag{2.190}$$

where the Heaviside step function $H(t)$ is added to account for the fact that before $t = 0$ the function shall equal zero.
An n-fold convolution of $y(t)$ with itself results in

$$y_n(t) = \frac{t^n}{n!}\, e^{-at} \tag{2.191}$$

where the constant factor $1/n!$ may be omitted, since an additional prefactor is needed anyway to scale the function to the observed pulse height.
This results in an output pulse shape of

$$V(t) = t^n \exp(-at) \tag{2.192}$$

where n is the number of dynodes.

The resulting pulse shape contains all relevant features – a finite rise time which is shorter than the fall time and an FWHM which increases with the number of stages – and closely resembles anode pulses observed in reality. This confirms the assumption of an exponential arrival distribution. Interestingly, the pulse shape becomes more symmetrical with rising n and for $n \to \infty$ results in a Gaussian. This is not surprising when considering that (2.191) is identical to a Poisson distribution (see eq. (2.11)) except for the factor a.

Following [170; pp. 369–373] and [267; p. 61], the pulse width can be characterized by the variance w around the centroid. This can be calculated from the sum of the variances of the transit times between the individual stages v_{ij}:

$$w \approx v_{12} + v_{23} + \ldots v_{n-1,n} + v_{na} + v_a \tag{2.193}$$

v_a is from the current induction at the anode (Shockley-Ramo theorem), v_{na} is often ignored since $t$na is small.
Note the difference to (2.150) for the TTS. Only fluctuations of the centroid of the pulse affect the transit time, and thus the variations of the *mean* interstage transit time of all electrons, which quickly decrease for later stages, enter the TTS equation. In (2.193) the variations in transit time of *all* electrons enter the width of the pulse shape and each stage contributes with full weight. On the other hand, v_{k1} plays no role for the pulse shape since its variation shifts the whole pulse, whereas it is the dominating factor for the TTS.
If all v_{ij} are assumed to have equal values v_d and the last two terms in (2.193) are neglected, one obtains

$$w \approx (n-1)\, v_d \tag{2.194}$$

This simple approximation already shows that the pulse gets broader with rising n.

The v_{ij} and v_d can be estimated from the variance of the interstage transit time distribution of eq. (2.190):

$$\mathrm{Var}\,(y(t)) = \frac{2}{a^3} - \frac{2}{a^4} + \frac{1}{a^5} \tag{2.195}$$

For time shift effects, these considerations have the consequence that the pulse shape is effectively independent of the gain δ except for statistical sampling of $y(t)$, which is not included in (2.193) and may play a role at low δ. Skipping one or two stages should only introduce small changes to the pulse shape, since the majority of the v_{ij} terms still contribute, and will slightly *reduce* pulse width (save for coarser sampling). This explains why the APS fit works so reliably for all TSE pulses.

Influence of voltage divider and cabling

For fast PMT pulses, overshoot and ringing can occur after the pulse from resonance in the electron multiplier and voltage divider as well as reflections in the electronics (e.g. on cable ends through impedance mismatch) [169; pp. 51, 101]. These effects are clearly visible in fig. 2.50 and complicate the search for pileup and EAPs.

To reduce ringing, damping resistors can be used in the voltage divider (VD) design, see p. 149. Reflections in the electronics can be located through the cable transit times.

Pileup from early afterpulses

One large source of contributions to the APS are early afterpulses from time shift effects as is visible by comparing fig. 2.50a (hard cuts eliminating TSEs) and 2.50c (no cuts). The falling flank becomes much broader and the first overshoot disappears if no cuts are applied. This *considerably* increases the fall time from 4.3 to 11.2 ns. Basically the shape of the complete delay distribution enters the pulse shape. This has the greatest influence on the pulse shape shortly after the peak. Integration over the APS curves with and without pileup removal shows that the EAPs increase the APS charge by no less than 60.3%.
Many EAP effects occur with rates per primary pulse far above one, which makes them statistic contributions to the APS and inseparable through cuts – they literally occur for every primary pulse and can thus be considered part of the APS. This indicates that an even larger part of the asymmetry between falling and rising flank is from unrecognizable EAPs. The shape of the APS in fig. 2.50a matches the γk pulse in fig. 2.49b very well, which shows that *non-statistic* EAPs were eliminated reliably.
The afterpulses are thus an essential part of the PMT response and already "planned in" by manufacturers and experimentalists, since otherwise the gain would only be 0.62 times as high [138]. They can thus be seen as part of the PMT response to a photon, though through other gain paths than foreseen by design.
The exclusion of EAPs reduces the asymmetry between t_r ($3.5 \rightarrow 2.7$ ns) and t_f ($11.2 \rightarrow 4.3$ ns) but does not completely remove it. This is attributed to variations in SE interstage transit times, which are inclined to lateness. This agrees with the prediction of a slight dominance of Lnn and Ln scattering versus Enn and En from TST and the fact that L-effects introduce larger lateness than E-effects earliness: Late effect contribute to the pulse shape more often and with larger deviations, which doubly accentuates the falling flank.

For some important time shift effects, example waveforms are shown in fig. 2.52 and can be compared to the γk pulse in fig. 2.49b. These waveforms were selected using cuts on time, delay, charge and charge fraction according to TST calculations and measured distributions, and in regions where no other effect overlapped. Due to the extremely high EAP rate of 108.8% (see 2.5.2.2) many waveforms contain three or more pulses. The shown waveforms were selected to show the effects in the pure form without additional pulses.
In all instances the waveforms clearly show secondary pulses, which for short delays are present in form of pileup. For short delays and large charge differences the alterations are subtle and hardly visible by eye without help of the fit functions. However, fits with two instead of one APS in all instances significantly improve the reduced χ^2 and yield acceptable values. This verifies the findings of TST in 2.3.6 and 2.5.2.2, which are hereby shown to not be caused by artifacts of the fit procedure.

[138] The total charge in the APS is the sum of 100% pe from the γk pulse plus an average charge of 60.3% pe from EAPs, in total 160.3% pe. Without the EAPs only the fraction $\frac{100\%\,\text{pe}}{160.3\%\,\text{pe}} = 0.62$ of the charge remains.

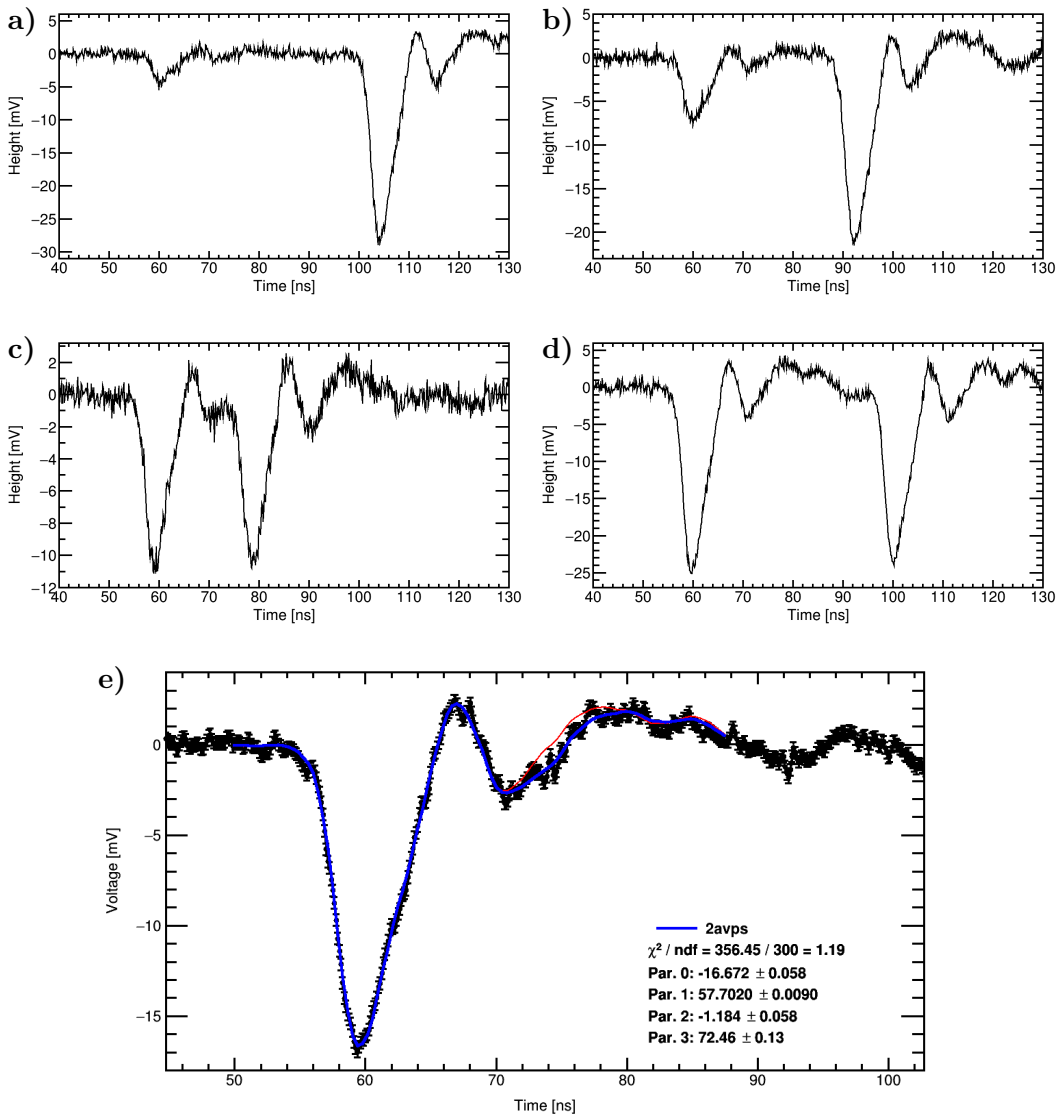


Figure 2.52: Typical pulse shapes of time shift effects producing early afterpulses. Isolated through cuts on time, delay, charge, and charge fraction for primary pulses and afterpulses. Effects shown in order of decreasing maximum delays (BL1, γ5k, γ51, BL11, L2, BE1). For short delays and small APs, the fits for one (red) and two pulses (blue) are shown to illustrate the difference to the average pulse shape (see fig. 2.50); height and time of a pulse are parameter 0+1 and 2+3; y error bars correspond to the baseline standard deviation. Note the scale change on the y-axis between plots.
a) - c) BL1 with delays around 43, 32 and 19 ns: The charge–delay dependence of primary pulse (PP) and AP is clearly visible.
d) γ5k: Both pulses have regular height and charge.
e) γ51: The small AP charge of about 7% pe is visible as an extension of the first oscillation (red. χ^2 2.54 for one pulse, 1.19 for two pulses).
Continuation on next page.

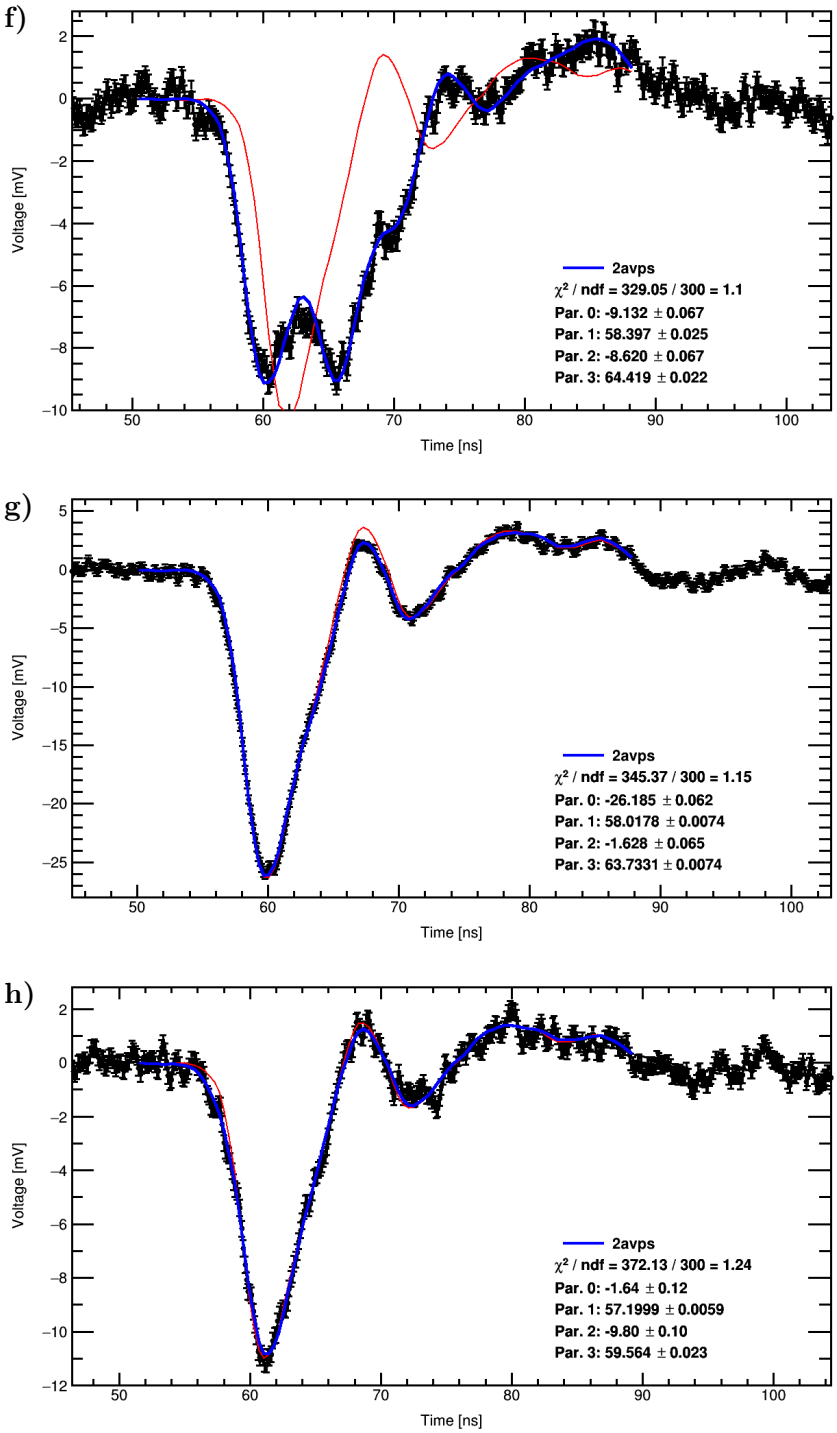


Figure 2.52: Typical pulse shapes of time shift effects producing early afterpulses (continued):
f) BL11: The pileup on the falling flank of the PP and the charge branching are well visible (red. χ^2 52.6 vs. 1.1).
g) L2: The AP reduces the amplitude of the first overshoot (red. χ^2 3.18 vs. 1.15).
h) BE1: The early pulse adds to the rising flank of the regular pulse and slightly reduces the first overshoot through its first oscillation (red. χ^2 1.74 vs. 1.24).

Variation with pulse origin

Underamplified pulses could be slightly narrower due to skipped amplification stages. This could be the case for TSEs like $\gamma 1$ and E1, as well as dynode dark counts and interdynode ionic afterpulses (iiAPs). To analyze this, it would be necessary to create APS for these effects using cuts based on TST.
For some ionic afterpulse (iAP) types, multi-pe pulses are observed, which accordingly should broaden the pulse shape somewhat due to the extra amplification stage (the cathode), which adds an additional term with large variance in (2.193).

Other dependences

The pulse shape varies with the number of detected photons λ, since due to the TTS the pulses are not stacked at perfectly matched transit times, which broadens the pulse. As for the TTS, this effect averages out with growing λ, though.
As discussed, the variations of the interstage transit times have a large influence on the pulse shape. Since the $t$ij decrease with rising interstage voltages, so does their jitter, and the pulse width, rise time and fall time all decrease with growing overall voltage V_t [169; p. 50]. The high voltage (HV) dependence of t_r, t_f, t_w, t_l and the APS is studied on pp. 522–526 based on measurements performed in this work.
Pulsed and direct current linearity (see 2.2.5), as well as the dynamic range of preamplifiers can cap the charge and thus modify the shape.

2.5 Intrinsic background

2.5.1 Dark count rate

Dark counts refers to pulses in the photomultiplier (PMT) output which are produced without incident light. The dark count rate (DCR) – also called dark noise rate or simply dark count [170, 301] – denotes the rate per second with which such pulses are observed.

Origins

Dark count pulses originate from thermionic emission from cathode and dynodes, afterpulses following other dark count pulses (a high afterpulse (AP) probability increases the DCR), particles from intrinsic or external radioactivity hitting the PMT, and cosmic rays striking it [169; pp. 69–72]. These contributions were already discussed starting on p. 112.
Naturally, time shift effects (TSEs) also can occur for dark count pulses, which can affect the DCR through the creation of EAPs and threshold effects. Mere shifts in the transit time (e.g. through L1), however, do not affect the DCR.
In addition, light which is created inside the PMT with long emission times (e.g. phosphorescence, see 2.6.1) can no longer be correlated to its external source or primary pulse, if dark count pulses occur between cause and emission, and effectively contributes to the dark counts.
The same applies to APs from previous photon pulses if the next acquisition window lies within the AP delay window, as occurs in experiments with random interaction times.

Charge

These effects produce pulse charges (unit pe, photoelectron) with:

- Sub-pe values: thermionic emission from the dynodes, early afterpulses (EAPs) (including photonic afterpulses (γAPs) on the dynodes), interdynode ionic afterpulses (iiAPs)
- Regular size: thermionic emission from the cathode, EAPs (including γAPs on the cathode), some cathode ionic afterpulse (iAP) types
- Multi-pe values: radioactivity, cosmic rays, iAPs on the cathode

With rising gain, the contributions to the dark count charge spectrum increase in all regions except for very high charges [170; p. 261], since the number of electrons in the cascade increases and the collection efficiency (CE) improves with higher voltages.

Threshold

The DCR coarsely scales with the inverse of the threshold [170; p. 259].

Voltage and voltage divider

The dependence of dark count on high voltage (HV) was discussed shortly on p. 134 and is characterized by an increasing amount of sources becoming detectable (fig. 2.28).

Initially, due to very low gain only a few multi-pe events from ionizing radiation are visible, before the thermionic pulses from k with 1 pe charge become detectable, both showing a plateau behavior. At high voltages the undersized pulses rise above the threshold. Since an increasing number of skipped stages can contribute, no plateau is observed, and the curve shows a smooth transition to the breakdown region caused by γAPs as visible in fig. 2.28.

A voltage divider which applies a negative voltage to the cathode can substantially increase the DCR (see p. 145).

Background for measurements

Dark counts are a background for all photon counting measurements, be it of PMT properties (see p. 91, 160, 161) or in an experiment, since they introduce fake triggers. Therefore the DCR should be as small as possible. At the same time the signal-to-noise ratio (S/N) must be maximized (p. 123), which is the basis of gain adjustment (p. 134) and is highest during the "plateau" from thermionic emission.

Material

The dark count rate also depends on cathode material and thus quantum efficiency (QE) due to the lower potential barriers for thermionic electron emission, as discussed on p. 56. An improved QE is connected to an increase in DCR. Similar considerations apply to the secondary emission yield (SEY) resp. QE of dynode materials.

Temperature

Naturally, thermal emission is strongly temperature-dependent (see p. 56 and eq. (2.6); [169; p. 71] shows the dependence of DCR and dark current on temperature). However, only contributions to the dark count charge distribution (SER) from thermionic emission increase with temperature; radioactivity and iAPs are unaffected [170; p. 262].
The DCR can thus be greatly reduced by cooling the PMT, until temperature-independent contributions start to prevail around 0°C [169; p. 71]. The few available measurements at cryogenic temperatures show that the DCR typically first falls with decreasing temperature, but then rises again to values even above that at room temperature [170; p. 69]. [290] observes a slow linear rise below −25 °C down to −80 °C. [187] measured a slow rise from 270 K to 170 K, followed by an abrupt increase, a subsequent stronger rise down to 50 K, and ending with a strong upward jump at 35 K. The DCR of 300 Hz at room temperature increased to over 1600 Hz at 35 K. [369] reported an exponential increase between 250 and 4 K with dark counts appearing in bursts. A possible explanation for the rising DCR with falling temperatures was given by [370] through a decreasing photocathode lattice energy with lower temperatures, which results in an increased electron escape probability.

Time in dark

Finally, the DCR is extremely dependent on the time in the dark and the preceding illumination of the cathode.
After exposure to light, the DCR is greatly increased and slowly subsides with time. The DCR decay follows t^{-n} between 10 to 10^4 s, with $n = 1.1 \pm 0.2$ and slightly increasing with the duration of illumination [170; pp. 449–453]. The rate increases

proportional to the exposure time up to moderate intensities and the enhancement is mainly from single-photon (spe) pulses, which indicates the cathode as origin.
Wright notes that all window materials except quartz (fused silica) fluoresce, and that the intensity strongly depends on material. The high K-content borosilicate glass Soveril 747 produces 100 times as much light as Pyrex (also borosilicate). The decay follows $t^{-0.5}$ and is thus much slower than the overall decay, which must contain other contributions.
Other possible explanations are a metastable excitation mechanism of the cathode [180], supported by [332] which considers this and not glass phosphorescence the major mechanism for UV excitation of PMTs (see 2.6.1).
Also, the increase of free ions in the vacuum after exposure to UV radiation results in a progressive gettering on the surfaces when turning on the PMT [180].
Furthermore, in semiconductors charge carriers can remain up to days after illumination due to the so-called persistent photoeffect [371].

Handling and storage recommendations

In order to avoid long DCR cool-down times, light exposure during commissioning should be minimized [170; pp. 449–453] [169; p. 70], and ideally PMTs should be stored in the dark [232]. Despite improvement of the settling time in recent years, it is still recommendable to place the PMT in the dark at least half an hour before commencing a measurement, possibly far longer if a low DCR is required. Turning on the HV during settling might reduce noise levels faster, and the DCR only levels out after weeks in dark.
Under no circumstances should a PMT be exposed to ambient light while turned on, as this in all likelihood will destroy the device. In some cases operation can be restored, but most often with deteriorated performance.

Values

At typical gains, the DCR is dominated by cathode emission due to the large size of the photocathode and the threshold suppression of dynode pulses. Thus the DCR scales roughly with cathode surface. For large box-and-line Hamamatsu [139] PMTs with 10 dynodes, standard QE and diameters from 8 to 20", the typical DCR/area lies between 10.5 and 14.4 Hz/cm^2 after 15 hours in dark. [212]. This corresponds to a DCR between 4 kHz (8" R5912) and 25 kHz (20" R7250).

Dark current

The dark count rate must be distinguished from dark current, which is the output current of the PMT without light input, and is relevant if current is measured instead of pulses.
The term dark noise (DN) is ambiguous and can refer to both dark counts or dark current. In the scope of this work, dark noise refers to dark counts.

Origins

Dark current is caused by all DCR effects plus sources which contribute only current and no measurable pulses. This comprises ohmic leakage currents between the anode

[139] Hamamatsu Photonics K. K.

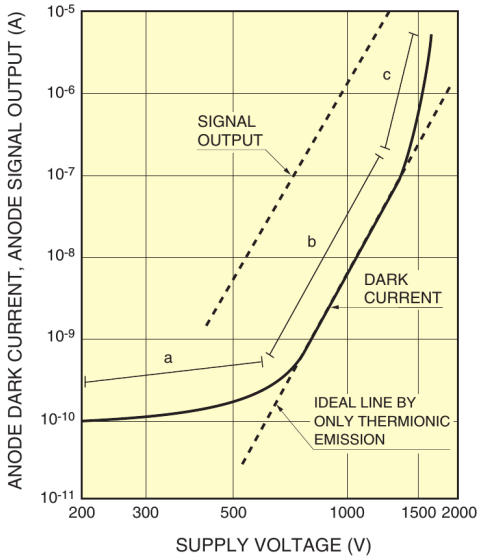


Figure 2.53: Dark current vs. voltage, from [169; p. 69]. Region a is dominated by leakage current, b by thermionic emission from the cathode, c is the breakdown region.

and other electrodes, thermionic emission from later dynodes, and for very high voltages field emission [169; pp. 69–74]. Therefore the dark current is always larger than the product of DCR and mean dark count charge alone.

Voltage

The dark current increases with HV and shows three regions (fig. 2.53). Initially, it is dominated by ohmic leakage, which scales linearly with voltage. At medium voltages, thermal emission becomes dominant, which itself is mostly from the cathode due to its far larger area and gain than the dynodes. Thus the current from thermionic emission is basically proportional to the total gain [170; p. 259], which depends on HV with a power law (eq. (2.102)) and at some point outweighs ohmic leakage. At very large voltages, breakdown begins to set in with an even steeper increase with voltage. [169; p. 70] attributes this to field emission, and γgk or γmk, i.e. photon feedback from pk which hit the glass or dynode mount, whereas [170; pp. 264, 438] ascribes it to light feedback from the anode region. In view of the findings of this work, light feedback from the last dynodes seems more likely, since for the Hamamatsu R5912 the calculated probability of γgk was undetectably low and for γmk only 0.010% even for an extremely low threshold of 4.5% pe, while γAPs had a detectable probability of about 20% (see p. 380).

Measurement

When measuring the DCR, it is important to always record the gain, threshold, time in the dark and temperature as well, since otherwise the measurements are completely incomparable and of little use.

The DCR can be measured by connecting the PMT to a discriminator set to an adequate threshold (e.g. 0.25 pe), which increments a scaler. By setting a fixed

acquisition time on the scaler, the DCR r_d can be obtained by simply dividing the recorded number of pulses n by the measurement time t_d.
The acquisition time should be adapted to the DCR to attain an acceptable uncertainty $1/\sqrt{n}$ from counting statistics.
Furthermore, the dead time of the discriminator should be as small as possible. For a recovery time τ and a non-paralyzable counter, the measured counts n_m are related to the true counts n_t by [170; p. 192]

$$n_m = \frac{n_t}{1 + n_t \cdot \tau} \tag{2.196}$$

When τ is known, this formula can be inverted to obtain a dead time correction. For typical DCR values, dead times of several ns, and reasonable acquisition times, the rate decrease through dead time, however, is negligibly small compared to the statistical error.
With this method, the DCR can, however, only be obtained before or after a measurement with the light source turned on. Due to DCR decay with time in the dark and the drift of PMT and light source it is recommendable to measure the DCR before and after the measurement and interpolate the DCR.

Obtaining the DCR from recorded waveforms during a measurement with a light source has several advantages. This shows the real DCR during measurement, is dead-time free, and is available as a side-product of a measurement. However, it requires offline analysis, since the DCR has to be fitted in the transit time distribution (TTD) or delay distribution (DD).
In the TTD this can be done most reliably in the region before the $\gamma 1$ pulses, since no response from the PMT affects it (TSEs, iAPs, oscillations on the pulse shape). This should be allocated in the acquisition window with sufficient datapoints for a good statistics fit, e.g. 10–20 ns for 8 gigasamples/s. Since the DCR is a random background, the fit function is a constant if an unshadowed distribution (see p. 164) or the delay distribution of *all* APs (here no shadowing occurs) is used. Without unshadowing, the visible fraction of DCR pulses follows an exponential decay and eq. (2.130) should be used as fit function.
For the constant fit, the DCR is obtained similar to the scaler-method by dividing the fitted dark counts per bin b_d by the total time window t_d in which they were recorded. $t_d = n_l \cdot w$ with n_l the number of light source triggers and w the bin width. n_l can be obtained from the number of entries n_t in the TTD and the occupancy o through $n_l = \frac{n_t}{o}$, which results in the DCR obtained from the TTD:

$$r_{d,\mathrm{TTD}} = \frac{b_d}{t_d} = \frac{b_d}{w \cdot n_l} = \frac{b_d \cdot o}{b \cdot n_t} \tag{2.197}$$

In the delay distribution of all ionic afterpulses, the DCR can be fitted with a constant after the afterpulses have disappeared, usually after several 10 µs for large PMTs. If the DD is plotted only for the first APs, this eliminates higher-order APs and reduces the time until the pure DCR is visible, but then unshadowing or use of the adapted fit function is required.
If the waveform acquisition triggers on primary pulses, the DCR for fitting a constant in the DD is simply

$$r_{d,\mathrm{DD}} = \frac{b_d}{w \cdot n_t} \tag{2.198}$$

where n_t is the number of pulses in the TTD, since in this case the window is opened by a primary pulse (PP) and not the light source trigger. If a coincidence of PP and AP is used to speed up acquisition, a similar equation as for the TTD follows, but with the coincidence probability in place of the occupancy.

2.5.2 Afterpulses

Afterpulses (APs) are secondary pulses following a primary pulse, which are caused by processes inside the PMT, with characteristic delays ranging from nanoseconds to many microseconds.
Various alternate names such as satellite pulses, spurious pulses, follower pulses or post-pulsing are used, and the terms already imply the central problem: If for a detected photon not one, but two pulses are produced, this introduces background in an experiment, which can deteriorate energy resolution, pulse shape discrimination, event reconstruction and background discrimination [123, 154, 276, 288]. This can be mitigated through measurement of the afterpulse probabilities, delays and charges, which allows statistic subtraction or event-by-event exclusion through cuts in data analysis.
Since any pulse can cause afterpulses, they are not only found after photon-induced pulses but also in dark noise, in the wake of a muon hit or even after other APs (higher-order afterpulses).

Afterpulse effects can be categorized phenomenologically by delay in three groups:

- **Early afterpulses (EAPs)**, with delays from zero to around[140] $2\,T_0$.
  The end point can be defined by the second-order γak (γak triggering another γak). Since $d_{\gamma\mathrm{ak}} \approx T_0$ (slightly higher), this is roughly $2\,T_0$. For the R5912 a value of about 120 ns results, while EAPs can be unequivocally observed to at least 100 ns delay.
- **Ionic afterpulses (iAPs)**, with delays from about 100 ns up to 15–30 µs. They are the longest known and best studied delay class and thus taken as reference point for the appellation of the other classes.
- **Late afterpulses (LAPs)**, whose delays exceed the range explainable by iAPs and can extend to over 100 µs

The following AP production mechanisms have been identified or suggested so far, and shall be labeled according to the (quasi)particle triggering the secondary pulse:

- **Ionic afterpulses**: Residual gas atoms or molecules can be ionized by fast photoelectrons or secondary electrons to different ionization states. The positive ions drift in reverse direction to the cathode or dynodes where they knock out electrons and trigger a second cascade.
  The large ion masses result in slow drift velocities and long delays in the microsecond range.

[140] T_0 is the mean transit time of the regular pulses (γk); these and subsequent terms are explained in 2.3.5 and 2.3.6.

AP charge ranges from sub-pe (dynode hit) to multi-pe (multiple secondary electrons (SEs) emitted from the cathode).

- **Photonic afterpulses**: Photons can be produced inside the PMT by a cascade through various processes, either by electron bombardment (bremsstrahlung, cathodoluminescence, excitation/ionization of residual gas atoms/molecules), transition radiation [321], or other photons (photoluminescence). An additional cascade can then be started through photoeffect on a sensitive structure in the PMT. Literature describes anode and electrode glow; here, this is extended to all sensitive structures.
  The majority of subtypes produces APs with short delays up to a few ten ns which thus contribute to the EAPs. However, phosphorescence or cathodoluminescence exciting similar states with long decay times can extend to beyond 100 µs, which could contribute to LAPs.
  Afterpulse charges range from undersized (hit on dynode, deck, grid or Al) to regular (cathode).

- **Electronic afterpulses**: Cascade branching can occur due to (in)elastic scattering or if a secondary electron is emitted with above-average energy or in a different direction, which creates a second pulse.
  Virtually all types are very early with delays < 10 ns, most have small charges. If the branching creates an early pulse (EP), the regular pulse becomes the afterpulse which then has near-regular charge and very short delay. Only a few possible processes were discussed before this work, which attempts at a comprehensive treatment of this effect class (see 2.3.5 and 2.3.6).

- **Trapping/Exoelectronic afterpulses**: An electron in the cathode or a dynode, which becomes trapped in a defect, can be released upon disappearance of the defect (e.g. thermally induced annihilation of vacancy and interstitial defects) with sufficient energy to be emitted into vacuum (exoelectronic emission) [372]. This might be a possible mechanism contributing to LAPs (see 2.5.2.4).

In silicon photomultipliers (SiPMs), afterpulses can result from at least three processes, with delays ranging from picoseconds up to the microsecond scale [373]:
If a charge carrier from the avalanche becomes trapped, its release after typically several ten ns produces an AP. This type is simply called “afterpulse” in SiPM terminology.
When a photon produced in the avalanche is absorbed in the substrate, the generated minority carrier can diffuse to the active region of a neighboring cell, where it produces a pulse (delayed crosstalk) with time scales up to several hundred nanoseconds [288].
(Prompt) crosstalk – a photon produced in the avalanche hits the active region of an adjacent cell – is also an AP type, though with only ps delays.
Crosstalk and delayed crosstalk can be considered subtypes of photonic APs, SiPM “afterpulses” are a form of trapping AP but with internal instead of external emission as for exoelectronic APs.

Since the AP charge varies tremendously between a fraction of a pe and over 30 pe, a measurement cannot be compared without knowing the used threshold and is of little use, as was already pointed out by [374] (1973).

2.5.2.1 Measurement

Afterpulse rate

Definitions

In the following, the **AP probability p_a** shall denote the probability to observe one or more afterpulses following a primary pulse. p_a is thus the fraction of primary pulses producing afterpulses and is always smaller than 1 (or 100%).
Opposed to this, the **AP rate r_a** is defined as the mean number of afterpulses per primary pulse, which – depending on definition – can be larger than one (see below).
When all APs following a primary pulse (PP) are independent, r_a and p_a are connected in the same way as the mean number of photon pulses λ and occupancy o in eqs. (2.15) and (2.17):

$$p_a = 1 - e^{-r_a} \tag{2.199}$$

$$\rightarrow \quad r_a = -\ln(1 - p_a) \tag{2.200}$$

In contrast to the AP rate, the **mean afterpulse number** n_a shall refer to the average number of afterpulses per PP producing APs, and cannot be smaller than one. Assuming no correlations between APs, n_a follows from the other two properties through

$$n_a = \frac{r_a}{p_a} = \frac{r_a}{1 - e^{-r_a}} = \frac{-\ln(1 - p_a)}{p_a} \tag{2.201}$$

Dependence on occupancy

All of these quantities depend on the occupancy, since each PP which is generated by a photon can produce afterpulses. Since the PPs are created by independent photons and their cascades do not influence each other except for very large occupancy (see 2.2.5), AP creation for different PPs should also occur independently, in which case the observed number (or charge) of APs scales linearly with occupancy o. This is in fact observed for iAPs up to light intensities of a few hundred pe per PP (e.g. [289, 375]), where linearity starts to affect results, and is expected for EAPs and LAPs as well.

Because the number of APs increases with occupancy, the quantities r_a, p_a and n_a – which are defined for single PPs – cannot be measured directly anymore, since even for very small o some 2 pe contamination occurs. This problem can be resolved in two ways.
Either one uses ultra-low occupancies (typically smaller than 0.2% [170; p. 183]), but this raises the dark count rate (DCR) contribution tremendously (see p. 91) and introduces large uncertainties.
Or the influence of the occupancy on AP quantities is corrected statistically, which will be used in the following.

Occupancy correction for pulse counting

The quantities for N PPs shall be denoted by $r_a(N)$, $p_a(N)$ and $n_a(N)$, the measured quantities are then for $N = \lambda$, the mean number of PPs per light pulse. For the low occupancies used here, the APs of separate PPs should be independent, and then

$$r_a(N) = N \cdot r_a \tag{2.202}$$

Dependences between APs of the same photoelectron (pe) are part of r_a and play no role.
The expected rate for λ PPs follows by summing over the $r_a(N)$, weighted by the probability $p(N)$ to observe N PPs, and dividing the result by the probability to observe one or more PPs, since only then APs can be observed [336]. Using eq. (2.202) and (2.11) this reads

$$r_a(\lambda) = \frac{\sum_{N=1}^{\infty} p(N) \cdot r_a(N)}{1 - p(0)} = \frac{r_a \cdot e^{-\lambda} \sum_{N=1}^{\infty} \frac{\lambda^N}{(N-1)!}}{1 - e^{-\lambda}} = \frac{r_a \cdot \lambda}{1 - e^{-\lambda}} \tag{2.203}$$

From this follows

$$r_a = \frac{1 - e^{-\lambda}}{\lambda} r_a(\lambda) \tag{2.204}$$

This is equivalent to the PP charge gain correction in (2.35). For $\lambda \ll 1$, $r_a \approx r_a(\lambda)$; for $\lambda = 2\%$ the correction factor is 0.990.

In a similar manner, the AP probability for multiple PPs can be derived:

$$p_a(N) = 1 - e^{-r_a(N)}$$

$$p_a(\lambda) = \frac{\sum_{N=1}^{\infty} p(N) \cdot p_a(N)}{1 - p(0)} = 1 - \frac{e^{-\lambda}}{1 - e^{-\lambda}} \sum_{N=1}^{\infty} \frac{\lambda^N}{N!} e^{-N r_a}$$

$$= 1 - \frac{e^{-\lambda}}{1 - e^{-\lambda}} \left(e^{\lambda e^{-r_a}} - 1 \right) = \frac{e^{\lambda} - e^{\lambda e^{-r_a}}}{e^{\lambda} - 1} \tag{2.205}$$

$$\rightarrow \quad r_a = -\ln \left(\frac{\ln \left(p_a(\lambda)(1 - e^{\lambda}) + e^{\lambda} \right)}{\lambda} \right) \tag{2.206}$$

$$\rightarrow \quad p_a = 1 - \frac{\ln \left(p_a(\lambda)(1 - e^{\lambda}) + e^{\lambda} \right)}{\lambda} \tag{2.207}$$

Eq. (2.206) allows to reconstruct the AP rate from the measured AP probability. Using (2.199) then the AP probability can be obtained directly with (2.207). Alternatively, the measured probability can be compared to $p_a(\lambda)$ of (2.205) using the corrected r_a from (2.204).
The derivation of $n_a(\lambda)$ is even more complicated but luckily unnecessary, since the true n_a can be obtained using the corrected r_a from (2.204) or (2.206) together with (2.201).

Occupancy correction through pulse charge

An alternative way to eliminate the occupancy dependence is to use the charges (integration, unit pe) instead of counting the PPs and APs (counting).
The AP rate is then obtained in the following manner:

a) **AP charge rate $r_{a,q}$** [pe/pe]: The charge sum of all APs is divided by the charge sum of all PPs.

b) **AP number–charge rate $r_{a,nq}$** [1/pe]: The total AP number is divided by the PP charge sum [376].

c) **AP charge–number rate $r_{a,qn}$** [pe]: The AP charge sum is divided by the PP number.

$r_{a,q}$ and $r_{a,nq}$ automatically eliminate multi-pe PPs, since their mean charge scales with λ.
The first method (AP charge rate $r_{a,q}$) produces the mean AP contribution to charge. But it makes no statement about the number of APs following a pulse, since the AP charge varies between effects and in most cases is greatly different from 1 pe.
The second definition (AP number–charge rate $r_{a,nq}$) yields the mean number of APs per PP charge and is similar to the counting method; in one case the average charge is used to statistically obtain the PP number, in the other PPs are counted and statistically corrected by the occupancy. Thus comparable rates and errors are expected from both methods. They are not completely equivalent, however, since in $r_{a,nq}$ also the charge variation of single photon (spe) PPs is included (small PPs are weighted less), whereas in r_a all PPs weight equally. Because the PP charge depends on the preceding time shift effects (TSEs) which determine whether APs occur (see 2.3.6.3 and 2.3.6.4), and a larger PP charge corresponds to more electrons in the cascade after d1 [141] which increases the probability of EAP and iAP effects, this dependence of AP occurrence on PP charge introduces differences between r_a and $r_{a,nq}$.
The third method – AP charge sum divided by PP number – is labeled AP charge–number rate $r_{a,qn}$, and shows the average AP charge per PP. Its values are similar to $r_{a,q}$ but slightly different, for the same reasons as argued for $r_{a,nq}$.

The values of the charge-based AP rate types depend on the definition of 1 pe. If the mean charge of clean γk pulses is taken (excluding underamplified pulses), the unit pe is used, whereas $\overline{\text{pe}}$ denotes the use of the mean PP charge. The latter is used by default.

The AP probability p_a and mean AP number n_a are only defined if APs are counted, since the AP number cannot be obtained from AP charge.

An AP charge rate $r_{a,q} > 1$ is not possible, since otherwise the device would operate in Geiger mode [142]. A value above 1, however, is feasible for the counting definition r_a, when many small pulses occur – which in fact is the case. From L3 alone on average $n_a = 1.7$ is expected, for L4 the value is 14.7, and this only increases for later dynodes. One might justifiably ask, whether these small effects should be considered APs or rather part of the average pulse shape; the transition is fluent. On the particle level, however, these interactions are undoubtedly present, which would raise r_a and n_a to astronomical numbers and make p_a indistinguishable from 1, if such small charges could still be resolved.

Whether the counting, charge, or one of the mixed definitions is employed depends on the use case. When pulse shape analysis is intended, the counting method should be used. This allows to determine the occurrence probability of effects independent of their charge. Furthermore, pileup recognition shows the strong charge variations of PPs, which can be eliminated through pulse counting.
When the pulse shape is not recorded, or only the charge contribution through APs is of interest – for example in a detector response function – the charge method can be used.

[141] dn denotes the n-th dynode.
[142] Theoretically, this *is* possible, if the APs have no feedback loops which can trigger the same AP type again.

Since pulse shape analysis yields the charge information as well and with the above-mentioned improvements, where not stated otherwise the pulse counting definition was used for both PPs and APs.

Units

The dependence on PP number and its linear relation are the reasons why the AP rate typically is quoted as %/pe, or short % pe. The unit already indicates that a clean spe measurement or occupancy correction is required for the value to have any meaning.
To avoid confusion, here the AP rate and probability are quoted as % when PPs and APs are counted, as % pe if AP charge and PP number are used, and as %/pe for AP number and PP charge. The rate is given as % pe/pe if charges are employed for both. The mean AP number is stated without unit for counting, and as 1/pe for AP charge and PP number.

Distributions for all afterpulses versus first afterpulses

For all AP histograms, there is the choice whether to plot all afterpulses occurring after a primary pulse, or only the first one. A distribution for the first APs shall be denoted by the subscript 1, the one for all APs by the subscript a. For example, for the delay distribution (DD) – see p. 197 for an overview and explanation of the used distribution types for PPs and APs – this is DD_1 and DD_a.
While the all-APs plot describes the real, complete behavior as in an experiment, it has the disadvantage that higher-order APs, i.e. APs produced by previous APs, are included as well, which covers later processes and hampers identification of effects. In order to resolve this, a complicated deconvolution of the spectrum from itself is required.
These problems can be avoided by plotting the first-AP histogram, which automatically excludes higher-order APs, unless the first AP was too small to be detected. Since both EAPs and iAPs can occur with very small charges, a small threshold should be chosen for AP measurements anyway. On the other hand, the first-AP histogram is subject to shadowing, but this effect can be corrected with the algorithm given on p. 164 and is small for low total AP probabilities (for iAPs and LAPs – *not* for EAPs). The unshadowed first-AP distributions are marked by the subscript 1u (e.g. DD_{1u}).

Correlations

It was established that APs from separate photons are statistically independent for all occupancy values relevant here.
This is not the case for afterpulses from the same photoelectron, since correlations can occur. An AP effect can either prevent a later effect (negative correlation), or cause/favor a secondary AP (positive correlation). For example, after a BE1 no BL1 can occur anymore, since the pk has already proceeded to d2. On the contrary, if an L2sh2E2 afterpulse is observed, there always will be a second since this effect produces three pulses.
In addition, for all APs – whether from separate photons or the same – applies that if two pulses are too close to be resolved individually (pileup), the second pulse will not be detected, which also leads to a negative correlation.

Unshadowing can only correct the statistic exclusion of independent events, and not the reduction or increase of events from correlations, which have physical causes. The shape of DD_{1u} thus only matches that of DD_a, if all APs occur independently and no correlations exist.
If an effect prevents a later effect physically or through pileup (negative correlation), the number of events of the second effect will be smaller in the DD_a than in the DD_{1u}.
On the other hand, if an effect causes a later interaction (positive correlation), DD_a will show an excess. This is the case for higher-order APs, which are produced by previous APs.
If both types of correlations occur, they can partially cancel and cover each other, which further complicates analysis.

It is important to note that after unshadowing a distribution of first APs, the integral over the curve is *not* the AP probability anymore, but instead becomes the AP *rate* of the first afterpulses, which can have values above 1.
This becomes immediately clear if one assumes a hypothetical $p_a = 99\%$. Unshadowing is equivalent to an occupancy correction for the APs (equivalent to the inverse of (2.35)), and thus will multiply p_a by the factor $r_a/p_a = 4.65$ (using (2.201)) to a value of 461%.

Multiple convolution of DD_{1u} with itself only reproduces DD_a, if a) APs can create other APs exactly like a PP, and b) no other correlations are present. Measurements show that b) is not the case; the results of time shift theory (TST) demonstrate that a) is neither (see 2.3.5). The auto-convolution of DD_{1u} only gives roughly correct estimates for effects based on AP-feedback loops to the cathode (e.g. γnk). The probability of any higher order AP can, however, be estimated from calculation of the respective higher-order TSEs.

Unshadowing thus only produces a theoretical curve, which ignores occurring correlations. If one is only interested in the correlated background introduced by the PMT into an experiment, and not in understanding the underlying processes, the distribution of all APs should be used and unshadowing of the first-AP curve is not required.

Measurement methods

Early afterpulses

EAPs can be measured with an flash analog-to-digital converter (FADC) just as the TTD. The timescale is the same and for this also very low occupancies are required to avoid contamination from 2-photon detection (see p. 92). This would introduce fake afterpulses, whose distribution in the transit time T follows the TTD, with particularly severely contributions around the main peak and L1. Counting methods using a scaler or a time-to-amplitude converter (TAC) as for the iAPs (see next paragraph) are unsuitable since they are sensitive to pulse shape oscillations and cannot find overlapping pulses, which imposes a lower delay limit of about 10–20 ns. Instead, waveform acquisition and pulse-shape analysis should be used to find EAPs and uncover pileup.

Ionic afterpulses

When measuring ionic afterpulses the occupancy can be set either very small or very large, as discussed on p. 93. Since the iAP probability amounts to only a few percent, a large occupancy has the benefit of a far lower measurement time (or higher statistics) and a better signal-to-noise ratio (S/N) of the APs to the DCR. On the other hand, this introduces large errors in the unshadowed first-AP distributions for long delays, since these are greatly suppressed, the physics might be slightly different in the last stages (linearity), and the AP response after spe PPs corresponds to the use case in most experiments. Many groups thus shun the multi-pe method (see p. 93).

The easiest method to measure iAP rate is by counting APs and PPs [170; p. 434]. This can be done by opening a gate with the light source trigger with a delay which excludes PPs and EAPs. The PMT signal is amplified (optional) and fed to a constant fraction discriminator (CFD), and all coincidences of gate and CFD output are counted with a scaler. The PP number can be obtained from the measured occupancy (same setup with adapted delay and gate, or in parallel with a second branch) and the number of light triggers. Dividing the AP number by the PP number results in the AP rate within the gate. This can be corrected by subtracting the DCR (same measurement but without light source) and by accounting for multi-pe pulses using eq. (2.204).

A delay-resolved measurement can be carried out with a start–stop setup as for the TTD measurement in fig. 2.48a [170; p. 434]. Here the delay should be increased to exclude PPs and EAPs, and the PP and DCR numbers have to be determined as well.

Waveform acquisition with a digital oscilloscope or FADC in addition allows more sophisticated analyses, such as obtaining charges and measuring all-APs and first-AP histograms at the same time. This can be done as for the TTD (fig. 2.48b) by triggering on the PP and recording a sufficiently long waveform. For spe PPs this leads to low AP statistics, since most waveforms contain only a PP; multi-pe measurements suffer from the problems discussed above.

Both good statistics and unaltered spe behavior can be achieved through a coincidence circuit for PP and APs as shown in p. 460. However, if the AP signal triggers acquisition, the PP cannot be unambiguously identified in the presence of dark counts. By adding an additional coincidence and delay the waveform acquisition can be started with a fixed offset from the light source trigger, which allows PP identification and cuts on the TTD. In addition, the APs and PPs must be counted by a scaler to scale the measurement with the AP rate. The iAP background from dark noise can be reduced by vetoing acquisition if a DCR pulse occurred within the typical iAP delay range (several µs) before the recording window [276]. An exponentially increasing bin width in the delay axis can help to improve statistics at large delays and uncover small effects, but bin contents and uncertainties must be normalized [276].

Late afterpulses

LAPs can be measured similarly to iAPs. A high λ helps to raise the small effects above the DCR.

The DCR limits the LAP measurement window to a timescale not much larger than $d_d = 1/\mathrm{DCR}$ [288]. After a delay of d_d on average 1 dark count pulse is observed, but the AP probability is always < 1 (otherwise the device would operate in Geiger mode) and the AP rate is finite. While one can fit and subtract the DCR (see pp. 161 and 163), for delays larger than d_d AP effects become increasingly stretched in delay since the variation scales with the mean, and reconstruction becomes increasingly difficult. For a DCR of 5 kHz the DCR limit is $d_d = 200\,\mu\mathrm{s}$. For 10 kHz this already affects the typical LAP timescale and impedes measurement, which is another reason to use multi-pe acquisition to lower the influence of dark noise.

Evaluation

The delay distributions for first APs have to be unshadowed (p. 164) as a first step, which is unnecessary if the all-APs distribution is used. Afterwards, the DCR, which is present as uncorrelated background in the delay distributions, is fitted (p. 350) and subtracted (p. 161) to reveal small effects. Integrating over the complete DD yields the AP rate. To improve accuracy, the region where the AP rate is compatible with zero can be excluded from the integral [276].
For the EAP delay distribution possibly no region without AP effects is present in the measurement. In this case the DCR should be determined through a fit in the TTD with (2.197), and the DCR per bin in the DD can then be calculated and subtracted by transposing (2.198).

In the 2D histograms where the charge (fraction) is plotted over time or delay, unshadowing should also be performed if only the first APs are included.
DCR subtraction is more complicated here, since also the dark count single electron response (SER) has to be known. This can be approximated with the PP-SER, although this will overestimate dark count charge. [276] subtracts the DCR from the charge fraction–delay distribution (FDD) by calculating the dark count charge fraction from the distribution of two random draws from the SER (one the PP, one the dark count), which as expected is symmetrical around a peak at 50%. This distribution is scaled with the expected number of dark counts per time bin and subtracted bin-wise from the histogram. Negative bins are suppressed in the subsequent plot.
For 2D histograms, DCR subtraction not always yields much benefit – for small bins and low DCRs nothing changes, since the average dark count per bin is $\ll 1$ and dark counts are present as either 1 or 0 bin counts. Only for large DCRs and at least moderate bin size a subtraction can improve the histogram. In this work small bins were used to better distinguish effects, so DCR subtraction was not required.

2.5.2.2 Early afterpulses

Due to their delays of typically a few up to several ten nanoseconds, early afterpulses in large detectors can be detrimental to energy resolution, event reconstruction, and pulse shape discrimination [123, 190], which requires to measure them in order to estimate the influence on the physics case and analyze possible remedies.

At least two mechanisms contribute to early afterpulses: Photonic and electronic afterpulses.
Electron trapping could also produce EAPs, if the release time scale is in the ns-regime – as observed for SiPMs – and *in addition* the release energy of these short-lived states is sufficiently high to allow escape into vacuum. The latter should at least be the case for negative electron affinity materials. Short-lived electron trapping thus might warrant further studies.

The mechanisms of photonic afterpulses (γAPs) and electronic afterpulses (BAPs) were extensively described in 2.3.5, along with the nomenclature used in the following; the probability, time and charge of time shift effects producing afterpulses was discussed in 2.3.6. In fact, one of the prime reasons for the development of time shift theory was to better understand the diverse EAP effects. The previous state of knowledge was summarized on p. 209.
Individual EAP effects of both types were already listed and discussed in 2.3.6, and summarized in table 2.2 and 2.3. The sources of photonic APs are further analyzed in 2.6.1.
The section at hand summarizes these results and gives an overview over EAP effects by discussing the contributions to the EAP distributions.
The probability types in the TST nomenclature correspond to AP rates as defined here. Therefore the calculated values of the tables for EPs and late pulses (LPs) can be compared to the measured distributions for all afterpulses, but not to distributions for the first afterpulse only.
The following plots and numbers are based on the calculations in 2.3.5 and 2.3.6, and use the same reference PMT: a Hamamatsu R5912 (8") with E7694-01 voltage divider at $g = 1.3 \cdot 10^7$, measured with $o = 2.9\%$, $h_s = 4.5\%\,\mathrm{pe}$, and $h_f = 1.44\%\,\mathrm{pe}$.

Higher-order APs are events, where two or more APs follow a PP. This can have three causes.
Effects where two or more successive interactions occur can produce three or more pulses (e.g. BL1-BL1, L2L3 or E2sh3E3). Many secondary interactions, however, produce no additional APs, as in the case of BL1L1. Also, not all combinations of effects are possible. An effect can only be followed by the same effect, if its origin stage is passed again, e.g. through backward scattering or photon feedback. For example, BE1-BE1 is not possible since the pk has already continued to d2, but BL1-BL1 and γ5k-γ5k can occur. Furthermore, the second effect can alter the properties of the first AP. For instance, for BL1-BL1 the charge of the first AP is smaller than for BL1 alone due to the second branching.
Alternatively, two independent effects can occur for the same PP without a causal relationship, such as two s1L1 with slightly different T, or an s1L1 which occurs together with an sh1E1.
Finally, if more than one photon is detected on the same waveform, the additional PPs and APs can become afterpulses.
Since higher-order APs created by successive interactions can be subject to correlations, one cannot simply convolute an unshadowed first-AP distribution with itself to obtain the all-APs distribution. Rather, all relevant effects have to be calculated using TST.
Higher-order APs are not visible in first-AP distributions, except when the first AP escapes detection.

Discussion of time shift theory results

Afterpulse rate

In the literature typically EAP rates of only a few percent are reported [276,313,347]. At first glance this is surprising, considering the plethora of EAP effects listed in in table 2.2 and 2.3 and their in some cases extremely high AP rates of over 1.
The reason for this discrepancy, however, immediately becomes apparent when looking at the charge–delay distribution (CDD) in fig. 2.56. The red box in the upper right shows the range typically accessible in an EAP measurement if no shielded setup and pileup search algorithms are employed – which excludes most effects and all regions with high probability densities.
Application of these measures increases the observed AP rate r_a to a value of 108.8% for the reference PMT (R5912). This tremendous value is in good agreement with the observable probabilities predicted by TST, as can be seen at a glance from fig. 2.54, where the measured curve is reproduced by the calculated probabilities.

For the reference PMT, after unshadowing, DCR subtraction and also subtraction of the shadowed PP from the AP number, the AP counting rate r_a is 108.8%. This means that on average there are 1.1 APs per PP – for smaller thresholds the values would be even higher, since APs from later dynodes would become detectable as well. The large r_a value, however, does not cause the device to operate in the autogenerating regime, since the AP charge rate $r_{a,q}$, which represents the mean ratio of AP charge to PP charge, is 48.7% $\overline{\text{pe}}/\overline{\text{pe}}$ and thus smaller than 1.
The AP charge–number rate $r_{a,qn} = 50.1\%\,\overline{\text{pe}}$ describes the average AP charge following a PP, independent of the PP size. Here the mean PP charge (1.44 pC) was used as 1 pe, which is why the number is very similar to $r_{a,q}$. This changes, if 1 pe is defined as the charge of γk pulses (2.40 pC), in which case $r_{a,q} = 47.2\%\,\text{pe}/\text{pe} > r_{a,qn} = 28.6\%$, since many underamplified PPs are found through pulse fitting. For the AP charge rate $r_{a,q}$ the different charge unit mostly cancels (except for the DCR and PP correction), but the mean AP charge after a PP $r_{a,qn}$ is inversely proportional to the size of 1 pe.
Finally, the AP number–charge rate $r_{a,nq} = 105.6\%/\overline{\text{pe}}$ lists the number of APs per mean PP charge, which is very similar to r_a. Using the mean γk charge a far higher value of 1.798/pe results, since the value scales with the charge of 1 pe.

Afterpulse probability

The AP probability p_a lies at 72.0% before unshadowing. After unshadowing, p_a becomes the first-AP rate with a value of 125.3%. This is higher than r_a, which can only happen due to pileup or effects excluding each other.

Mean afterpulse number

The measured mean AP number n_a is 1.527. When this value is calculated only from p_a using (2.201), it increases to 1.767, which shows that negative correlations prevail, since less APs are observed than predicted from the first APs. Calculation via r_a produces a somewhat smaller value of 1.641, which is still too large, because the negative correlations reduce the estimate for p_a from (2.199) which is used in (2.201).

Delay distribution

In fig. 2.54 the measured DD_a (AP rate r_a vs. delay d) is plotted together with calculated p_c values from TST for the studied time shift effects which produce EAPs. For afterpulses the delay distribution is more helpful than the afterpulse transit time distribution (AP-TTD), since every effect in the PP-TTD produces clusters in the AP-TTD according to the possible AP effects from the DD, which creates overlaps and makes it harder to pinpoint effects. This ambiguity is resolved for the delay distribution (see p. 197).
The combined observable probability p_c corresponds to the AP rate, and makes the assumption that multiple simultaneous occurrences of the same effect cannot be resolved and are detected as one pulse. This can underestimate the AP rate for effects with large delay ranges and high occurrence probabilities.
The studied TSEs include APs of higher order. If the DD exceeds the calculated values, this can be either due to missing effects, underestimated probability factors or delay spreads. Too high TST values can be caused by too large probability factors or uncertainties of the thresholds.
The DD_1 suffers from shadowing and thus yields too small values when compared to TST calculations. However, unshadowing is not reliably possible (see p. 366), which is why it is not reasonable to compare TST values with DD_{1u}. A qualitative comparison of DD_a and DD_1 resp. DD_{1u} is drawn on p. 366, which allows further insights in the origin of EAPs.

Effects

In fig. 2.54, the first peak at 1.5 ns is from the sum of many effects with small to very large probabilities, which results in the largest peak in the distribution.
Since the pulse fit routine could only resolve pulses with at least 0.8 ns distance, the region before the peak is strongly suppressed, which likely creates a peak where otherwise a continuous rise to $d = 0$ might occur. In addition, the large probabilities combined with the short delays lead to a strong suppression of the DD in the first nanoseconds due to unresolved pileup. This agrees with TST calculations, which predict a total *detectable* EAP rate from all effects of 5.86 (0.57 from EPs, 5.28 from LPs), whereas the DD_a contains a rate of "merely" 1.09. Since the DD is in good agreement with TST values above around 10 ns, the missing rate must originate from shorter delays. The unresolved pileup will also lead to too large AP charges for delays up to about 8 ns.
This means that the EAP rate strongly increases towards smaller delays, since most effects occur for small d, and without pileup recognition the average pulse shape (APS) will show large contributions from EAPs on the falling flank. This is not necessarily a bad thing. The additional pulses can be interpreted as part of the signal response of the device, and add to the charge contained in the pulse. However, they broaden the pulse shape and deteriorate the time resolution of the sensor or a detector employing it.

BL3 and BL4 are expected to show dropping probabilities towards smaller delays (see p. 319) and thus should contribute more to the second peak at 5.5 ns.
For L2, L3 and L4, judging from plot, the delay likely is systematically too high. Such a mismatch was already observed for the earliness of EP effects, which was too small by a factor of 1.3–2 (p. 258). This is not in tension, because the true field layout can modify both the forward and reverse transit times. If the delay is scaled down

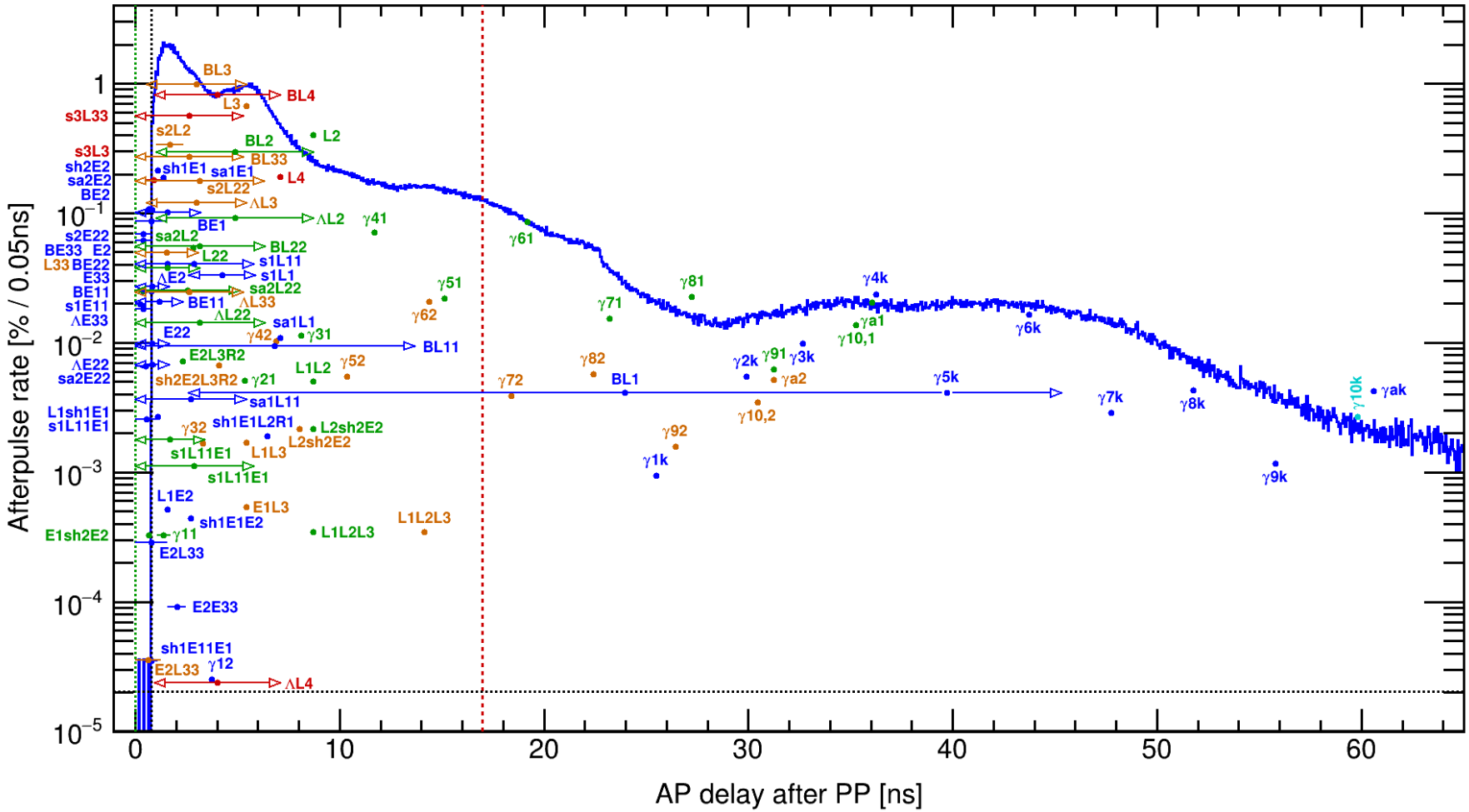


Figure 2.54: Delay distribution (DD) of early afterpulse (EAP) effects for all afterpulses (DD_a) from measurement and calculated with time shift theory (TST). The DD shows the rate r_a (number of afterpulses per primary pulse) over the delay d (transit time difference between an afterpulse (AP) and the primary pulse (PP)).
Data for a Hamamatsu R5912 (+1425 V, $g = 1.3 \cdot 10^7$), recorded with SPAX and pulse fitting (see p. 460) at 2.9% occupancy, $h_s = 4.5\%$ pe (hardware threshold), and $h_f = 1.44\%$ pe (pulse fit threshold) for $4.9 \cdot 10^6$ PPs, resulting in $5.5 \cdot 10^6$ APs and an exceedingly large $r_a = 108.8\%$. Detection limits from 1 entry/bin and the minimum delay resolvable by the fit as black dotted lines. Zero delay as green dotted line. The red dashed line marks the first zero-crossing of the average pulse shape (APS) obtained without pileup-discrimination (fig. 2.50c); in a pileup-agnostic (classic) analysis, only events at larger delays can be detected reliably, which excludes most effects (shown in fig. 3.19d for same measurement).
Effects calculated with TST drawn in with labels, markers and arrows/lines for the peak observable rate p_m over the delay of effects. Markers indicate center of distribution, arrows/lines denote a spread over a delay range. Effect color represents AP charge: blue/cyan $> {}^2/_{\delta 1} = 13.3\%$ pe, green $> {}^2/_{\delta 1 \delta 2} = 1.5\%$ pe, orange $> {}^2/_{\delta 1 \delta 2 \delta 3} = 0.16\%$ pe, red $> {}^2/_{\delta 1 \delta 2 \delta 3 \delta 4} = 0.02\%$ pe.
Large delays $\gtrsim$11 ns are dominated by photonic APs hitting the cathode and d1, at short delays mainly electronic APs (branching) are found. The most probable effects have very small charges and only become detectable through their large mean occurrence per pulse; however, this strongly depends on threshold. The extremely high values at very short delays show that EAPs normally significantly contribute to the average pulse shape. All in all, the delay distribution can be well explained. Detailed discussion see text.
The DD with additional AP charge information (CDD) is shown in fig. 2.56, with AP charge-fraction η_q (FDD) in fig. 2.61.

by a factor of about 1.5, this would match the observed sub-peaks at 5.5 ns (L2), 4.5 ns (L4), and 3 ns (L3) very well, although they can also be caused by BL3 and BL4.

Photonic effects from initially γn1 and then γnk dominate the DD after about 11 ns and thus in most experimental situations based on a classic analysis without pileup-recognition. In general, the values roughly reproduce the DD curve, with some exceptions (see also discussion on p. 324), which suggests that the curve is indeed created by the suspected effects. Since the calculation of these effects involves many unknowns, but these almost only depend on optics and not fields, tuning the probability factors should allow a near-perfect match with data. The calculated delays of the individual γn1 and γnk effects are based on interdynode transit times as well, and thus will show deviations from the measured delays. However, since the total transit time T_0 to the anode was measured and used as input, these differences will not be large.
The oscillating probability in γnk and γn1 between successive origin dynodes (e.g. γ4k, γ5k, γ6k) is due to the assumed dynode stack geometry, which increases the solid angles of earlier dynodes for even n. This is not reproduced by the device, possibly due to a slightly different dynode layout or a flawed solid angle estimation.
γ41 and γ51 lie too low, likely since some probability factors were estimated incorrectly. γ51 is additionally lowered by the solid angle oscillation; in data γ51 has a higher probability than γ61 and γ41 and is the γn1 effect with the highest probability. This can be interpreted as an initially rising γAP probability from a growing number of SEs before the photon losses gain the upper hand and the probability drops again. The same increase and decrease is found for γnk.
In the values of γ10k, γak, γ10,1 and γa1 external paths were included (see fig. 2.43c), which is why their probability is a little higher than expected from photon losses. If external paths are including with the same magnitude into other effects, this would raise their values slightly. The fraction of photons taking external paths, however, could be overestimated: The cluster of γ4k, γa1 and γ10,1 would lead to a too high total value compared to the curve.
The total probabilities for photonic APs are discussed on p. 379.

For BL1, contributions are expected mainly for small and high delays. For the former, the effect is covered by larger photonic APs, for the latter an increase towards larger delays is expected from the FDD in fig. 2.61.

The ledge near 22.5 ns could stem from γ82, since the CDD (fig. 2.56) indicates that these events have up to 2% pe.

General discussion

Three regions can be identified. At delays up to about 11 ns, electronic APs prevail (electronic region). This is followed by a range dominated by photonic APs created on d1 (γn1 region), which extends to about 30 ns, with a fluent transition to the region where photonic APs on the cathode are most frequent (γnk region).

As a whole, the DD curve can be said to be well understood. The probability variables which were used are very general in nature and can be phenomenologically finetuned even without an exact understanding of the processes gained through electrodynamic simulations. Only the delays can not be reproduced completely

based on analytic calculations alone. For effects with small charges (e.g. L4, BL4, ΛL4), the observable probabilities are very sensitive to the exact thresholds, in which case p_c can easily be off by a large factor. This is clearly visible for BL4, which has the same charge as ΛL4 but a mean occurrence N per PP of 39.9 instead of 10.0, which makes BL4 the most likely effect in the DD, whereas ΛL4 is just above the detection threshold.

In a classic analysis without pulse shape analysis, only events with delays above the red dashed line are visible, i.e. a fraction of the γAP and BL1. This could explain, why observation of the other effects has not been reported up to now.

For a detailed comparison of measurement and theory, one would have to calculate the sum curve of all effects including the spread in delay. As a second step, a global fit of all effects would allow to determine the (many) probability factors and improve the understanding of all device parameters – to the extent achievable with analytic calculations. This might make it a worthwhile goal at a later stage.

Delay distribution of first afterpulses compared to all afterpulses

Figure 2.55a shows the DD for all afterpulses and using only the first AP.
The distributions are congruent up to ≈ 3ns due to pileup, since the fit can only resolve pulses which are at least 0.8 ns apart. If a second AP is too close to the first to be found in the fit, it does not enter DD_a. This means that merely a fraction of the effects with delays of only a few ns are found and the AP rate actually is far higher. After about 3 ns, DD_a shows an excess from additional APs following the first.

When DD_1 is unshadowed and the DCR is subtracted from both DD_{1u} and DD_a, fig. 2.55b results. Due to the unresolved pileup at the lowest delays, the DD_{1u} shows an excess up to 7 ns. This shifts the complete curve to too high values compared to the all-AP distribution since unshadowing assumes that all events are independent, which complicates direct comparison. The DD_{1u} would normally lie below the features at 15–22 and 30–50 ns. Nevertheless, general tendencies can still be recognized.
The DD_a shows positive correlations after each large AP population, which stem from secondary APs. This is the case after 7 and 23 ns caused by the peaks at 1.5 and 5.5 ns following the previous clusters. It also occurs after the broad ridge from 30 to 50 ns due to the first two peaks as well as the features around 15 and 21 ns.
For this reason, several features are better visible in the (unshadowed) first-AP distribution. The peak at 15 ns (γ51) and the drop at 22.5 ns (γ71 and γ82) are sharper, the peak of the γnk region is at 41 ns, and after 49 ns an abrupt drop is observed. This indicates that both for the γnk and the γn1 chain the most likely effect is from d5 (γ5k and γ51) and the last effect to occur with notable probabilities is from d6 or d7 (e.g. γ6k and γ61). This shows that the values for γAP from late stages were overestimated in the calculations and the photon loss parameters have to be adapted. The peak at 34 ns in the DD_a is also less pronounced in the DD_1; the excess could originate from γ51 following γ51.

The maximum depicted delay in fig. 2.55a and 2.55b is larger than in fig. 2.54 and thus shows the behavior of higher-order photonic APs. The second-order γnk should be visible up to about 100 ns, third-order up to 150 ns, and so on.

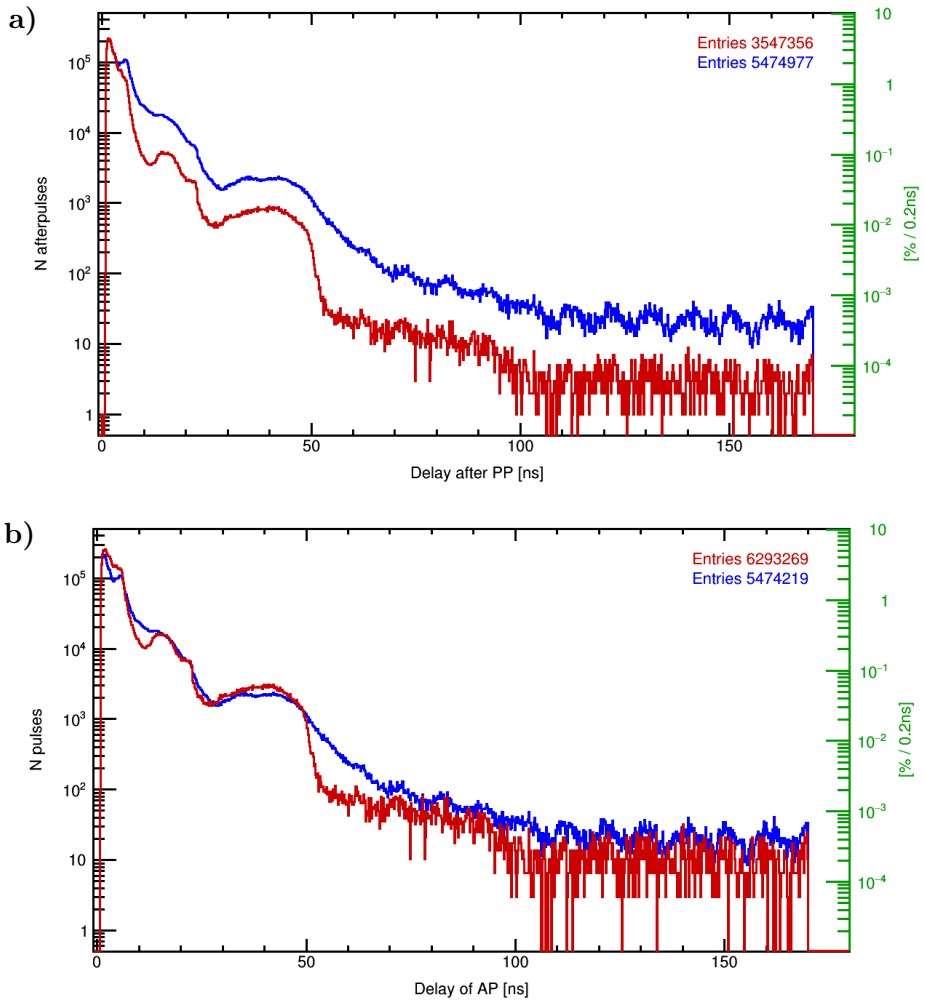


Figure 2.55: Early afterpulse delay distribution: Left axis AP number per bin, right axis AP rate r_a per bin. Same measurement as in fig. 2.54. Discussion see text.
a) All afterpulses (DD_a, blue) and first afterpulses (DD_1, red).
b) DD_a (blue) and the unshadowed first-AP delay distribution (DD_{1u}, red), both after subtraction of the DCR.

The DCR was subtracted in fig. 2.55b from both graphs, but is negligible anyway (0.60/bin). It is important to stress that the flat region after about 100 ns delay can *not* be explained by dark counts. Setup fluorescence should be insignificant, since it only will enter the DD if two photons are detected [143] and the occupancy is small, as is the fraction of the TTD from fluorescence ($\mathcal{O}(1\%)$).
This means that *everything* after γnk, which ends around 50 ns, is from EAPs of higher order, e.g. electronic or photonic APs, which follow a photonic AP.

Interestingly, such effects are also visible in the first-AP distribution. This is only possible if either the first AP escapes detection by being smaller than the threshold or it is delayed by other effects. Since γnk pulses have regular charge and EAPs following other, earlier EAP types cannot reach such large delays, the former is unlikely, which points to other origins. γ6k-L1 or γ6k-BL1 (primary pulse below threshold) could explain the events between 70 and 100 ns. The CDD for first APs

[143] One as PP and a second photon from fluorescence as AP

(not printed) shows that the pulses between 50 and 70 ns have γn1 charges, which indicates γnk-ΛL11 as possible origin; the events from 100 to 170 ns have low charges as well and originate from a combination of multiple-interaction EAPs, interdynode ionic afterpulses (iiAPs) and in part pulse subtraction artifacts (see p. 372). The latter are the reason for the oscillatory behavior at large delays.
In the DD_a, additional contributions from higher-order γnk and their secondary interactions are expected. Second-order electronic APs, γn1 and γnk will contribute up to 100 ns, and the flat distribution at large delays is increased by 3^{rd} (and 4^{th}) order EAPs.

Charge–delay distribution

The charge–delay distribution (CDD) shows the AP charge over the delay after the PP and is plotted for all afterpulses (CDD_a) together with calculated TST effects in fig. 2.56 and 2.57. The measured CDD for extended delays is given in fig. 2.59, where also higher-order γnk effects are visible. The CDD for the first afterpulses only (CDD_1) is shown in fig. 2.58 without unshadowing, since this would introduce errors.

In fig. 2.56 the effect of using the SPAX-setup [144] (see p. 460) and the PEST [145] analysis software (p. 481) are shown. With a good unshielded setup and a classic pulse search based on thresholds, only the region inside the red dashed lines is accessible. The hardware threshold is lowered to the black dashed line (h_s) by performing the measurement in a Faraday cage. Through pulse fitting, the threshold can be further lowered to the black dotted line (h_f) for effects with another pulse $> h_s$ in the waveform. At very small charges and delays, the limits of the fit algorithm become visible in form of an empty triangle around 2 % pe up to 2.5 ns: Small pulses at very short delays after a large primary pulse cannot be distinguished from statistical fluctuations of the PP.
To the author's knowledge, no CDD with such low limits in delay *or* charge has been reported so far, and the EAP effects in this region are observed here for the first time – that means all plotted effects except BL1, γnk, and possibly some γn1. This allows to confirm the predictions of TST for most relevant effects. It also explains, why most interactions with delays below 10 ns so far have not been described, in contrast to BL1 and the photonic APs.

Artifacts

The plot shows that the setup fluorescence plays no large role for APs, since it is absent at 100 % pe between 8 and 20 ns. In the afterpulse charge–time distribution (AP-CTD) in fig. 2.60 on the other hand the fluorescence is visible, since it shifts the PP and all APs following it backwards. Since the CDD is the charge-resolved DD, no large contamination by fluorescence will occur in the DD either.
The clusters at the 2 % pe level around 34, 44 and 54 ns are caused by the APS oscillations combined with pulse subtraction of a slightly too small or high charge. In this case either a maximum or minimum in the APS can produce a small artificial peak in the waveform, when a noise spike, which sits on top of the extremum, extends above the threshold. The peak at 22.5 ns, however, can *not* be explained this way

[144] Shielded Photosensor Analysis eXperiment
[145] Photosensor Evaluation Software Toolkit

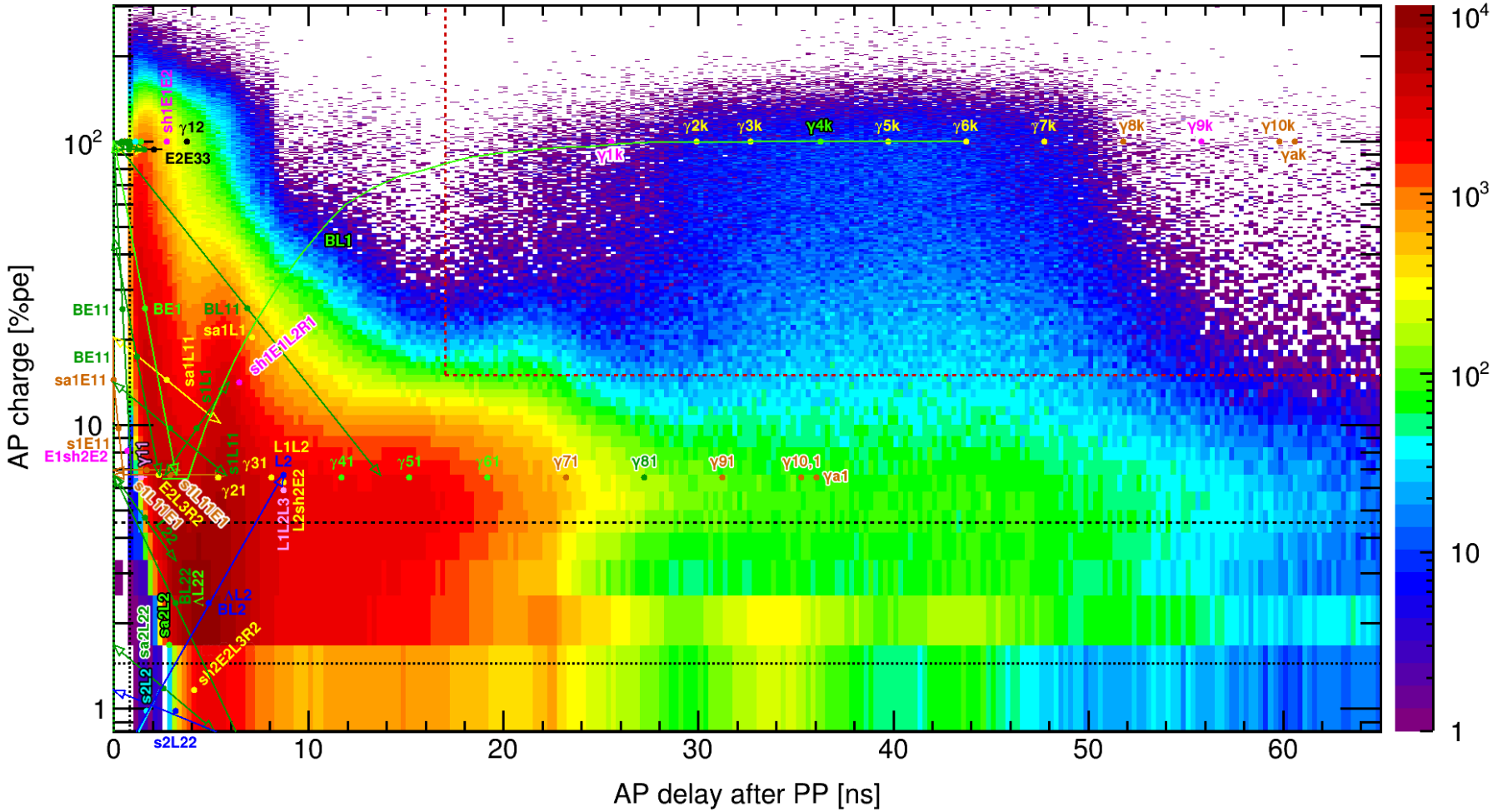


Figure 2.56: Charge–delay distribution (CDD) for early afterpulse effects showing all afterpulses (CDD_a) from measurement and TST calculations. The CDD depicts the relation of AP charge (y-axis) and occurrence rate (z-axis) to the delay after the PP (x-axis). — Data recorded for a Hamamatsu R5912 at +1425 V ($g = 1.3 \cdot 10^7$) with SPAX and pulse fitting for 2.9% occupancy, instrumental threshold $h_s = 4.5\%$ pe, fit threshold $h_f = 1.44\%$ pe, and bin widths 0.25 ns and 0.02 pC; $5.5 \cdot 10^6$ entries, $4.9 \cdot 10^6$ PPs. 1 pe = 2.40 pC (charge of clean γk pulses). Number of pulses per bin indicated by z-axis (colors). h_s long dashed line, h_f dotted line; effects below h_s can only be detected if a pulse $> h_s$ triggered acquisition. The lowest bins are only partially filled and have reduced bin contents. Black dotted line at 0.8 ns denotes minimal resolvable pulse distance in fit. Red dashed lines mark the first zero-crossing of the average pulse shape without pileup discrimination (17 ns, see fig. 2.50c) and a typical good charge threshold of 15% pe. Without a shielded setup and pileup recognition only the section within the red box is reliably detectable for this voltage divider, which includes only γnk, a part of BL1 and possibly a small fraction of some strong γn1 effects (shown in fig. 3.20d for same measurement). — Effects calculated with TST drawn in with labels, markers and arrows/lines, whose color corresponds to the predicted observable rate p_c: > 10% cyan/blue, > 1% bright/dark green, > 0.1% yellow/brown, > 0.01% pink/magenta, < 0.01% red/black, below 10^{-5} omitted. Markers show calculated center of distribution. Systematic charge and delay ranges of effects displayed by arrows/lines. Diagonal arrows indicate the tendency of the charge–delay dependence; exact curve exemplarily drawn for BL1. — An enlarged view of the region around 100% pe for short delays, where a large number of effects is concentrated, is shown in fig. 2.57. — All clusters of the CDD can be explained by TST. For detailed discussion see text. — When the CDD is integrated over the charge, the DD results (fig. 2.54). In the FDD, instead of the AP charge the AP charge fraction η_q (definition in eq. (2.169)) is plotted (fig. 2.61). See p. 197 for disambiguation of plot types.

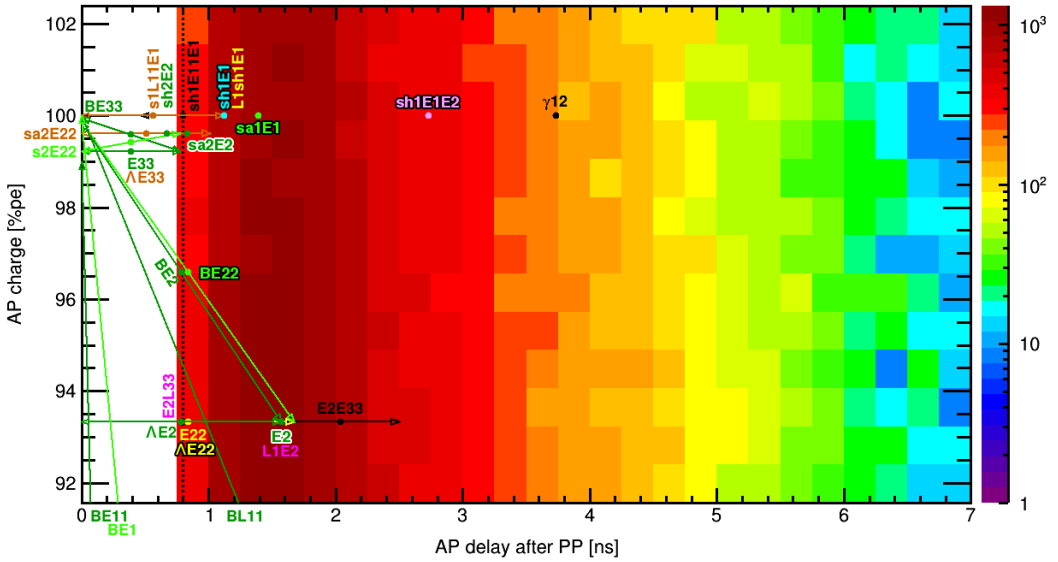


Figure 2.57: Enlarged section of the EAP-CDD in fig. 2.56 for charges near 1 pe and short delays. There is a clustering in this region from the regular pulses (RPs) of EP effects. Due to the large charge resolution of the PMT the effects are smeared and not distinguishable. Due to the minimum resolvable delay of 0.8 ns of the fit routine (black dotted line) no pulses can be detected at shorter delays.

and might be caused by $\gamma 82$, which lies directly below it and has a p_c of 0.27%, which is approximately equivalent to the excess in the DD at this delay.
The sharp line at high charges directly before 8 ns is caused by the detection window of the fit routine. If a pulse is found, the algorithm searches for other pulses up to 8 ns before it, and sometimes discovers a small, previously unrecognized PP before a large AP, in which case the latter enters the CDD. After 8 ns delay no further AP can be found this way, so the CDD appears empty compared to smaller delays. In addition, this leads to a condensation near the detection window border from pulses which partially extend into the window with their trailing flank.
The mentioned artifacts are small and yield no strong contributions to the CDD.

Effects

Just as the DD, the CDD can be well explained by TST and no unresolved effects remain.

From fig. 2.56 and 2.57 follows that the first peak in the DD at 1.5 ns is caused mostly by BE1, BE11, s1L11, sh1E1, sa1E1, E2, BE2 and BE22 (first "finger" in the CDD). The second finger, which results in the peaks around 5.5 ns is mainly formed by L2, BL2, ΛL2 as well as s1L1, sa1L1 and the early part of BL1. s1L1 and sa1L1 should follow a very similar curve as BL1.
Primary pulses from a second photon can filter into the CDD. The strongest influence is expected from the main peak which would become visible at very short delays and over nearly the complete charge range. Due to the low occupancy of 2.9%, however, at most only a rate of 1.5% multi-pe events are expected (eq. (2.20b)), which is far less than the observed events in this region.
The highest AP rate density is encountered near h_f around 4.5 ns where BL2, ΛL2, BL22 and ΛL22 cross (actual curves are concave).

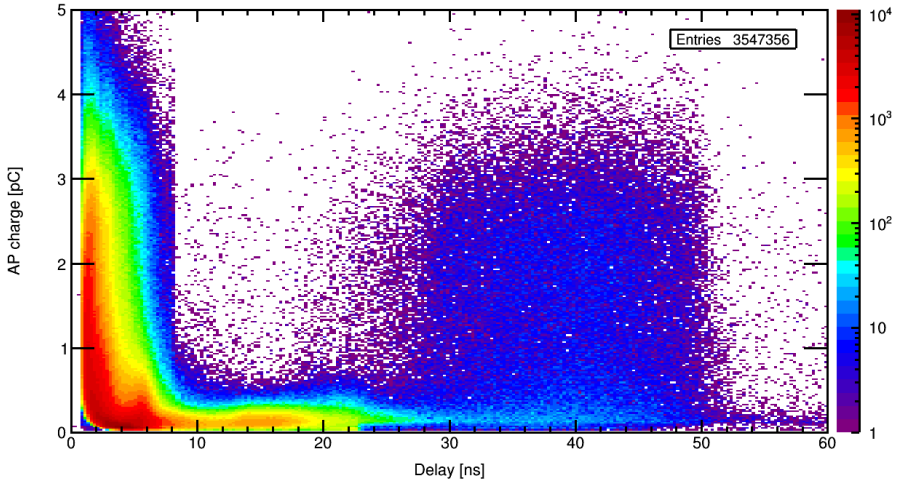


Figure 2.58: CDD for first afterpulses only (CDD_1), without unshadowing; same measurement and binning as in fig. 2.56, but with linear ordinate with absolute charge, 1 pe = 2.40 pC. Many secondary effects, which are visible in fig. 2.56, disappear. Discussion see text.

As found for the DD, the delay of L2 should be shorter by about a factor of 1.5 to match the measured cluster around 6 ns. L3 and L4 are unobservable individually, but have high mean expectancies which extends their sum charge above the fit detection threshold.

The charge–delay function of BL11 will also be curved, which should reproduce the observed green contours above the line. Unfortunately, an exact curve can only be calculated with precise knowledge of the fields. For increasing delays, likely at some point the pk energy becomes too low for refocusing onto d1, and the pk is drawn directly to d2, which would lead to a charge drop (p. 299) as observed for several effects in the charge–time distribution (CTD). This is supported by the far lower number of events at 20–30% pe between 10 and 14 ns in the first-AP CDD, which points to pileup from multiple EAPs as origin of the events in the CDD_a in this region.
The events for regular charge and delays up to 7 ns can be explained by a combination of BL11, multiple EAPs, γ12, and 2 pe late primary pulses (cathode time spread) filtering in from the TTD.

The γnk chain can be seen very well in the CDD_a and CDD_1. γ1k has a far lower probability than γ2k, which is why the onset of the latter is clearly visible. The maximum probability is reached around γ5k, after which the effect strength starts to drop and nearly completely vanishes after γ7k (fig. 2.58). The events in fig. 2.56 after 50 ns thus likely stem from γnk triggered by earlier APs. The early stages can produce cathode hits without or with very few reflections, so the increase of the SE number dominates initially, but for later stages the light path suppression prevails – be it for internal or external paths. This of course must be the case by design, or else the effective light feedback from late stages with $\mathcal{O}(10^7)$ SEs would render the device unusable without electronic reset (Geiger mode). In fact this is what sets the upper limit of the usable voltage range (see pp. 134, 350, 379, 408 and 418). The estimated times of γ10k and γak have a rather small error, since this is just the (precisely measured) total transit time plus the photon transit time to the cathode (estimated at about 0.7 ns).

For the most probable γnk effects γ4k to γ7k even some pulses with 2 pe charge due to pileup of two γnk (or 2 pe emission) are present. This is better visible in the persistent plot in fig. 2.51.
For γnk the behavior from 0 to 16 ns repeats as an echo through second-order APs, which is clearly visible between 50 and 65 ns at medium and low charges, since these events are missing in the CDD_1. This also causes the excess at charges from 20 to 60% between 30 and 48 ns in CDD_a; γnk events have regular charges and would not cause higher bin contents for smaller charges.

The γn1 effects show up in form of a band extending from 5 to around 30 ns (see fig. 2.58. After γ51 the effects decrease in probability, before this effect BL11 contributions prevent comparison of their strength. γ91 and beyond are strongly suppressed, if present at all.
In CDD_1 the γn1 band is still visible after the delay calculated for γa1. Only a few explanations can cause such long delays: Either a first γn1 AP is smaller than the threshold but produces a second-order γn1 which is detected, or a γnk is subject to a second interaction (e.g. γnk-E1). Likely both effects occur.
The events in CDD_a after 36 ns thus hail from a combination of second-order γn1 and dual-interaction γnk. Between 28 and 36 ns, first order γn1 overlap with these effects, which makes it hard to discern whether γ91 and later γn1 occur or not.

For γnk the cathode can not only be hit through internal photon paths but also via external ones (see fig. 2.43c), which is not the case for γn1 or only with greatly reduced probabilities. This could lead to a probability increase (or less steep drop) of γnk compared to γn1 for the last stages. However, this is not observed, which can be attributed to the light shields placed around the last stages with the explicit purpose of preventing light feedback to the cathode from anode glow. Instead, both γnk and γn1 show similar behavior: γn1 effects stop after d8, and γnk after d7, that is, in both cases after seven crossed stages; for more steps the losses dominate and impose a rather abrupt cut.
Combination of these observations lead to the conclusion that in case of the R5912 the internal light paths dominate for both photonic AP chains. Photographic proof of PMT-generated photons hitting the cathode through internal paths for the same PMT is shown in 2.6.1, while no light emission via external paths could be measured despite extended trials.

Finally, for BL1 the calculated true curve is exemplarily drawn (see p. 243). This is indistinguishable from γnk at large delays, shows a gap around 16 ns, and becomes visible again below ≈ 13 ns.

Long delays

The CDD in fig. 2.59 shows additional clusters after the end of γak, which can only stem from higher-order and multiple-interaction EAPs. The events at regular charges between 50 and 100 ns can be caused either by second-order γnk and γnk-L1. At larger delays, statistics are too low to unambiguously determine whether third-order γnk are present. At γn1 charges, events can be found until the end of the window at 170 ns delay. This could be explainable through secondary interactions of higher-order γnk. The cluster between 112 and 150 ns would match with the delays of third-order γnk followed by E1.

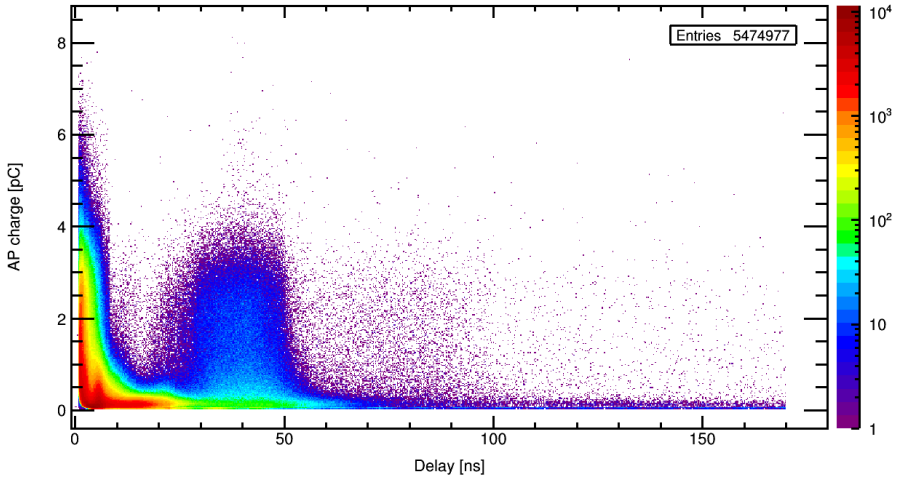


Figure 2.59: CDD$_a$ with extended delay range, same measurement and binning as in fig. 2.56, 1 pe = 2.40 pC. Higher-order EAPs are clearly visible at delays above 50 ns. Discussion see text.

Since only very few pulses with regular charge are observed after 100 ns delay, it is more likely, however, that these events are the first interdynode ionic afterpulses. This is studied and verified in 2.5.2.3. Fig. 14 in [276] shows the CDD for ionic APs for the same PMT type, in which iiAPs can be observed with small charges down to at least 250 ns and possibly to 100 ns delay.
Most likely, both effect groups overlap, and a fluent transition between EAPs and iAPs occurs.
In the lowest charge bins artifacts from APS peaks and non-ideal APS subtraction become visible with intervals of roughly 10 ns. While negligible compared to true EAPs for small delays, near the end of the extended window these events represent a substantial contribution, and EAP (resp. iiAP) rates cannot be reliably extracted anymore.

Afterpulse charge–time distribution

The charge–time distribution for EAPs is shown in fig. 2.60 for all afterpulses.
As for the AP-TTD (not shown), in the AP-CTD the different possible PP times (see TTD in fig. 2.36) cause their afterpulses to overlap, but here the charge information helps to distinguish AP effects to an extent.
Despite this, the AP-CTD is helpful when compared to the TTD and CDD, since it shows which PPs and APs can cause which effects in the CDD. In addition, it reveals that the RPs of EP effects in fact become afterpulses. Otherwise the region around 100 pe at $T = T_0$ would be nearly empty due to the low occupancy and EP probability; the expected rate of 2 pe PPs of which one is an EP and one is an RP is very small and only $\mathcal{O}(0.1\%)$ for the shown measurement (see values in table 3.3 and (2.18)).

The AP-CTD is generally very similar to the CDD and can be said to be very well understood, with no unexplained effects remaining. Other than in the CDD, however, the features of the TTD become present, such as the L1 peak around 50 ns in form of its secondary effects (represented by L1sh1E1 and L1E2). This leads to an increase of the secondary APs between 50 and 65 ns compared to the CDD.

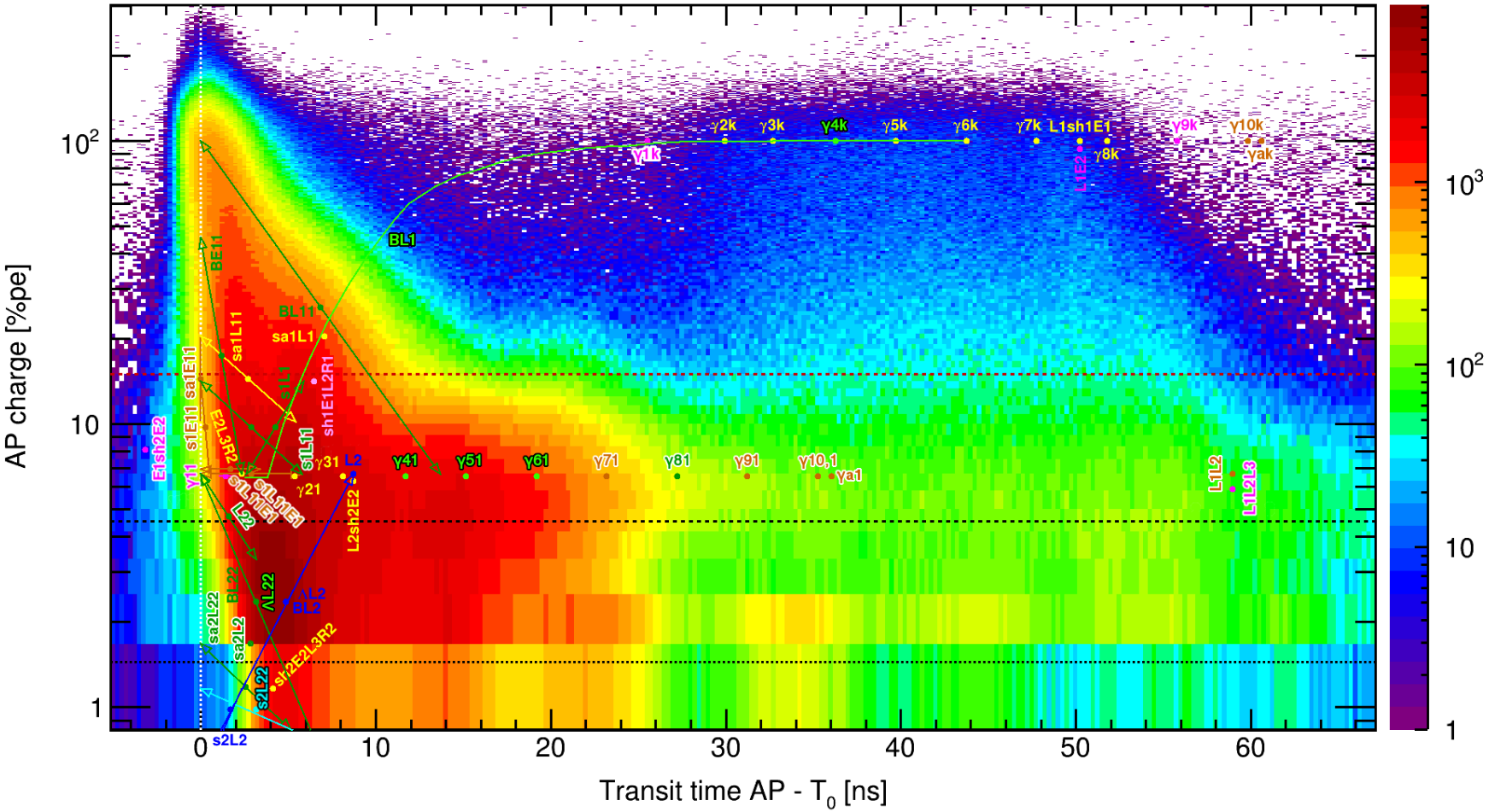


Figure 2.60: Afterpulse charge–time distribution (AP-CTD) for early afterpulse effects from measurement and TST calculations. This plot shows the charge and probability of APs vs. their transit time.

Data for a Hamamatsu R5912 at +1425 V recorded with SPAX and pulse fitting using $h_s = 4.5\%$ pe, $h_f = 1.44\%$ pe, and bin widths 0.25 ns and 0.02 pC for $4.9 \cdot 10^6$ PPs, resulting in $5.5 \cdot 10^6$ APs. Relative transit time T_r to γk on x-axis. 1 pe = 2.40 pC (clean γk charge). Number of pulses per bin indicated by z-axis (colors). h_s long dashed line, h_f dotted line. Lowest bins partially filled. The threshold of a good unshielded setup without pileup-recognition is marked by the red dashed line (15% pe); everything below is undetectable. See fig. 2.56 for more information on the measurement.

For effects calculated with TST (drawn as labels, markers and arrows/lines) the color corresponds to the predicted observable rate p_c: > 10% cyan/blue, > 1% bright/dark green, > 0.1% yellow/brown, > 0.01% pink/magenta, < 0.01% red/black, below 10^{-5} omitted. Markers show calculated center of distribution. Ranges in time or charge shown by arrows/lines. Diagonal arrows imply the tendency of the charge–time dependence; exact curve exemplarily drawn for BL1. Effects where a regular pulse (RP; possibly with reduced charge) becomes an AP are omitted to avoid clutter (very many overlapping effects).

A point in the CDD (which depicts AP charge and rate vs. *delay*, see fig. 2.56) is smeared in the AP-CTD, since the transit time of the PPs which can create this AP type varies (cf. the PP-CTD in figs. 2.45 and 2.47); see p. 197 for an overview of the distribution types. However, the AP-CTD shows the time information, which is lost in the CDD. Through this it can be verified that RPs which occur after an EP become APs (region at $T_r = 0$ and charges between 6.7 and 100% pe).

The AP-CTD can be said to be completely understood, and no unexplained phenomena are found. Discussion see text.

APs can occur with negative times if they follow a PP with large enough earliness. This can be seen at 2–10 % pe for T_r between −4 to 0 ns for some small effects based on E1, similar to E1sh2E2 which is drawn in exemplarily.
As already mentioned, between 8 and 20 ns delay around 100 % pe the setup fluorescence becomes visible, since the APs following it are delayed. The artifacts from unideal pulse subtraction at 2 % pe are visible again, since the majority of PPs have times close to γk, so no strong smear occurs.

Charge fraction–delay distribution

The charge fraction η_q was defined as the charge of an afterpulse divided by the charge sum of the AP and its PP in (2.169). In fig. 2.61 and 2.62 the charge fraction is plotted over the delay (charge fraction–delay distribution, FDD) for all afterpulses (FDD_a) together with TST calculations. Fig. 2.63 shows the FDD for the first afterpulses only (FDD_1) and for extended delays.
The FDD is similar to the CDD, but makes a systematic dependence between AP charge and PP charge visible, such as for BL1, which cannot be distinguished from γnk in the CDD. It also helps to disentangle effects which have the same AP charge but different PP charges. Furthermore, since the charge fraction has an upper limit of $\eta_q = 1$, events are condensed at large charge fractions, which better shows their presence (e.g. γ12).
For small delays and either small or large η_q the pulses cannot be resolved by the fit, since a small pulse is indistinguishable from statistic fluctuations of a large pulse shortly after or before it. Thus no events are detectable in these regions.

For L2 also in the FDD a shorter delay of around 6 ns is required to match the observed high-density cluster.
The tail of the second finger (around 5.5 ns) extends up to charge ratios of 1, which can be understood by an initially slow drop of BL11 with rising delay and unresolved pileup from frequent effects. Since η_q is capped at 1, a high density ridge is observed at the upper border which results from statistic condensation. The events with high charge fractions around 10 ns delay likely also originate from pileup.
For the γn1 effects it is clearly visible that γ51 has the highest probability. Its delay is estimated slightly higher than the center of the cluster.
BL1 is indistinguishable from γnk in the CDD, but due to its correlation of AP and PP charge it is well resolved in the FDD and separated from the γnk, whose charge is independent of PP charge. For better comparison with data the exact curve was calculated, which shows a complicated shape (discussed on p. 244). Similar shapes are expected for other effects with anticorrelations of PP and AP charges. Fig. 2.63 is plotted with a linear ordinate, which better depicts the dropping η_q at large delays.

The artifacts from non-ideal pulse subtraction occur at the same delays as in the CDD. However, they are more problematic here. If a small PP is followed by a large AP (e.g. E2), unideal subtraction of the latter can cause the artifacts from the APS peaks to have far larger charge ratios. This is the cause of the stripes extending upwards from the artifacts clusters. One can even discern that the delays slightly rise for larger charge fractions, which is expected from EP-events with large APs. The subtraction routine was heavily optimized to minimize this behavior; no better performance than shown could be achieved. The remaining contributions from subtraction artifacts are small and can be easily recognized by their characteristic shape, respectively be excluded by charge cuts.

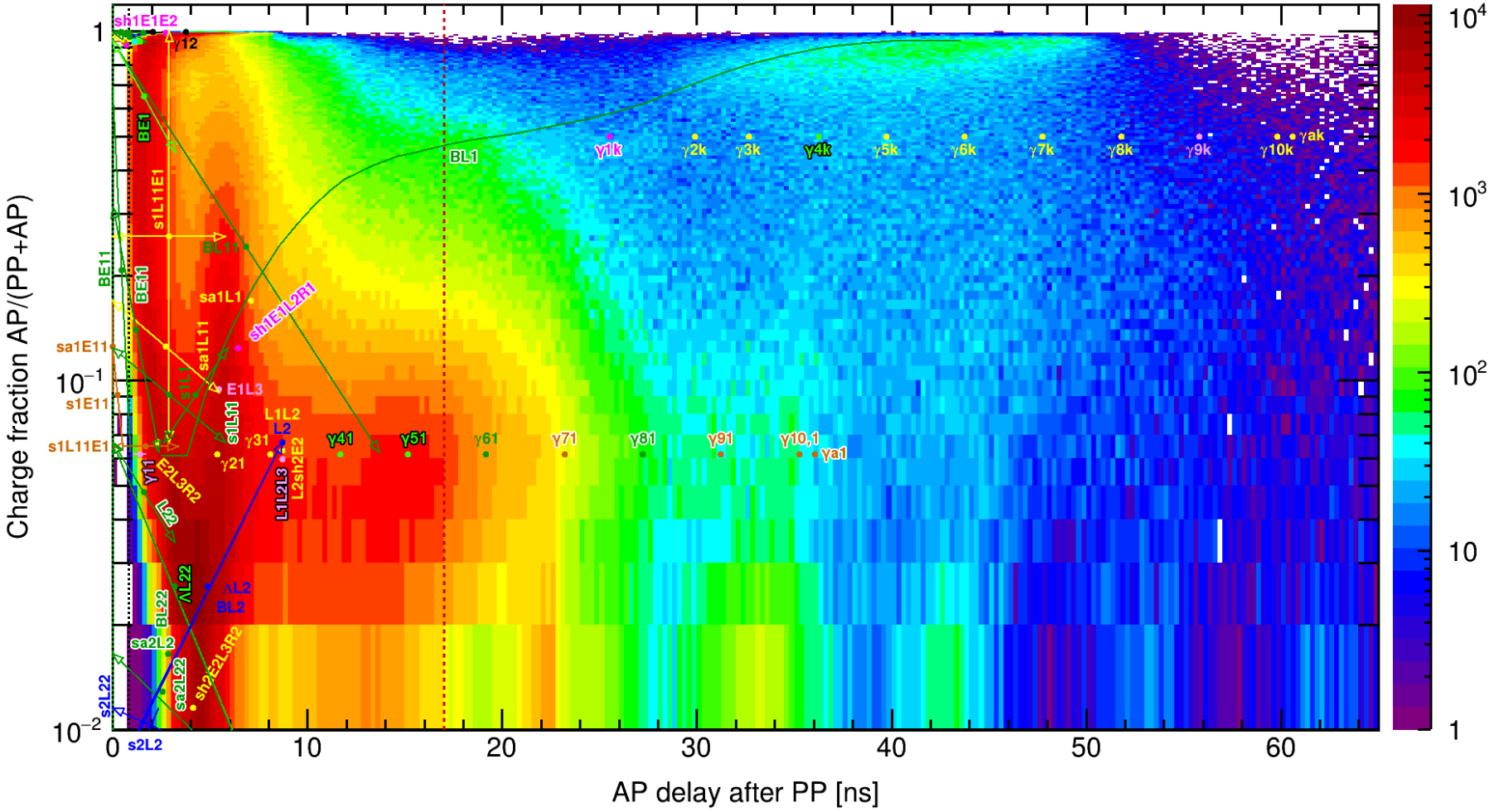


Figure 2.61: Charge fraction–delay distribution (FDD) for early afterpulses from data and TST predictions. This distribution depicts the AP charge fraction η_q (charge of an AP divided by charge sum of this AP and the PP) and occurrence probability over the delay (transit time difference between AP and PP).
Measurement for a Hamamatsu R5912 at +1425 V recorded with SPAX and pulse fitting using $h_s = 4.5\%$ pe, $h_f = 1.44\%$ pe, and bin widths 0.25 ns and 0.01 for $4.9 \cdot 10^6$ PPs, resulting in $5.5 \cdot 10^6$ APs. Number of pulses per bin indicated by z-axis (colors). The red dashed line marks the first zero-crossing of the APS containing pileup (fig. 2.50c); in a classic measurement, only some of the afterpulses with larger delays can be detected reliably, which excludes most effects (shown in fig. 3.21d for same measurement). See 2.56 for additional information on measurement.
For effects calculated with TST (drawn as labels, markers and arrows/lines) the color corresponds to the predicted observable rate p_c: $> 10\%$ cyan/blue, $> 1\%$ bright/dark green, $> 0.1\%$ yellow/brown, $> 0.01\%$ pink/magenta, $< 0.01\%$ red/black, below 10^{-5} omitted. Markers show calculated center of distribution. Ranges in time or charge shown by arrows/lines. Diagonal arrows imply the tendency of the charge–delay dependence; exact curve exemplarily drawn for BL1.
An enlarged view of the region around $\eta_q = 1$ for short delays, where a large number of effects is concentrated, is shown in fig. 2.62.
All clusters of the FDD can be explained by TST. Detailed discussion see text.
The FDD yields complementary information to the CDD (fig. 2.56); e.g. BL1 is clearly separated from the γnk effects here. Integration over η_q produces the DD (fig. 2.54).

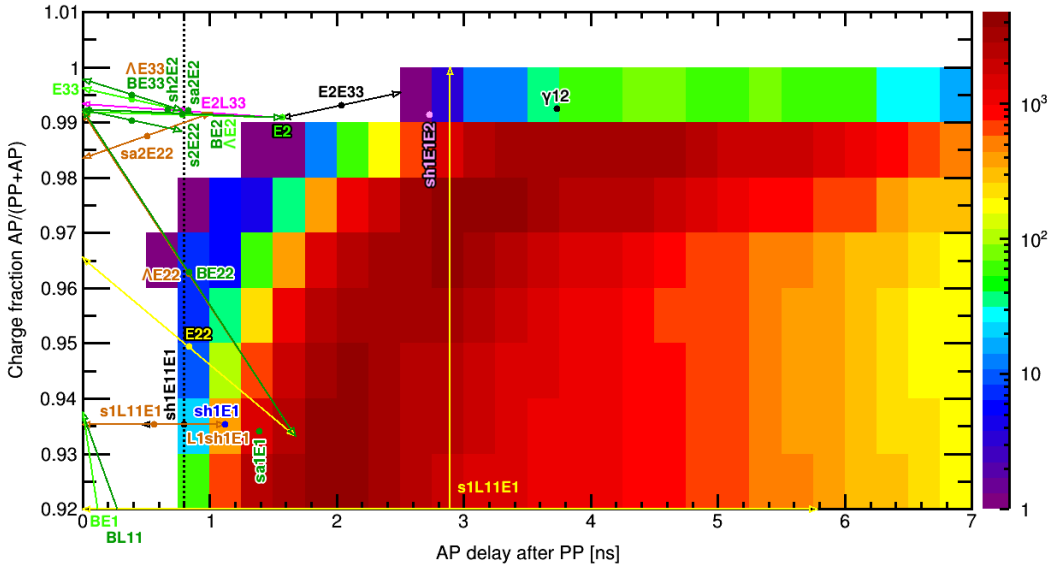


Figure 2.62: Enlarged section of the FDD in fig. 2.61 near 100% pe and short delays. As for fig. 2.57, EP effects cause a cluster of RPs here. Since for the charge fraction η_q a value of 1 is a hard limit (cf. eq. 2.169), the effects are condensed and better visible, although the region closest to $\eta_q = 1$ and $d = 0$ cannot be resolved by the fit routine. The black dotted line denotes the minimum detectable delay.

The FDD_1 shows a marked decrease at high charge fractions for delays of 4 ns and higher, which confirms that pileup is the main origin of these events in the FDD_a. $\gamma 51$ is very clearly visible. The γnk are centered around 50% as expected, show the highest probability for delays corresponding to $\gamma 5$k, and abruptly disappear after 50 ns, which is in accordance with the findings from the DD.
At delays from 60 to 100 ns, BL1L1 and BL1BL1 become visible. For even larger delays starting around 110 ns an isolated low charge cluster appears. This matches the events discussed for the CDD on p. 372. Since virtually no events with medium charge fractions are present in the FDD_1 and in addition the second-order γnk are not accompanied by events at low η_q values, third-order γnk followed by E1 can be excluded as origin, which leaves very early iiAPs as the only possible cause.

Dependence of afterpulses on threshold

The CDD (or CTD) is very helpful to determine which effects can be expected in a use case depending on the selected threshold. Above the usual cut at 25% pe not many effects remain, and the greatest part of the AP rate is excluded.
The dependence of the remaining fraction of the EAP rate on threshold is shown in fig. 2.64 together with the remaining PP fraction. After an initial drop, which is steeper for the EAPs, both curves become flatter.
The ratio of residual PPs to APs, which can be interpreted as a type of signal-to-noise ratio, continues to grow until a threshold of 1 pe, however, and thus does not provide a natural choice of the threshold point. Typical thresholds between 15 and 25% pe offer a good compromise between AP suppression (46.6–34.0% remaining) and PP preservation (80.3–71.0%). On p. 542 a more sophisticated method to select the threshold, which best discriminates EAPs while preserving PPs, is introduced.

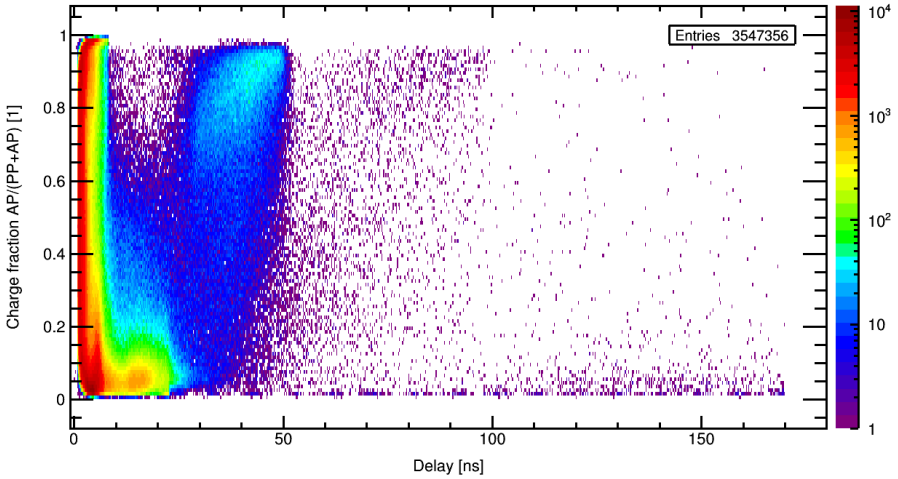


Figure 2.63: FDD for first afterpulses only (FDD_1), same measurement as in fig. 2.61, linear y-axis and extended delay range. Discussion see text

Dependence on voltage and series

On p. 528 the influence of HV, the variation with PMT type, and the dependence on whether a classic or a pulse fitting analysis is used is covered for the DD, CDD, and FDD as well as the EAP rate, probability, mean charge and height, and the γAP rate (see p. 379).

Comparison to literature

In [335] for the same PMT series as studied here (R5912), early afterpulses are reported to occur according to Hamamatsu between 8 and 60 ns with a rate of 3%, ionic APs between 100 ns and 16 µs. Many new PMTs initially (after 12–24h operation) had higher EAP rates than this, which dropped with longer operation.
The delay range agrees well with the results shown here but is likely limited at short delays due to PPs on the waveform. If instead of the first zero crossing a threshold of 25% pe is used to determine the pulse end, in the APS in fig. 2.50c this is already reached at 8.5 ns, which agrees with the reported minimal delay. The very low EAP probability most probably originates from use of such a high threshold which only accepts γnk, BL1 and possibly a fraction of BL11.
The dependence of EAP rate on operation time is intriguing; no simple explanation presents itself.
The very early onset of iAPs at only 100 ns delay supports the hypothesis that the events observed here after about 110 ns with small charges are in fact ionic interdynode afterpulses.

In [276], R5912 PMTs of the high quantum efficiency (HQE) variant were studied at gains comparable to here and with threshold of about 10% pe. In this publication, the FDD was measured down to about 10 ns delay and clearly shows BL1 from 50 ns down to 10 ns with the same shape as calculated here. In addition, some γnk events centered around 50% are present, as well as some events from pileup of multiple γn1. The γn1 effects themselves are not visible, likely due to threshold.
However, another event class between 40 and 70 ns and $\eta_q = 0$–5% is visible, which is centered around 55–60 ns and 2.5% and is not observed here. No TST effects with

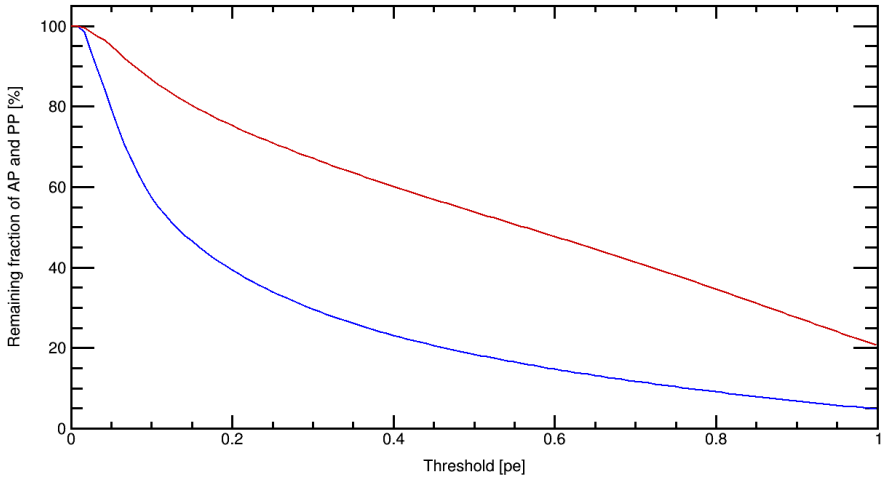


Figure 2.64: Dependence of EAPs and PPs on threshold. The remaining fraction of all APs (blue) and PPs (red) is plotted over the threshold. 1 pe = 2.40 pC (from clean γk pulses). Data from the same measurement as in the previous plots. Discussion see text.

high probability are expected in this region. These events could originate from APS subtraction (see above) or reflections in the electronics.

A similar PMT (Hamamatsu R7081-HQE, 10") was studied in [347] with a threshold of 0.3 pe. A CTD is plotted, which unfortunately is mostly covered by a strong presence of dark counts. The only visible features are the main peak with a trailing flank at regular charges (possibly setup fluorescence), L1 and ΛL1, and the upper part of Rm is vaguely perceptible. E1 and γ1 are completely missing due to the high threshold. In addition, a DD and a CDD are shown for delays > 25 ns. Due to the large minimal values for delay and charge, only γnk and BL1 are expected to be visible. Most EAPs are observed between 25 and 40 ns with regular charges and a peak at 25 ns. This suggests that the γnk peak occurs at even earlier dynodes than for the R5912 and that the photon transport losses between dynodes are larger.

Photonic afterpulses

The possible mechanisms of photon emission inside the PMT were summarized on p. 353 and are discussed more extensively in 2.6.1. Using TST (see 2.3.5.3) allows to calculate the probabilities of emission and transport onto a photosensitive structure, where the photon can create a photoelectron through external photoeffect. This can occur on the cathode, another dynode, the same dynode as for photon emission (in this case an additional electron is added to the cascade which increases its charge), the aluminized glass, the focusing grid if present, and possibly also the dynode mount, since the latter two are coated with cathode material as a side effect of the cathode production. Amplification of this photoelectron results in an additional pulse, which depending on total transit time either is an AP or a PP whose mother pulse becomes the AP.

Electrode glow

Early afterpulses resulting from internal photon production have long been known and were reported by many authors, including [332] [173; p. 56] [195, 232] [172; p. 4-40] [190, 203, 264, 321, 347] [170; pp. 435–439, 458]. In the literature this is

referred to as anode glow, dynode glow or electrode glow and corresponds to the γnk effects (discussion see p. 324), especially of the last stages. Its origins likely are a combination of bremsstrahlung, cathodoluminescence and excitation/ionization of residual gas (see p. 379). Electrode glow increases with gain due to the higher number of impact particles [232] [203; p. 21]. If the signal is intense enough, photon emission at the anode and the last dynodes can even be observed in the dark with the naked eye [170; p. 438].
The delay after the PP is reported as roughly equivalent to T_0 or T_{L1} [170,232]. More exact measurements performed by [264] with an external observer PMT show that for the ET[146] 9350 (8") the peak photon emission occurs later than both these times, which is in agreement with TST predictions for γak.
Electrode glow is considered the reason of electrical breakdown at high voltages by [170; p. 438]. To reduce γAPs and extend gain range, manufacturers use photon traps such as shields around the dynode stack and opaque support wafers [173; p. 56] [203] [170; p. 438]. The enclosures around the dynodes typically are at the potential of a middle dynode to also serve as ion traps [170] to reduce the iAP rate. Based on measurements performed during this work together with the Cherenkov Telescope Array (CTA) collaboration, Hamamatsu extended the dynode stack light shield of the R11920-100 PMT (1.5"), which greatly reduced the light emission through external paths (see p. 429) [181,182].

Rate

The AP rate from γn1 effects can be estimated through cuts excluding electronic effects, γnk pulses, secondary effects and artifacts. A rate of 13.61% is obtained for delays between 11.5 and 28 ns and AP charges from 2.5 to 20% pe. γ11 to γ31 cannot be included due to strong overlapping electronic effects but have small rates anyway. Furthermore, a small contribution from stage-skipping BL11 could occur, which are inseparable from the main γn1 effects. Despite these reservations, the rate is in very good agreement with the observable AP rate of 13.23% predicted by TST (see also p. 110).

Obtaining a clean rate for γnk is more difficult, since at low charges secondary electronic and γn1 effects overlap, and at high η_q BL1 is present, which excludes a large fraction of γn1. For delays between 24 and 52.5 ns, AP charges > 40% pe and $\eta_q < 0.68$ thus only a rate of 0.29% results, which is far from the predicted value of 3.66% and can only serve as a lower limit.

γn2 effects are expected between 0.6 and 30.5 ns with a total calculated rate of 2.80%. These events are difficult to separate from artifacts, for which reason no measured rate could be obtained.

Between the ranges dominated by electronic APs (up to ≈ 11.5 ns) and iiAPs (above ≈ 110 ns), a total AP rate of 21.36% is found. This is dominantly from γAPs, but BL1 (TST-predicted rate 4.1%) and a small fraction of BL11 (about 0.4% predicted) are also contained in this number. If the values from TST are subtracted, in total a rate of about 16.9% is expected to be caused directly or indirectly (through additional interactions) by photonic APs. When including only APs with delays between 11.5 and 110 ns, charges > 2.5% pe and $\eta_q < 0.68$, a rate of 16.38% is measured, which is in good agreement with this number.

[146] Electron Tubes Ltd.

2.5.2.3 Ionic afterpulses

Ionic afterpulses (iAPs) occur when an atom or molecule of the residual gas in the vacuum is ionized by an electron of the cascade [170; pp. 440–449]. The electron loses part of its energy through ionization and one or more additional electrons are knocked out. All electrons are then accelerated onto the next stage where they produce SEs. Due to its positive charge the ion is accelerated in the opposite direction, towards earlier stages, where it ejects one or more secondary electrons on impact (ion-induced secondary electron emission). These SEs start individual but overlapping cascades, which are registered as one afterpulse. Due to the large ion mass the drift times are far longer than for electrons and iAP delays for large PMTs range from about 100 ns to several tens of µs, depending on ion type, position of creation and impact stage. Electronic and photonic APs on the other hand have delays from 0 ns up to ca. 100 ns.

In addition to the problems any source of afterpulses introduces (see p. 352), if a coincidence of signals from multiple PMTs is used as event trigger – as in all large neutrino detectors so far – the large charge of iAPs can cause fake triggers and sets a lower limit to the detector threshold [377] or requires to use an algorithm to exclude this behavior.
In [170; pp. 440–449] a summary of the present knowledge of the mechanisms producing iAPs is given. Extensive studies of this AP type were carried out as soon as in the early 1950s, and most findings already could be established in them [378–380].

In the following, TST will be extended to include iAPs and complement the existing iAP theory by additional effects. Different types of iAPs can be distinguished based on region of creation, impact stage, ion type, secondary interactions, and even target element in case of scattering. Therefore first a nomenclature is introduced to classify iAP behavior and facilitate discussion. Please refer to 2.3.5 for the used notations. Numeric values are based on the same reference PMT as for EAPs (Hamamatsu R5912 at +1425 V).
No complete treatment of the effects as for EAPs was carried out, since many variables were not known (secondary emission yield (SEY) for ion impact, ionization cross-sections), the delays strongly depend on the exact field layout, and deviations from simple field shapes hinder effect identification. The focus is put on analytic delay calculations; charge and probability are only discussed qualitatively. The material presented here can be the basis of further studies.

Nomenclature

Analogous to photonic APs, the term ionic AP shall refer to all afterpulse types based on ions, including those with other secondary interactions.

Effect names are based on all relevant interactions. Since the ion origin plays a great role, after every ionization Φ its location is noted (vacuum v_{xy} between stage x and y). In the effect label the origin region is shortened to the two location letters xy, and the ionization Φ is replaced by the created ion i, since this allows to specify the ion species (see below).

For example, if a pk creates an ion between k and d1, the topology notation is

$$\begin{array}{l} \mathrm{pk}\Phi\mathrm{v_{k1}} \rightarrow (\mathrm{pk{+}sik}){:}\mathrm{R}{:}1 \\ \qquad\qquad\quad \searrow \mathrm{ik1{:}R{:}k} \rightarrow \mathrm{skR1} \end{array}$$

leading to the effect label pk:i:k1-Rk (pk creating ion in v_{k1}, which regularly creates SEs on k). Note the creation of an additional electron through the ionization in v_{k1} (si:k, short sik).
The standard ionizing particle for v_{k1} is pk, for $v_{n(n+1)}$ it is sn. The origin region can be described without ambiguity through the earlier stage x alone (k instead of k1). The standard amplification of the ion is Rx. Standard components can be omitted, so the label can be abbreviated to ik.
An sn creating an ion in $v_{n(n+1)}$, which creates SEs on dn, has the topology

$$\begin{array}{l} \mathrm{sn}\Phi\mathrm{v_{n(n+1)}} \rightarrow (\mathrm{sn{+}sin}){:}\mathrm{R}{:}(\mathrm{n{+}1}) \\ \qquad\qquad\quad \searrow \mathrm{in(n{+}1){:}R{:}n} \rightarrow \mathrm{snR(n{+}1)} \end{array}$$

and the label sn:i:n(n+1)-Rn, abbreviated as i:n with a colon to distinguish it from the preposition. Again, an additional electron is created by the ionization within $v_{n(n+1)}$ (si:n, short sin).

For different interactions on the ion impact stage than regular amplification, the interaction is noted down in place of R, as usual. The standard ion for a stage can be abbreviated with i. For k this is ik1, for dn it is in(n+1). For example, pk:i:k1-Lk-Rk describes a pk producing an ion in v_{k1}, which is elastically backward scattered in the cathode, refocused onto k, and produces SEs there. This can be abbreviated as iLk, since pk is the standard generating particle for ik1, ik1 is the standard ion for k interactions, and Rk can be omitted.

For all effects, subtypes with different ion species can exist. This is denoted by replacing i with the actual ion, such as in H^+k (pk:H^+:k1-Rk).

If an effect requires specification of the scattering partner of the ion, the target atom is added as subscript to the interaction symbol. For instance $H^+L_{Cs}k$ is short for pk:H^+:k1-L_{Cs}k-Rk, i.e. a pk creating an H^+ ion in v_{k1}, which is backward scattered on a Cs atom in the cathode, accelerated back onto k, and generates SEs on the re-hit.

Effect groups summarize effects depending on ion origin or impact location:

- For **predynode iAPs** (**ikRx**, x = k or f) the ion is created in the space v_{k1} between k and d1 and can thus only hit the cathode or the focusing grid.
- For **interdynode iAPs** (**inRx**, x is a previous stage) the ion is generated in the regions v_{xy} between dynodes and can hit either a previous dynode or the cathode.
- The term **cathode iAP** (**ixRk**, x = k or n) describes effects where the ion hits the cathode, independent of origin (v_{k1} or $v_{n(n+1)}$).
- For **dynode iAPs** (**ixRn**, x = n or later region) the ion hits a dynode and thus must originate from interdynode space.

Effects

Initially, the origin of ions was not clear and the hypothesis of an adsorbed layer of residual gas on the dynodes was discussed [170; pp. 440–446]. This was refuted by [381] through the argument that the field at the dynode surface cannot allow the escape of both electrons and positive ions at the same time. The currently accepted theory of iAP origin is ionization of residual gas in the vacuum between stages.

ik

The dominant effect is ik with its different ion subtypes. An iAP is only observable if at least 1 sk is emitted, in which case it has regular charge, but the charge can extend up to several ten pe for large SEY. The slow ion drift velocity causes delays between several hundred ns (see table 2.4) and around 20 µs [225], depending on ion mass and charge, PMT size and voltages.

Ion impact can generate shk and sak, which traverse x_{k1} 4.4 and 5.5 ns faster than sk. Without pulse shape analysis, these pulses will, however, overlap with the sk-pulses and be detected as one broad pulse. Such transit time differences of the individual cascades due to the SE energy distribution can occur for all iAP types.

Interdynode ionic afterpulses

Interdynode iAPs [232, 382] [172; p. 4-40] [170, 180] typically will hit the next earlier dynode (i:n), but depending on origin and fields the ions can also diffuse between dynodes and hit earlier stages (inRx, x < n), including the cathode (inRk). The latter is especially problematic in linear-focused designs unless the dynodes are carefully aligned [180, 382]. inRk could be measured by Coates up to d5 for different ion species (H^+, He^+ and presumably O^+, although Wright [170] considers N^+ or $CH_4{}^+$ more likely).

In addition to sn, ions can also be created by pk passing d1 (such as in p3) through the effects pk:i:1-Rx (x = k or 1) and pk:i:2-Rx (x = k, 1 or 2).

The AP charge for a dynode impact will be reduced by the δ of the missed gain steps, and typically is smaller than 1 pe, since $\delta 1$ is very high. The delays are also shorter than for ik, since the interdynode distances are far smaller and the field gradients are higher. The interdynode ion creation rate increases with the number of SEs in the cascade and is highest for the last stages, but the AP charges of i:n are also smallest there.

Impact on other structures

Ion bombardment of an accelerator grid (ikRf or inRf) creates no iAPs, since the sf cannot escape (sfTf). If the grid is at a potential between k and d1, ikRf is possible and will have very short delays compared to ik.

The deck (ikRm) should not be hit, since it typically lies at d1 potential and positive ions are accelerated away from it.

The aluminized glass can be hit (ikRg or inRg), as it has the same potential as the cathode. The resulting delays of ikRg should be similar to ik, but likely at

most one sg will be emitted, since the SEY for electrons already was very low. This should result in about regular charges and suppress the occurrence probability.

In principle also an ion hit on the inside of the dynode mount (inRmi) could produce one SE on occasion, which would result in an undersized AP.

Negative ionization

Negative ionization is also theoretically possible for very low electron energies. In this case, the ion drifts in the same direction as an electron.

If the ion is created in v_{k1}, the pk disappears and the ion produces a *primary* pulse on d1 with very long transit time (i^-kR1).
Since the cross sections for negative ionization are largest for energies in the eV range [383], the ion origin drift distance is $\approx x_{k1}$ and the transit time can be calculated with the equations for the electron (e.g. (2.141)) by substituting the ion mass. Using the mean t_{k1} of linear and quadratic fields, assuming an initial ion energy of 1 eV (t_{k1} is infinite for $U_0 = 0$ in a quadratic field) and neglecting the pk transit time to the ion, this results in immense transit times to d1 between about 1086 ns for H^+ and 25050 ns for Cs_3Sb^+ (4109 ns for N^+). Since the SEY of ion and electron impact on d1 are different, the PP charge will differ from γk.

For ion creation in v_{12} the ion will hit d2 (i^-1R2) and produce an AP, since the rest of the s1 electrons create the PP. The charge will be reduced due to the skipped $\delta 1$ and possibly due to the altered SEY. Using the same assumptions as for i^-kR1, the transit time to d2 will range between ca. 229 and 5207 ns.

For later $i^-nR(n+1)$, the charge will likely be too small to be detectable.

Secondary interactions

Ions can also be subject to elastic (ε) or inelastic (Λ, B) scattering. Forward (E) and backward (L) scattering are defined with regard to the normal movement direction of a particle, and for positive ions thus *reverse* to electrons.

Elastic scattering losses

Unlike electron backscattering, the elastic scattering loss is now *not* negligible anymore since the masses of projectile and target are similar. The energy fraction which the backscattered particle retains (kinematic factor K) for Rutherford scattering is [337; p. 60]

$$K = \left(\frac{\cos(\theta) + \sqrt{\left(\frac{m_2}{m_1}\right)^2 - \sin^2(\theta)}}{1 + \frac{m_2}{m_1}} \right)^2 \tag{2.208}$$

where θ is the scattering angle, m_1 the mass of the projectile particle, and m_2 the mass of the target atom. The backscattered energy thus depends on the target atom, the ion and θ. K increases with $\frac{m_2}{m_1}$ and decreases with θ.
If $m_2 \geq m_1$, the maximum scattering angle θ_m is [337; p. 61]

$$\theta_m = \arcsin\left(\frac{m_2}{m_1}\right) \tag{2.209}$$

For 180° backscattering on a K_2CsSb photocathode or a Cs_3Sb dynode and typical ion species, the maximum K results from scattering on Cs, and amounts to 0.970 for H^+, 0.886 for He^+, 0.655 for N^+, and 0.289 for Ar^+; for Cs^+ $\theta_m = 90°$ and no backscattering can be observed. For each ion subtype of a scattering-based effect, therefore different variants are expected depending on the target element.

Backscattering

Time shift effects for ions can be derived in the same manner as for electrons. Since iAP rates are far smaller than EAP rates, it suffices to treat one additional interaction and neglect ion impacts beyond d2.

The most common effect should be from elastic backward scattering from the cathode towards d1, after which the ion is decelerated and refocused onto the cathode, where regular SE production takes place. This can occur through ikLkRk (short iLk) and inLk. To calculate the delays of these effects, a transit time formula for ions moving in reverse directions is required, which will be derived shortly. As will be seen, for a quadratic potential and scattering energies smaller than U_{k1} (eq. (2.220)) the transit time is independent of the backscattered energy, but depends on the ion mass and charge. The backscattered energy also affects the impact energy and thus AP charge.

For inelastic scattering through ixBLk, two pulses with charges correlated by their delay are produced (three pulses in total), for ixΛLk only one pulse with delay-dependent charge is created. Inelastic scattering can also occur for all following effects.

inLkR1 (n $\geq$1) are not expected to occur with notable probabilities, since the angular phase space for which d1 is hit is very small and the ions are most likely focused away from d1.

Scattering on the grid will have small probabilities due to the fill factor. ixLfRk should introduce some delays compared to iLk effects, inLfR1 (n $\geq$ 1) is late compared to inR1, and ixEfRk is also later than ixRk due to the loss of kinetic energy.

Ion scattering on dynodes can occur via ixL1R1, ixE1Rk, ixE2R1, inL1R2 (n $\geq$ 2), which all introduce lateness compared to the effects without scattering.
Intradynode ion scattering of the type iL11 and iE11 should also be possible and introduce extra delays compared to i1.

Luminescence

The ion impact is also expected to produce photons through luminescence, which can cause second-order APs either preceding or following the iAPs.
For ixRγk1 two pulses are created, where the first is small and earlier than the ion pulse by the same time as γ1 ($\approx$ 24 ns), and the second has regular iAP charge.
ixRγ1k starts two pulses, of which the first is undersized, and the second has regular charge and arrives with a delay of $t\gamma$1k + tk1. The probability should be far higher than for ixRγk1 due to the larger electron number in the cascade and the favorable solid angle of the cathode.
ixRγ12 creates two pulses, the first very small and a few nanoseconds earlier, the second undersized. This effect will have a very small probability due to threshold.

Time shift effects

The secondary electrons created by an ion impact can be subject to all possible TSEs, including EAPs. This will be present as additional effects following all iAP populations with short delays and typically small charges.

Ion types

The ion species causing iAPs can be studied by introducing known gases into the PMT vacuum and observing which iAP peaks increase [170; pp. 440–449]. This can be done by connecting the PMT vacuum to a gas handling system [380, 381]. Alternatively, the PMT can be exposed to helium which diffuses through its glass and increases the corresponding peak. This is possible by comparing of iAP spectra before and after exposition through an enriched atmosphere [375] or through long-term diffusion of He contained in the (laboratory) atmosphere [382]. The delay calculations for different ions can be compared to the observed peaks either by value or by the delay ratio between peaks. This is aided by unambiguous identification of at least one peak through gas introduction.

Reported ion species comprise H^+, $H_2{}^+$, He^+, N^+, $CH_4{}^+$, H_2O^+, CO^+, $N_2{}^+$, Ar^+, $CO_2{}^+$, Xe^+, Cs^+, and Hg^+ (ordered by mass) [170, 375, 376, 380–382, 384].
Not all ion species are observed in all devices; usually only a couple of dominant peaks are present. In most cases either H^+ or $H_2{}^+$ in addition to He^+ and N^+ (or an ion with similar mass) are found.
Some authors also report O^+ or $O_2{}^+$ peaks, but Wright [170] considers this debatable, since oxygen is highly reactive and unlikely to exist in free form, and attributes this to either N^+ or $CH_4{}^+$, which have similar atomic weights, where the latter is said to be a well-known residual in vacuum devices.
An Hg^+ peaks stems from the vacuum pump used to evacuate the PMT.
Wright [170] considers the latest discernible features in [380] to possibly result from Sb–Cs compounds (dynode resp. cathode materials). Since [376] also reports a Cs^+ peak, it is possible that Sb^+ also causes iAPs.
Morton [381] and Wright [170] observed groups of later APs to disappear after a short time of operation at moderate light level, exposure to ambient light, or too high voltages. Wright observed that the iAP types always reappeared after switching off the device and attributes this to a gettering action.

In the following analyses, only ion species previously discussed in literature and with single charges are taken into closer consideration. The presence of doubly charged ions such as N^{2+} is unlikely due to low generation probabilities and such types are only considered in exceptional cases.

Influence on primary pulse

One point which is frequently neglected is that during ionization one or more additional electrons are produced (depending on ionization state) – e.g. through $A + e^- \rightarrow A^+ + 2e^-$ – which subsequently are focused onto d1.

The kinetic energy of the pk, minus ionization energy and a small momentum transfer to the ion, is split between all electrons. Therefore the PP actually consists of two or more parallel cascades. Depending on where between k and d1 the

ionization takes place, the particles gain additional energy in the field, and the sum $\delta 1$ can overcompensate the ionization loss and make the PP *larger* than without AP production; the SEY per energy is higher for small energies, so this constitutes an extra amplification step. For an ion created near the cathode, the PP charge approaches 2 pe, near d1 it will be close to 1 pe [170; pp. 445–446]. This was already experimentally verified by [385].
As will be shown, for quadratic potentials the ion drift time to k does not depend on its origin in v_{k1}. However, for deviations from a quadratic field the dependence of q_{PP} on position might allow to statistically reconstruct the origin for individual effects.
The possible PP charge range and the occurrence probability also depend on ionization position, since the ionization cross section is energy dependent and thus the ionization probability varies with the distance to k.

The PP should also occur somewhat later than T_0 due to the energy loss and subsequent re-acceleration of the pk. In addition, the PP should often show a broader shape because of the energy range and emission direction of the knocked out electron(s), which cause time differences of the PP cascades. Furthermore, the transit time spread could be slightly reduced due to the averaging effect described on p. 169.

Delays

For simple field functions the transit time of ions can be calculated analytically. Since positive ions traverse the PMT fields in reverse direction than electrons, the formulas derived for electrons in 2.3.4 and 2.3.5.5 cannot be employed, but the calculation of the equations for ions is analogous. The subsequent derivation for linear and quadratic fields in forward direction is based on [170; pp. 442–443] which itself follows [382, 386]. All other formulas were derived for the present work.

Forward movement

Let the ion have charge $Z\,e$, mass m, initial energy $E_0 = \frac{1}{2} m\, v_0^2$ with v_0 the initial velocity. The ion shall move along a one-dimensional potential $U(x)$ where x is the distance from the impact surface (e.g. the cathode) and x_0 denotes the position where the ion is created. Then the kinetic energy after traversing the potential difference $U(x) - U(x_0)$ equals

$$\frac{1}{2}\, m \left(v^2(x) - v_0^2\right) = Z\, e \left(U(x_0) - U(x)\right)$$

The initial kinetic energy corresponds to an effective potential U_0 which relates to the initial velocity v_0 through

$$Z\, e\, U_0 = \frac{1}{2}\, m\, v_0^2$$

The velocity v then reads

$$v(x) = -\sqrt{\frac{2Ze}{m}} \cdot \sqrt{U(x_0) - U(x) + U_0}$$

where v is negative, since the ion drifts towards $x = 0$.

The time t required to reach a position x then is

$$t = \int_{x_0}^{x} \frac{dx}{v} = -\sqrt{\frac{m}{2Z\,e}} \int_{x_0}^{x} \frac{dx}{\sqrt{U(x_0) - U(x) + U_0}} \tag{2.210}$$

Linear field

If a linear field $U(x) = U_1 \frac{x}{x_1}$ is assumed, where $U_1 = U(x_1)$ and x_1 is the distance of the the next stage (e.g. d1) to the impact stage, combination with (2.210) results in the ion transit time $t_{i,l}(x, U_0)$:

$$t_{i,l}(x, U_0) = \sqrt{\frac{2m}{Ze}} \frac{x_1}{U_1} \left(\sqrt{\frac{U_1}{x_1}(x_0 - x) + U_0} - \sqrt{U_0} \right)$$

$$\rightarrow \quad t_{i,l}(0, U_0) = \sqrt{\frac{2m}{Ze}} \frac{x_1}{U_1} \left(\sqrt{\frac{U_1}{x_1} x_0 + U_0} - \sqrt{U_0} \right) \tag{2.211}$$

$$\rightarrow \quad t_{i,l}(0, 0) = \sqrt{\frac{2m}{Ze}} \cdot \sqrt{\frac{x_1 x_0}{U_1}} \tag{2.212}$$

So the transit time to the surface $t(0, U_0)$ *increases* with the traversed distance x_0. Heavier ions arrive later and the drift time decreases with ion charge Z.
Equation (2.212) can be used for iAPs which hit the next stage, such as ik or in; the more general eq. (2.211) allows to calculate inRx. For example, for i1Rk this is done by first calculating $t_{i,l}(0, 0)$ for the d1–d2 region. Afterwards $t_{i,l}(0, U_0)$ for k–d1 can be determined by plugging in the final energy from the previous acceleration (depends on x_0!) as U_0 and approximating x_0 with x_1.

To obtain the average value over x_0 for $U_0 = 0$, for simplicity a uniform probability distribution over x_0 is assumed. This is only a rough estimate, since the ionization cross section depends on energy and close to the electron emitting stage no ionization is possible at all. Integration over x_0 from $x_{0,\min} = 0$ to $x_{0,\max} = x_1$ yields the average transit time

$$\bar{t}_{i,l}(0, 0) = \int_{x_{0,\min}}^{x_{0,\max}} \frac{t_{i,l}(0, 0)}{x_{0,\max} - x_{0,\min}} dx_0 = \frac{2}{3} \sqrt{\frac{2m}{Ze}} \frac{x_1}{\sqrt{U_1}} \tag{2.213}$$

Quadratic field

For quadratic fields $U(x) = U_1 \left(\frac{x}{x_1}\right)^2$ and from (2.210) follows $t_{i,q}(x, U_0)$:

$$t_{i,q}(x, U_0) = \sqrt{\frac{m}{2Ze}} \frac{x_1}{\sqrt{U_1}} \left(\arcsin\left(\frac{x_0}{\beta}\right) - \arcsin\left(\frac{x}{\beta}\right) \right)$$

$$\text{with} \quad \beta = \sqrt{x_0^2 + x_1^2 \frac{U_0}{U_1}}$$

$$\rightarrow \quad t_{i,q}(0, U_0) = \sqrt{\frac{m}{2Ze}} \frac{x_1}{\sqrt{U_1}} \arcsin\left(\frac{x_0}{\beta}\right) \tag{2.214}$$

$$\rightarrow \quad t_{i,q}(0, 0) = \frac{\pi}{2} \sqrt{\frac{m}{2Ze}} \frac{x_1}{\sqrt{U_1}} \tag{2.215}$$

Again, (2.215) is for iAPs which hit the next stage, whereas (2.214) is more general and allows calculation for inRx in the same manner as for linear fields.
Eq. (2.215) is *independent* of the distance x_0. No matter where the particles are created on the way from one stage to the next, the quadratic field completely compensates the traversed distance and all ions arrive at the impact stage at the same time. This is the reason why for many fast PMTs, where the field between k and d1 is nearly quadratic [170; pp. 443–444] [375], the peaks of different ion types in the DD and CDD are well-defined and not smeared over a large delay range. Since the distance x_{k1} and potential difference U_{k1} are constant, the delay only depends on mass and charge of the ion. This allows to distinguish different ion species and use the PMT as a crude mass spectrometer for its own residual gas, as was already pointed out by [380] (1952).

Cubic field

For a cubic field $U(x) = U_1 \left(\frac{x}{x_1}\right)^3$, and the transit time to the surface ($x = 0$) for $U_0 = 0$ results to

$$t_{i,c}(0,0) \approx 1.4022 \sqrt{\frac{m}{2Ze}} \sqrt{\frac{x_1^3}{U_1 x_0}} \tag{2.216}$$

which *decreases* for larger distances x_0 due to the stronger initial acceleration. The general expression $t_{i,c}(x, U_0)$ is rather lengthy and of not much relevance here.

The average transit time can be derived analogous to $\bar{t}_{i,l}(0,0)$:

$$\bar{t}_{ic}(0,0) = \int_{x_{0,\min}}^{x_{0,\max}} \frac{t_{i,c}(0,0)}{x_{0,\max} - x_{0,\min}} dx_0 = 1.4022 \sqrt{\frac{2m}{Ze}} \frac{x_1}{\sqrt{U_1}} \tag{2.217}$$

A clear relation between the exponent of the field and that of x_0 in the transit time is discernible from the three equations for the ion transit time $t_i(0,0)$ to a surface without initial energy. If the exponent of the field function increases by one, the transit time is multiplied by $\sqrt{\frac{x_1}{x_0}}$. The numerical prefactor decreases from 1.414 over 1.111 to 0.991.
The average transit time ($t_{iq}(0,0)$ does not depend on x_0 and is its own average) is *identical* for the used approximations except for a prefactor, which increases with field exponent from 0.667 over 1.111 to 1.402; the mean transit time is smaller for lower field exponents.

Reverse movement

If an ion is backward scattered off a surface (e.g. iLk), it moves in the same direction as an electron would but is decelerated by the fields. Depending on whether its backscattered energy E_s (effective potential U_s) is larger than the potential difference U_1 to the next stage, it either comes to a halt and is then accelerated back towards its scattering stage, or can reach (or pass) the next stage, for example if an ion from v_{12} is scattered off k.

The reverse ion transit times can be calculated in the same way as for electrons (p. 247).

Using the linear decelerating field $U(x) = -U_1 \frac{x}{x_1}$, one can obtain

$$U_s < U_1 : \quad t_{i,lrs} = \sqrt{\frac{2m}{Ze}}\, x_1\, \frac{\sqrt{U_s}}{U_1} \tag{2.218}$$

$$U_s > U_1 : \quad t_{i,lrg} = \sqrt{\frac{2m}{Ze}}\, x_1\, \frac{\sqrt{U_s} - \sqrt{U_s - U_1}}{U_1} \tag{2.219}$$

The quadratic decelerating field $U(x) = -U_1 \left(\frac{x}{x_1}\right)^2$ leads to

$$U_s < U_1 : \quad t_{i,qrs} = \sqrt{\frac{m}{2Ze}}\, \frac{x_1}{\sqrt{U_1}}\, \frac{\pi}{2} \tag{2.220}$$

$$U_s > U_1 : \quad t_{i,qrg} = \sqrt{\frac{m}{2Ze}}\, \frac{x_1}{\sqrt{U_1}}\, \arcsin\left(\sqrt{\frac{U_1}{U_s}}\right) \tag{2.221}$$

The reverse ion transit time $t_{i,qrs}$ for quadratic potential and smaller energy is identical to t_{iq}, since the latter did not depend on energy, and because electromagnetic movements do not change under time reversal.

For $U_s < U_1$ the time until return to the scattering stage is simply twice the result from the corresponding formula.
For an ion which is accelerated from rest from a position between the scattering stage and the next, the total delay of iLk for the quadratic potential is simply $3\, t_{i,q}(0,0)$, which is *independent* of scattering losses, since U_s does not enter (2.220).
For the linear field and without scattering losses the iLk delay is $3\, t_{i,l}(0,0)$; scattering losses reduce U_s and thus the delay.

Total delay

In order to obtain the total transit time, the time the cascade and the scattering electron take to reach x_0 has to be added to the ion drift time. This is *not* independent of x_0 (see (2.135) and (2.141); in these equations $x_1 \mathrel{\widehat{=}} x_0$). However, the electron transit times lie in the region of a few ten ns for large PMTs, whereas the ion drift times to the cathode are in the µs scale, so the fluctuations of the latter dominate and for quadratic fields no strong delay–position dependence results. This might be different for interdynode APs, which as was seen in the EAP section 2.5.2.2 can have delays of as low as 100 ns, and possibly also for light ions such as H^+, $H_2{}^+$ and He^+.
The transit time t_{k1} from k to d1 also is slightly smaller for the emitted sk (or even shk/sak) than for a pk due to the higher initial energy, which reduces the delay by 3.8 ns for the R5912 when assuming similar SE energies as for electrons. For ion impact on dynodes such a correction is not necessary.
Taking half of the electron transit time between stages as typical time until ion creation, for ik the extra delay apart from ion movement amounts to 8.8 ns, for i1 it is 27.3 ns, and for i2 30.8 ns.

Charge

Theory

The charge of ionic afterpulses depends on missed gain steps (e.g. i1 or i2) and the ion-induced secondary electron yield (ISEY), which itself depends on ion type,

surface material and impact energy. The latter follows from the potential difference between ion origin and impact stage.
Literature research could not produce ISEY curves for PMT materials and common ion types, but general tendencies can be found. The review article of [387] states that the ISEY generally falls with increasing nuclear charge Z of the impact ion. This is in accordance with [381], who found that H_2^+ had a typical AP charge of 4 pe, whereas this number had already dropped to 1 pe for N_2^+, and [180] where early iAPs are observed to be significantly larger than later for a number of PMT types. Furthermore, [387] notes that for insulators much higher ISEY are expected than for metals due to the absence of electrons in the conduction band. This should also apply to the semiconductors forming cathode and dynodes and is the same behavior as for the electron SEY (see 2.1.3).
The mean SE emission energy from ion impact should be similar to that from electrons, and sk should have comparable energies to sn (about 5 eV). If an ixRk produces several sk, the resulting pulse is typically larger than an RP; charges up to several ten pe are reported in literature (e.g. [276]). For ixRn the charge will be underamplified due to skipped stages in most cases, but for instance for ixR1, it can be larger than 1 pe if the ISEY is greater than $\delta 1$.
Finally, for each effect the energy dependence of the ionization cross-sections influences the shape of the charge distribution and its extent, since the crossed potential difference of the ion depends on its origin.
Since the crucial ISEYs were not known, the charge of effects unfortunately could not be calculated.

Probability

In order to predict the probability of a (sub)effect, the energy-dependent ionization cross-section (ICS), the composition and pressure of the residual gas, transport factors onto a structure, and the ISEY have to be known or estimated.
According to [388] the ICS for relevant ion types strongly varies within the relevant energy range. For single ionization the maximum is near 20 eV (Cs), 40 (Xe), 55 (Hg), 60 (H), 70 (Ar), and 100 eV (He, N). In addition, the minimal required energy is non-negligible for most species (Cs 3.89 eV, Hg 10.4, Xe 12.1, H 13.6, N 14.5, Ar 15.8, He 24.6) and introduces notable dead regions near the cathode and dynodes, where no ionization can take place due to too low pk resp. sn energies. Peaks in the iAP DD which were attributed to ions with double ionization have not been reported in literature, likely due to the small probability of occurrence.
The ICS energy dependence influences the probability of an effect, but also the distribution and ranges of its AP ands PP charges, and for deviations from quadratic fields the delay distribution as well. While the ICS can be incorporated in the theory, this is a matter for another work and probability values thus were not calculated.

If an energy-independent ICS is assumed, the ratio of the ionization probability in two interstage regions can be estimated from the ratio of electrons in the cascade multiplied with the ratio of distances between stages (the interaction volume is proportional to the crossed distance). For v_{12}/v_{k1} this is $\delta 1 \cdot \frac{x_{12}}{x_{k1}} = 2.1$, for v_{23}/v_{k1} it is $\delta 1 \cdot \delta 2 \cdot \frac{x_{23}}{x_{k1}} = 12.3$, and for v_{34}/v_{k1} 131.5. This indicates that more ions are produced in the space following the k–d1 region, which could indicate that instead of ik the dominant causes for ion hits on the cathode actually are i1Rk and i2Rk effects, if the transport factors out of the dynode chain are high. Due to the dynode layout

of the R5912 – which allows effects like p3 – contributions are chiefly expected up to i2Rk; for i3Rk, the ion trajectories would need to be strongly curved, which is unlikely.

Ionic afterpulse rates for large PMTs (see 3.1) usually lie between 2 and 10% for delay ranges typically from about 100 ns to 10–25 µs.

Analysis for R5912

The results of a low-occupancy iAP measurement for the R5912 are shown in figs. 2.65 (DD_a and DD_{1u}), 2.66 (CDD_a and CDD_1) and 2.67 (FDD_a and FDD_1). While the CDD and FDD help to understand the behavior and attribute effects to clusters, the DD better shows the total probability of effects since it is a projection of the CDD on the delay axis.
A coincidence between PP and AP was used to increase statistics. To avoid fake AP triggers from pulse shape oscillations following large PPs, delays below 83 ns were excluded. In order to allow LAP search, the coincidence window covered delays up to 100 µs. The susceptibility to noise spikes in the long iAP acquisition windows requires a larger threshold than in the EAP measurement. Since the focus was on the iAP occurrence, secondary EAPs from iAPs are of lower priority and a classic analysis sufficed.

Rates

The iAP probability p_a was determined to 5.77% from a simple iAP measurement using the same settings (reason explained on p. 549) for a software threshold of 34.1% pe. This value drops to 4.10% after DCR subtraction and unshadowing.
For the final iAP rate r_a a value of 5.36% was measured, whereas the charge rate $r_{a,q} = 10.46\%$ is far higher. The mean charge amounts to 1.87 pe, which shows that oversized pulses dominate over undersized iAPs. However, the analysis threshold excludes a large fraction of underamplified pulses, so the true value will be somewhat lower. In any case, judging from fig. 2.66 the widespread assumption that iAPs are larger than RPs is a pitfall if one presumes that iAPs can be reliably eliminated by a large threshold: The majority of iAPs have regular or undersized charges and only the high charges of several ixRk effects (identification see below) raise the mean charge above 1 pe.
The heightened charge rate results mostly from the high mean charge, with additions from the mean AP number $n_a = 1.31$ (after DCR subtraction and unshadowing). This figure is also of great interest, because it implies that despite the small iAP rate, the probability for secondary APs is quite high; higher than explainable without correlations. This can be attributed to the large charge iAPs which produce multiple sk electrons, each of which can induce EAPs and iAPs in its cascade. Thus, the total probability of secondary APs is greatly enhanced. This hypothesis is verified through comparison of the first-AP and all-APs variants of DD, CDD and FDD: Where large charge iAPs are present (mostly 0.4–3 µs), a strong increase of low-charge secondary events is observed in their wake (EAPs, inR1), while for regular-sized APs (4–6 µs) the enhancement is far less pronounced. Since the secondary APs are mostly undersized, their contribution to the charge rate is only small compared to that from the mean charge of primary iAPs, despite their large mean occurrence probability.

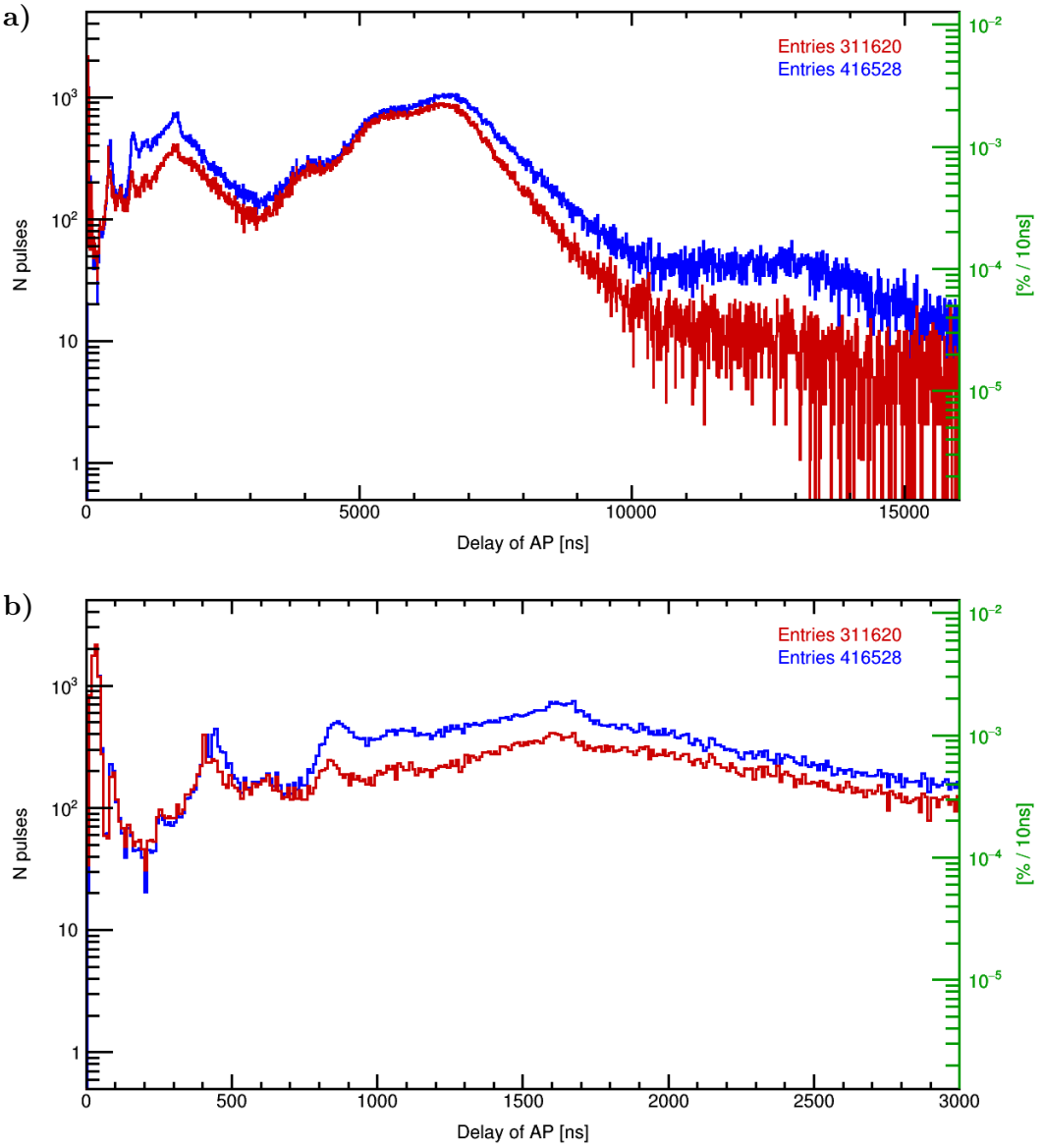


Figure 2.65: Delay distribution (DD) of ionic afterpulses (iAPs) for Hamamatsu R5912. Voltage +1425 V, gain $g = 1.3{\cdot}10^7$, occupancy $o = 3.5\%$, trigger on coincidence of afterpulse (AP) and primary pulse (PP), 10^6 waveforms, classic analysis, threshold 24.6% pe, 1 pe = 2.40 pC (γk pulses), DCR subtracted.
Distribution for all APs blue (DD_a), unshadowed distribution for first APs red (DD_{1u}). Delay range 0–16 µs in upper, 0–3 µs in lower plot; extended range of 0–100 µs shown in fig. 2.68c. Discussion see text.

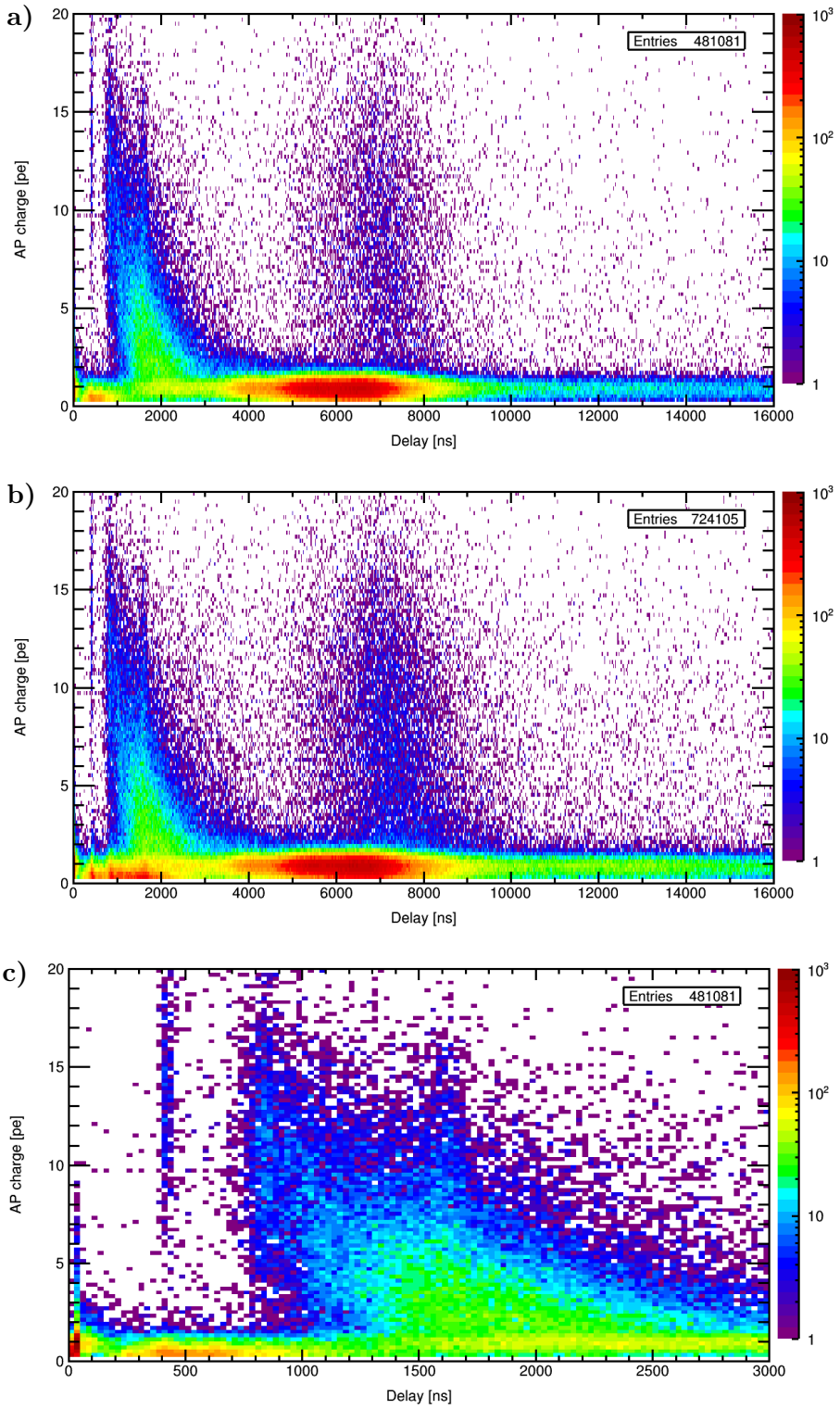


Figure 2.66: Charge–delay distribution (CDD) of ionic afterpulses for R5912. Number of pulses per bin indicated by z-axis (colors). Data as in previous figure.
a) First afterpulses only for 0–16 µs (CDD_1).
b) All APs for the same delay range (CDD_a).
c) CDD_1 for 0–3 µs. Discussion see text.

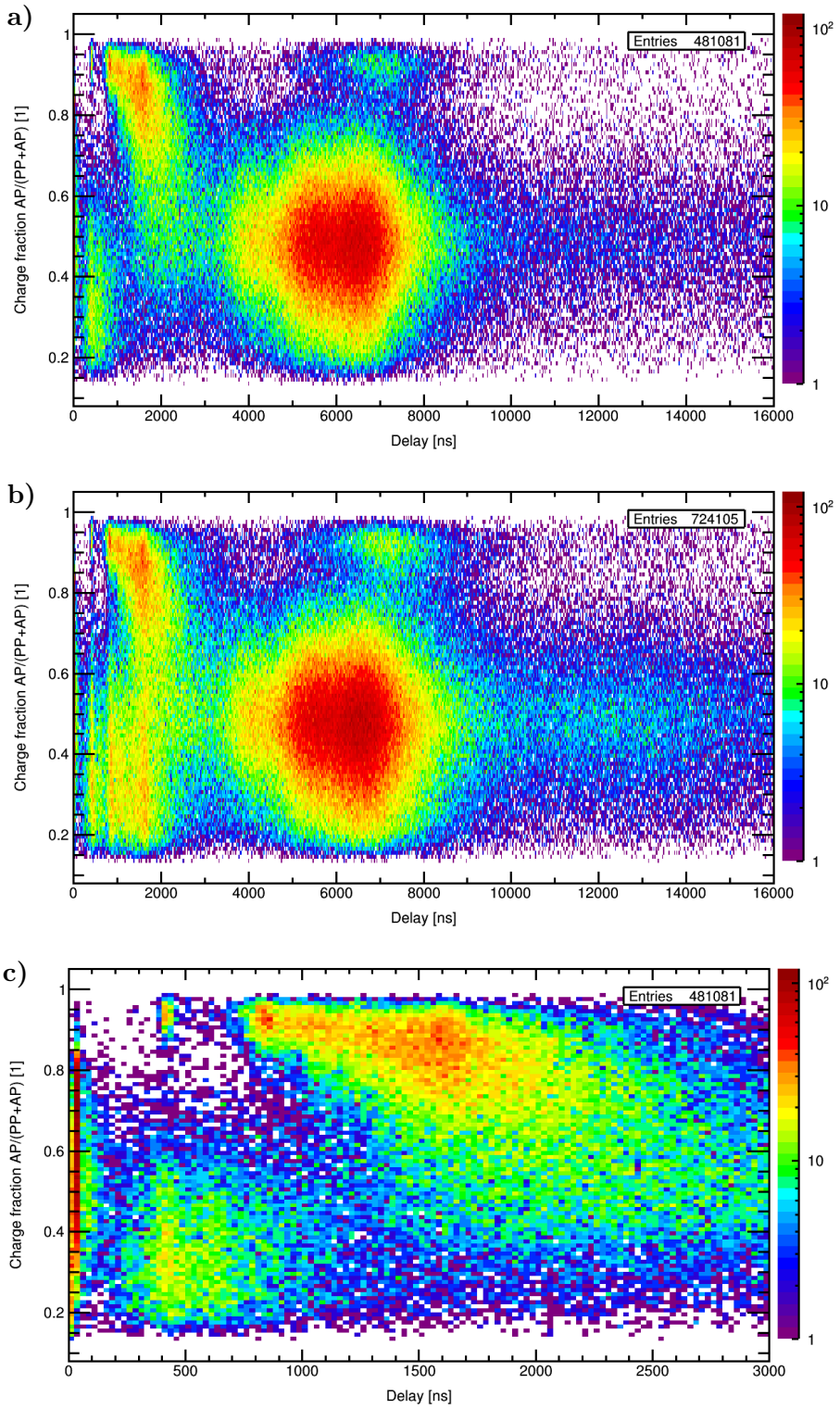


Figure 2.67: Charge fraction–delay distribution (FDD) of ionic afterpulses for R5912. Number of pulses per bin indicated by z-axis (colors). Data as in previous figures.
a) First APs only for 0–16 µs (FDD_1).
b) All APs for 0–16 µs (FDD_a).
c) FDD_1 for 0–3 µs. Discussion see text.

Delay calculations

As was discussed in 2.3.4 and 2.3.5.5, the fields of the R5912 appear to be similar to a combination of quadratic and linear components from studying electronic transit time. This could introduce some smearing of the peaks due to the linear contribution whose delay scales with the root of the covered distance. In addition, even for purely quadratic fields, for inRx the delays depend on the initial position according to (2.214) and decreases with growing x_0. This means that the closer the ion is created to the next stage, the larger is its traversed potential and thus its impact energy, and the shorter its delay, which would result in a falling charge–delay dependence. However, in the ionic afterpulse DD, CDD and FDD all observed peaks are very sharp and no smearing is observed; the structure between 1 and 3 µs at $q > 1$ pe in the CDD is actually caused through the overlap of several effects, whose mean charge decreases with delay. Likely the inRx charge–delay dependence is present but confined to small delay differences and thus not well discernible.
Nevertheless, sharp iAP peaks are not in tension with a combined field, because the movement directions of electrons and ions are opposite. Positive ions can only be created some distance away from k after the pk gains sufficient energy for ionization. The linear component is only expected to dominate near the cathode where it has the purpose of facilitating pk extraction. At larger distances the quadratic fields dominate and determine the transit times, since the largest part of the acceleration is from the quadratic component. Thus ions can have well-defined peaks due to quadratic acceleration, whereas electron behavior shows influence from the mixed fields.

The delays of ik, i1, i2, i1Rk, i2Rk and i2R1 were calculated for the quadratic fields for all expected ion types. For ik, i1 and i2 the delays were also calculated for linear fields. The results of are listed in table 2.4.
The calculations show that – depending on ion type and effect – iAPs can cover a large delay range from 100 ns (H^+2) to over 13 µs (Cs_3Sb^+k). Secondary and tertiary APs can extend even further which explains why in large PMTs iAPs are routinely observed up to 25 µs or more.
i1Rk has very similar delays to ik for quadratic fields despite the larger covered distance, because of the initial energy of the ion when crossing the large k–d1 region, which it acquired between d1 and d2. For the same reason i2Rk has actually *shorter* delays than ik.
The delays are far smaller for i:n than for ik due to the shorter distance, and smaller for i2 than i1 due to the reduced distance and larger interstage potential.
For the iAPs originating in $v_{n(n+1)}$ the delays depend on the exact field layout and are even more sensitive to deviations between model and device than for EAPs due to the slow ion movement. Therefore only coarse estimates are possible for i1Rx and i2Rx.

The calculations for ik peaks in the table for the $\bar{t}_{i,l}$, $t_{i,q}$ and $\bar{t}_{i,q}$ columns are in tension with some prominent peaks in the CDD and DD at 5.3, 0.4 and around 1.6 µs. There is a significant gap in ik delays between $CO_2{}^+$ and Sb^+, in which neither PMT materials nor atmospheric gases lie (except for the very seldom Kr), and a similar gap between He^+ and N^+. In addition, the earliest observed high-charge peak at 405 ns should be caused by an effect based on H^+ or $H_2{}^+$, since this is observed in most PMTs; however, the calculated values for ik, i1Rk and i2Rk are far larger. These discrepancies are likely caused by the linear field component. Since the ICS has its

Ion	ik			i1Rk		i2Rk		**i1**		**i2R1**	**i2**	
	$\bar{t}_{i,l}$	$t_{i,q}$	$t_{i,q,c}$	$\bar{t}_{i,q}$	$\bar{t}_{i,q,c}$	$\bar{t}_{i,q}$	$\bar{t}_{i,q,c}$	$t_{i,q}$	$\bar{t}_{i,l}$	$\bar{t}_{i,q}$	$t_{i,q}$	$\bar{t}_{i,l}$
H^+	0.51	0.60	**0.47**	0.62	**0.48**	0.55	**0.43**	**0.17**	0.15	**0.20**	**0.12**	0.10
$H_2{}^+$	0.72	0.85	**0.66**	0.86	**0.67**	0.76	**0.59**	**0.22**	0.19	**0.27**	**0.15**	0.13
He^+	1.02	1.20	**0.93**	1.21	**0.94**	1.07	**0.83**	**0.31**	0.26	**0.37**	**0.20**	0.17
N^+	1.89	2.23	**1.73**	2.24	**1.74**	1.97	**1.53**	**0.55**	0.47	**0.66**	**0.35**	0.30
O^+	2.02	2.38	**1.85**	2.39	**1.86**	2.10	**1.63**	**0.58**	0.50	**0.71**	**0.37**	0.32
$CH_4{}^+$	2.03	2.38	**1.85**	2.39	**1.86**	2.11	**1.63**	**0.58**	0.50	**0.71**	**0.37**	0.32
H_2O^+	2.15	2.53	**1.96**	2.54	**1.97**	2.23	**1.73**	**0.62**	0.53	**0.75**	**0.39**	0.33
CO^+	2.67	3.15	**2.44**	3.16	**2.45**	2.77	**2.15**	**0.76**	0.65	**0.92**	**0.48**	0.41
$N_2{}^+$	2.67	3.15	**2.44**	3.16	**2.45**	2.77	**2.15**	**0.76**	0.65	**0.92**	**0.48**	0.41
$O_2{}^+$	2.86	3.36	**2.61**	3.37	**2.62**	2.96	**2.30**	**0.81**	0.70	**0.98**	**0.51**	0.44
Ar^+	3.19	3.76	**2.92**	3.76	**2.92**	3.31	**2.57**	**0.91**	0.77	**1.10**	**0.56**	0.48
$CO_2{}^+$	3.35	3.94	**3.06**	3.95	**3.07**	3.47	**2.69**	**0.95**	0.81	**1.15**	**0.59**	0.51
Sb^+	5.57	6.55	**5.09**	6.55	**5.09**	5.75	**4.46**	**1.56**	1.33	**1.89**	**0.96**	0.82
Xe^+	5.78	6.81	**5.28**	6.81	**5.28**	5.97	**4.64**	**1.62**	1.38	**1.96**	**1.00**	0.85
Cs^+	5.81	6.85	**5.31**	6.85	**5.31**	6.01	**4.66**	**1.63**	1.39	**1.97**	**1.00**	0.86
Hg^+	7.14	8.41	**6.53**	8.41	**6.52**	7.38	**5.72**	**2.00**	1.70	**2.42**	**1.22**	1.04
$CsSb^+$	8.04	9.48	**7.35**	9.47	**7.35**	8.31	**6.45**	**2.25**	1.91	**2.72**	**1.38**	1.17
Cs_2Sb^+	9.92	11.69	**9.07**	11.68	**9.06**	10.24	**7.95**	**2.77**	2.35	**3.35**	**1.69**	1.44
Cs_3Sb^+	11.50	13.54	**10.51**	13.53	**10.50**	11.87	**9.21**	**3.20**	2.72	**3.88**	**1.95**	1.66

Table 2.4: Calculated total delays for ionic afterpulse effects for common ion types for quadratic and linear fields according to eq. (2.215), (2.213) and (2.214) for an R5912 at +1425 V. Values in µs. $t_{i,q} = t_{i,q}(0,0)$, $\bar{t}_{i,l} = \bar{t}_{i,l}(0,0)$, $\bar{t}_{i,q}$ is the mean sum delay from two or more regions, the subscript c indicates a correction with the factor of 0.776 (see p. 397). These values can be compared to the observed peaks in the DD, CDD and FDD. Bold values were used as best estimates. Discussion in text.

peak close to the cathode, where the linear component has its largest influence, this might notably reduce the delays.
It is common in literature to not match the calculated values to peaks in data but rather compare the delay ratios between effects to those in data. This assumes that all alterations to the usually used quadratic field can be subsumed in a factor. When this is done for the strongest peaks at 6.53 µs and 5.31 µs, an excellent agreement to Hg^+ and Cs^+ is found, which are both expected and frequently reported. If the calculated delays for ixRk effects are scaled with the ratio 0.776 between the computed delay $t_{i,q}(0,0)$ of Hg^+ and its peak center in data, all measured peaks are represented by calculated values from quadratic potentials, besides small remaining differences which might be due to the used approximations, such as assuming quadratic interdynode fields. For ik, i1Rk and i2Rk the corrected values, which were scaled by this ratio, are listed in the table. Interestingly, these numbers are even smaller than from linear fields alone. The origin of this unexplained acceleration would have to be studied with field simulations.

Effect identification

The following discussion is oriented towards the first AP distributions, since less secondary effects complicate the identification. The calculations are for quadratic fields and scaled to Hg^+ for cathode iAPs.

Early afterpulses

At the earliest delays, a few EAPs with regular charges are encountered. Since here a classic analysis was performed, only γnk and BL1 and secondary effects can be

recognized. Because the coincidence only accepted delays above 83 ns, for lower values only a minute fraction of the true events is detected when they occur as random additional events [147]. For these reasons the EAP rate is far lower than in the EAP analysis, but nevertheless the known grouping around 40 ns is recognized with charges up to 9 pe from multiple γnk and multi-pe emission. The cluster around 95 ns with sum charges up to 3 pe is already within the coincidence range and is attributed to second order γnk or BL1L1.

Dynode ionic afterpulses

Interdynode iAPs start around 100 ns, as verified by calculations. The CDD_1 (fig. 2.66c) shows that up to about 160 ns EAPs (regular charges) dominate, whereas beyond ixRn effects (undersized) prevail. According to calculations for quadratic fields, i1 effects are present between 0.17 and 3.2 µs, and i2R1 effects lie between 0.20 and 3.9 µs and are slightly later than i1, which should be visible as sub-peaks. i2 (= i2R2) extends from 0.12 to 1.95 µs, but should in most cases have too small charges to play a large role for the used threshold. Generally, i1 and i2 mimic the series of ik types at lower delays and charges. The mean transit times from linear fields are lower. A linear component will furthermore introduce a smear of the peaks.

In data most low-charge events occur in the first microsecond with charges up to about 1.5 pe and peaks below 1 pe, which matches with i1 and i2R1 effects. In many cases the overlap of several close effects likely causes an effective peak at a value corresponding to no single effect. Using linear fields or transit times obtained from field simulations in the interdynode spaces, and including the ICS energy dependence might lead to exact matches of peaks with single effects.
The earliest iAPs in CDD_1 are likely remnants of H^+2 (calculated center at 115 ns; large ISEY, see H^+2Rk). The small peak at 255 ns probably results from a combination of i1 of $H_2{}^+$ (0.22 µs) plus He^+ (0.31 µs) and i2R1 of H^+ (0.20 µs) and $H_2{}^+$ (0.27 µs). The strong peak at 415 ns likely is from He^+2R1 (0.37 µs) and He^+3R1 (probably higher). The cluster around 620 ns matches i1 of H_2O^+ (0.62 µs) or $CH_4{}^+/O^+$ (0.58 µs), or N^+2R1 (0.66 µs). The distribution which is centered on 850 ns should result from i1 of Ar^+ (0.91 µs) and possibly $O_2{}^+$ (0.81 µs), and i2R1 of H_2O^+ (0.75 µs) or $N_2{}^+/CO^+$ (0.92 µs).
The whole low-charge region between 350 and 750 ns is from overlapping effects with possible contributions from i1 or i2R1 of He^+, N^+, $O^+/CH_4{}^+$, H_2O^+, $CO^+/N_2{}^+$, Ar^+, and $CO_2{}^+$ (see table).
While Cs^+ and Hg^+ produce the largest ik peaks, these ions appear to be far less stronger in i1: In figs. 2.66c and 2.67c the underamplified pulses present at 1.63 µs (Cs^+) and 2.00 µs (Hg^+) are barely noticeable. Possibly, the large mass and corresponding slow velocity of these ions favors an extraction from the dynode chain.

For H^+, $H_2{}^+$ and He^+ the i2R1 populations appear to be larger than i1, whereas for heavier ions i1 effects seem to better agree with observed peaks. This could indicate a high transport factor from v_{23} to d1 for light ions due to their mobility. The dynode layout of the R5912 favors i2R1, which possibly even reduces i2.

[147] In the measurement a triple coincidence of light pulse, PP, and AP with delays between 83 ns and 95 µs was used, which excludes first order EAPs as triggering APs. However, since EAPs occur independently from iAPs, which can initiate a trigger, the PP *can* create a second AP in form of an EAP, which becomes visible in the recorded waveform. Since the iAP presence is required, but that of an EAP is not, the EAP rate is suppressed compared to the iAPs.

The most frequent ion types are difficult to establish. He^+ and H^+ are observed with high certainty, later delays are most likely dominated by N^+, $N_2{}^+$, H_2O^+ and Ar^+.

Cathode ionic afterpulses

ik effects show large charges and the predicted charge decrease with rising atomic number of the ion, which corresponds to growing delays. The calculated quadratic delays are virtually identical for i1Rk and ik, so both will contribute to the same peaks. i2Rk is notably earlier, since the speed gain for v_{k1} can outweigh the extra covered interdynode distance, and can better explain some peaks than ik/i1Rk. This might be due to approximations and can only be clarified with a field simulation.
The earliest line at 405 ns with mean charge $\approx$13.3 pe is attributed to H^+:2Rk (425 ns); the offset likely is from inaccurate transit time estimates between d3 and d1. For ik and i1Rk of the same ion (470/480 ns) notably less events are observed, that is, ik is the sub-peak of i2Rk and not vice versa. This either points to an effective extraction of the light ions from v_{23} or uncertainties in the computation. The mean charge is remarkably high and will lead to many subsequent EAPs, which feed the underamplified region up to ca. 500 ns.
For $H_2{}^+$, on the other hand, very few events – best visible in the FDD – are observed around 695 ns (ik and i1Rk: 0.66/0.67 µs) and 590 ns (i2Rk: 0.59 µs) with mean charges of about 11.5 pe. Both clusters have approximately equal populations.
The high-charge peak at 840 ns (mean charge 11.3 pe) is from He^+2Rk (0.83 µs). ik/i1Rk of this ion (0.93/0.94 µs) are also present as later sub-peaks with slightly lower mean charges, as is expected from the smaller impact energy.
The broad peak around 1.5 µs ($\approx$3 pe) might be from N^+2Rk (1.53 µs). The corresponding ik/i1Rk cluster is found around 1.7 µs (also about 3 pe) with calculated delay 1.73/1.74 µs.
In the range 1.8–3 µs several populations of ik/i1Rk and i2Rk overlap, most likely from N^+, $N_2{}^+$, and Ar^+, and possibly also from $CH_4{}^+$, H_2O^+ and $CO_2{}^+$ (values see table). These distributions show rapidly decreasing mean charges. While some ions still produce 2 pe or more, with increasing delays more and more events are compressed into the 1 pe bar, since production of at least one sk is required for cathode iAP detectability.
The peak around 4.1 µs has a mean charge close to 1 pe, albeit some events with 2 pe are still present. After this cluster no *primary* iAPs with charges commonly above 1 pe are encountered. This peak must be from a frequent ion species due to its large size. Nevertheless no good agreement is found for ik (unless Kr^+ with 4.22 µs is considered), the closest likely i2Rk candidates are Sb^+ (4.46 µs) and Cs^+ (4.66 µs). Since Cs^+ produces the second-strongest peak, Cs^+2Rk seems likely despite the large difference in delay. Maybe the transit times are more sensitive to the exact fields for such large ion masses. Alternatively, if double ionization is considered possible, Hg^{2+}2Rk (4.05 µs) would provide a good match. The corresponding Hg^+k (4.62 µs) might be present on the left flank of Cs^+k (see below). Cs^{2+}k (3.76 µs) could overlap with the 4.1 µs peak, but Cs^{2+}2Rk (3.30 µs) is directly in the valley between peaks and thus unlikely.
The largest peaks around 5.3 µs and 6.5 µs (fitted centers) are from Cs^+k and Hg^+k. An Hg^+2Rk distribution should be located near 5.72 µs, and would strongly overlap with ik of Cs^+ and Hg^+, which might cover its presence.
The last detectable cluster in the first-AP distributions is rather broad and encountered around 11–13 µs. This would match with ik from Cs_3Sb^+, which is the dynode

material of this PMT, and ik of other Cs–Sb compounds like Cs_2Sb+ and $CsSb^+$ could also be present on the flank after Hg^+k. Nevertheless, the presence of these compounds is difficult to establish because of well-founded second order iAPs which occur at the same delays (see below).

For ik the most frequent ion species are, somewhat surprisingly, the heavy ions Hg^+ and Cs+, likely followed by N^+, $N_2{}^+$, and Ar^+. He^+ and H^+ are certainly present, although more seldom. $H_2{}^+$ is observed but very rare. Possibly present species comprise $CH_4{}^+/O^+$, H_2O^+ and $CO_2{}^+$. Regarding molecular forms, cathode iAPs from H^+ in this PMT are far more frequent than from $H_2{}^+$, and more likely for N^+ than for $N_2{}^+$.
For light ions i2Rk tends to dominate over ik (H^+, He^+, about equal rates for $H_2{}^+$, though), for medium mass similar contributions are found (N^+), and for heavy ions (Cs^+, Hg^+) ik is stronger. As for i1 and i2R1, light ions seem to more easily escape the dynode chain.
In [382] a linear-focused PMT showed inRk sub-peaks after ik up to at least d5, where the peaks from later dynodes had larger delays. For the box-and-line R5912 here no series of subsequent peaks is observed for several reasons: i1Rk is congruent with ik according to calculations, i2Rk actually lies *before* both, and ion feedback from v_{34} or later should be strongly suppressed due to the dynode geometry. The different delay behavior observed here is most likely due to the smaller diameter (2”) of the PMTs (RCA 8850 and 8852) studied in [382], since x_{k1} is far smaller and thus the interdynode distances play a larger role for the delays of inRk.

Backscattering

In the region from 1 to 3 µs for large charges a falling charge–delay behavior is observed. This is reminiscent of electronic branching TSEs from backscattering. The later part of this range was interpreted as several overlapping ion types with gradually decreasing mean charge; however, between 1 and 1.4µs no compatible ions have ik/i2Rk delays in this range.
This might be explainable through doubly-charged ions: N^{2+}k (1.23 µs), N^{2+}2Rk (1.40 µs) and theoretically O^{2+} (1.09 µs); the latter would match with a peak observed at 1.09 µs (5.4 pe). However, both double ionization and the availability of free oxygen are unlikely – the former especially since except for the doubtful presence of Hg^{2+} no double ionization peaks were observed.
While cubic field functions would lead to a falling charge–delay behavior (see (2.216)) as found between 1 an 3 µs, a notable cubic component would affect all distributions in the CDD. Such field configurations would smear out the peaks, which is in contradiction to the observed sharp distributions. In addition, a cubic field component would affect the electron transit times, which disagrees with the results of the TSE and EAP analyses.
These considerations suggest to analyze whether backscattering effects provide alternative explanations for the events between 1 and 3 µs.

For iLk in a quadratic potential the extra delay from backscattering was independent of the backscattered energy U_s, see (2.220). Therefore the delay does also not depend on ion origin (varying impact energy) and ion mass (kinematic factor) and always amounts to $3\,t_{iq}(0,0)$. The AP charge rises with scattered energy (unless the SEY(E) peak is passed) and thus increases with x_0 and decreases with ion mass.

Since the delay is unaffected by scattered energy, this can, however, not explain a falling charge–delay behavior for a single ion type.
For a linear potential, the delay scales with $\sqrt{U_s}$ in (2.218) and thus is affected by both x_0 and ion mass. The mean delay is significantly smaller than for the quadratic field. From this a *rising* charge–delay function is expected for a single ion type, which is the opposite of the observed behavior.

Inelastic scattering via iΛLk follows the same extra delay equations with additionally reduced scattering energy, so in quadratic fields no charge–delay dependence results. In a linear potential for a fixed impact energy the delay is proportional to the root of the scattered energy, which amplifies the existing rising charge–delay behavior.
Branching in form of iBLk creates a first AP at the same time as the corresponding ik type but with smaller charge (at least 1 pe for detection), and a second AP with properties as for iΛLk, which would only appear in the all-AP distributions.

Ions from $v_{n(n+1)}$ hitting k (for example i1Lk) have extra energy compared to v_{k1}, which affects charge and possibly delay. If $U_s > U_{k1}$, d1 can be struck (e.g. i1LkR1) and the delay depends on U_s even for quadratic potentials according to (2.221). However, this produces an undersized AP and is very improbable – in the majority of cases the ion is lost.

Dynode scattering before a cathode impact through ixEnRk (e.g. i1E1Rk or i2E2Rk) will introduce delays compared to ixRk due to the scattering loss and the extra path. But the charge is still higher than ik, since the impact energy is above U_{k1}.

For iL1 and iL2 undersized pulses with larger delays than i1 and i2 will result. The behavior is the same as for as for iLk.

In data, no significant number of iLk events is found at the expected delays for a quadratic potential, namely three times the delay of the ik lines of H^+ (1.22 µs) and He^+ (2.52 µs). This might, however, be due to the linear component, which can severely reduce the delay from the loop.
A high-charge line is clearly visible at 1.595 µs in CDD_1 and as a small, focused peak on top of the broad 1.5 µs peak in the DD_1. This feature could result from He^+2Lk. The mean charge of the distribution is difficult to assess due to the overlap with other effects at lower charges, but appears to be lower than that of the He^+2Rk line (about 11.25 pe). For 180° scattering a reduction by a factor of 0.655 to 7.4 pe is expected from (2.208). The delay would also agree with second-order APs of type He^+2Rk-He^+2Rk, but these should not be visible in the first-AP distributions, since the initial AP has a very large charge and is certain to be detected.
If the events between 1 and 1.4 µs originate from backscattering, only the ions H^+, $H_2{}^+$ and He^+ are possible candidates since their ik peaks occur earlier. The peak around 1090 ns in the DD might be caused by He^+2ΛLk and some H^+2Lk visible at high charges in the CDD_1. The latter, however, would result in a different delay factor to ik (2.7) than found for He^+2Lk (1.9). He^+2ΛLk should show a rising charge–delay distribution extending up to He^+2Lk. While there are events which could be interpreted this way, this does not explain a focusing of the distribution into low delays.

The events between 1 and 1.4 µs are thus not yet completely understood. Double ionization events might be the best explanation, possibly combined with several backscattering effects.
While the peak at 1.6 µs likely is caused by backscattering, all in all and unlike for electrons, ion scattering seems to play no large role in the iAP distributions.

Secondary afterpulses

When comparing DD_{1u} and DD_a many features are nearly identical, namely the H^+2Rk spike, the 4.1 µs peak, Cs^+k and Hg^+k. No negative correlations (mutually exclusive effects, see p. 357) are observed except for a slight excess in some bins before 400 ns. On the other hand, there are a few regions, where pronounced positive correlations occur due to secondary EAPs and iAPs.
The excess in DD_a following the H^+2Rk peak and between 800 and 3500 ns is too high to be from statistic covering (shadowing) alone, since the sum rate up to these delays is far too low. The CDD shows that the discrepancy is from small charges, and at large charges a near-complete match is observed (see also FDD). This is caused by EAPs produced by the high sk number. When comparing CDD_1 and CDD_a, each high-charge iAP distribution is correlated with subsequent low-charge EAPs. This is especially clear for the sharp iAP lines at 405, 840 and 1595 ns, which are followed by an EAP peak with the expected delay scale. The EAP production decreases with mean ik charge and is notably reduced after about 4 µs, when only 1 pe iAP pulses occur anymore.
The prominent high-charge cloud around 7.2 µs in the CDD can be traced to second-order high-charge iAPs between 0.4 and 1.7 µs from H^+2Rk, He^+2Rk, He^+2Lk and other effects, which are created by primary iAPs of the 5.3 and 6.5 µs peak. Since the SEY generally falls with rising atomic number of the ion and after 4 µs all species only produce 1 sk, a primary effect can be excluded. The appearance of a small fraction of these secondary iAPs in the CDD_1 could be explainable by unrecognized primary iAPs with low charges.
Similarly, the smaller high-charge distribution around 5660 ns is likely caused by secondary iAPs of the same types following the 4.1 µs peak. There is a fluent transition to the 7.2 µs cloud.
Between 8 and 10 µs, an excess in 1 pe events and some 2–4 pe pulses is encountered (see DD and CDD). The former is caused by primary iAP in the 1.5 µs cluster followed by a secondary iAP of the 5.3 or 6.5 µs types (the opposite of the 7.2 µs events), the latter is from secondary medium-charge iAP types from between 1.5 and 4 µs occurring after Cs^+k or Hg^+k.
Finally, instead of Cs–Sb compound ions, the 1 pe events between 10 and 15 µs could be caused by secondary iAP of the large peaks at 4.1, 5.3 and 6.5 µs following a primary iAP of the same clusters. In the same manner, the bump in the falling flank near 9.5 µs could be from 4.1 + 5.3 µs or vice versa. The presence of several such events in the CDD_1 again could be caused by missed detection.

Comparison with literature

In [276], 255 devices of the HQE variant of the same series as studied here (Hamamatsu R5912) were characterized including a detailed iAP analysis. The findings are in excellent agreement with the ones presented here. Fig. 14 in the publication shows a CDD, fig. 16 the charge distribution for notable event groups. Between 100 and

200 ns APs with 1 pe dominate (i.e. EAPs). Undersized iAPs are found mostly from 0.2 to 2 µs, some down to almost 100 ns, with a peak near 500–600 ns and a mean charge of 0.6 pe in the range of 200 to 800 ns. The authors also interpret these events as positive ions striking d1 with one skipped amplification stage (i.e. i1 or inR1). A small population between 420 and 450 ns with charges above 10 pe is found in most PMTs and stated to be no artifact. Between 800 and 3000 ns, multi-pe iAP are found with charges up to 30 pe, the falling charge–delay behavior from 1 to 3 µs is also observed. From 3 to 10 µs spe pulses are encountered, with the exception of a high-charge spike around 7–8 µs with charges up to 30 pe.
Other publications which confirm the results shown here are [289, 309, 352, 389, 390]. In some cases the relative strength of peaks is different [289, 389], but this intra-series variation is well-known (e.g. [382]).

Dependences

Besides the dependences, which were already discussed, iAP observation depends on applied voltage, vacuum and diffusion of He, and – interestingly – the wavelength of the photons inducing the PPs.

As expected, the iAP mean charge rises with U_{k1} [180], which is from the cathode iAP fraction. In [170; pp. 447–448], the iAP rate is plotted against total voltage for the ETE[148] 9107B (30 mm). First a strong rise with HV is observed, since only multi-pe and high-delay spe APs are detected, before a plateau attributed to ixR1 is encountered. This is followed by another increase and a plateau suggested to stem from ixR2. After this, the rate rises again and the ixR3 plateau is still vaguely perceptible. Wright predicts a rate increase of $\delta 1 = 15$ and $\delta 1 \delta 2 = 75$ between detection of only spe pulses and the inclusion of ixR1 resp. ixR2, when these events rise above the threshold with growing high voltage (HV). The ratios simply correspond to the number of electrons in the cascade, and the predicted values agree well with data. [170] recommends to operate just below the knee of the first plateau.
On p. 544 the influence of HV is analyzed for the iAP DD, CDD, and FDD as well as the rate, probability, mean charge and mean height based on measurements performed for this work.

In order to reduce the iAP rate, the PMT vacuum is maintained at a value of 10^{-6} to 10^{-5} Pa [169; p. 72] by means of internal getters [259]. The normal He content in the atmosphere is only 5.2 ppm, but suffices to cause a noticeable increase in AP rate over several years [180] as He diffuses through the glass into the PMT vacuum (see [170; pp. 448–449] for theoretical treatment) and the iAP peaks from He, which are found in nearly all PMTs, get stronger.

According to [170; p. 446], a wavelength dependence of the iAP rate for PMTs of various manufacturers was observed in a single publication [391]. For blue light of 480 nm over 10 times more APs were detected than for 585 nm, and these were present in form of a quasi-exponential decay of ca. 10 µs length on top of the (unaltered) peaks. This was attributed to glass or cathode excitations which relax radiatively, whereas Wright surmises that it might be from photoeffect on d1 ($\gamma 1$) leading to i1. The peak wavelength of Cs_3Sb lies at 280 nm (see p. 55) and the QE drops

[148] Electron Tubes Enterprises Ltd.

rapidly in the visible range [169; p. 34], which is in agreement with Wright's theory; however, this does not explain the absence of the strong long-term exponential decay in all other measurements, including the ones performed here. It could be that a better crystallinity of the modern high-gain first dynodes reduces the fraction of phosphorescent radiative relaxation.

Reduction of ionic afterpulse rates

The inclusion of getters can drastically reduce the iAP rate, in case of the 2" ETE 9813 to 0.03% pe [180]. In [177] through additional getters and redesigning of the space between d1 and deck, for a 1.5" Hamamatsu prototype the rate could be reduced to about 0.004% pe for charges above 4 pe. However, the total iAP rate for a 1 pe threshold was as high as 10% pe, which indicates a strong ion production between dynodes leading to a large number of dynode iAP, but an effective inhibition of ion transport from v_{12} to k.

2.5.2.4 Late afterpulses

Late afterpulses (LAPs) are afterpulses occurring with delays beyond that of ionic afterpulses, which therefore must have other physical sources. This type of events was reported several times with delays ranging between 20 µs and several 100 µs and very low rates [170, 180, 232, 321, 392–396] (see also 2.6.1). Despite the scarcity of these events, they can nevertheless be problematic for low event rates in combination with long acquisition windows, as e.g. was the case in the Borexino detector (see below and fig. 2.68a) or possibly for LIDAR[149] [397].

In 2010, Quirin Meindl observed a peak around 100 µs delay after muon events in Borexino data (fig. 2.68a), which could not be explained by physics events and was attributed by the present author to LAPs of the ET 9351 PMTs [396] [154; p. 88]. The region between 30 and 650 µs delay could best be described by a model of a short exponential decay with $\tau = 9.2$ µs, a Gaussian with mean 98 µs and $\sigma = 42$ µs, a long exponential decay with $\tau = 116$ µs and a constant background from dark noise.
The presence of LAPs was confirmed by [321] for the predecessor model ET 9350 (called long APs in this publication). The authors studied delays up to 1 ms with $o = 500$ and observed a single LAP peak for the 9350 (see fig. 2.68b) and an 8" Photonis PMT; no other features were found. For the 6" FEU-49B no LAPs were encountered. In the 9350 the LAP peak center was near 110 µs with a mean delay of 120 µs. LAP delays ranged between 70 and 200 µs for the 9350 and the Photonis PMT, the charge was 1 pe and the AP rate was $\ll 0.1\%$ pe. For a smaller occupancy of 20 the LAP peak was almost invisible against dark noise for the 9350.
In the measurements carried out here, no LAPs could be observed unambiguously for the Hamamatsu R5912 and R11780 (see figs. 2.68c and 3.26; detailed discussion on p. 553). For the former this is in agreement with the pulse shape of the DEAP-3600 detector shown in fig. 2 of [398], which operates with 255 HQE R5912 PMTs and shows no sign of LAPs up to 160 µs. The absence in these measurements, however, might be caused by the low occupancy compared to [321] (here 0.035, in [398] up to 3.5), due to which the LAPs could be covered in dark noise.

[149] Light detection and ranging

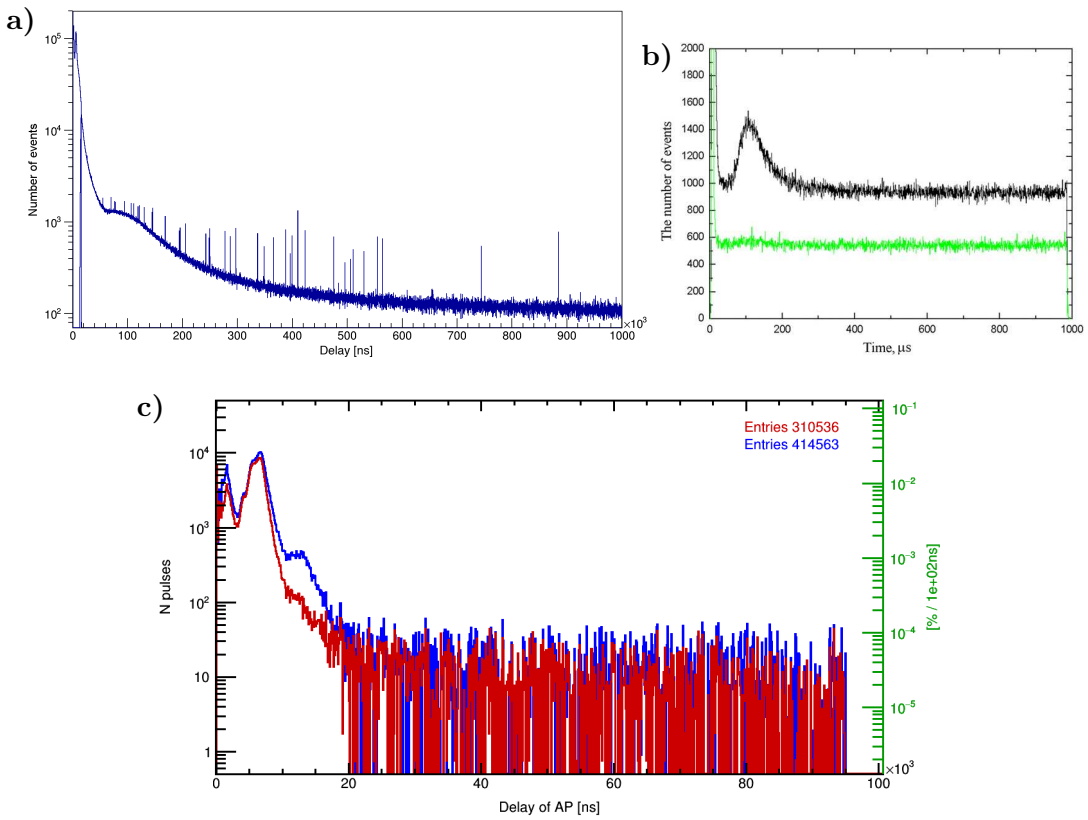


Figure 2.68: Late afterpulses (LAPs):
a) ET 9351 (8") signals in the Borexino detector after muon transit. The broad peak around 98 µs is from LAPs, the spikes are from neutron captures, and below 16 µs a different acquisition electronics is used. The distribution is discussed in [154; ch. 5]. Adapted from [396], courtesy of Quirin Meindl.
b) Thorn EMI 9350 (8"), black $o \approx 500$, green 20; from [321]. This PMT is from the same manufacturer as the ET 9351 and very similar. A LAP peak is visible around 110 µs.
c) Absence of LAPs in Hamamatsu R5912 (8") for small occupancy: DD_a (blue) and DD_{1u} (red) within 95 µs delay, DCR subtracted, $o = 3.5\%$, same measurement as in fig. 2.66. No events different from dark counts are visible in the CDDs and FDDs after 20 µs.

Possible causes

[395] proposed glass fluorescence through cathodoluminescence from ions produced near the anode, which escape the dynode chain and hit the glass bulb, as explanation for the emergence of LAPs. However, this only leads to an exponential decay extending up to 40 µs and no peaks are predicted [321].
Another possible model explaining APs up to 20 µs mentioned in [321] is through exoelectronic emission combined with ionic feedback. This is the case when a defect is formed with available energy, which traps an electron. When the defect disappears later under certain conditions, it releases the energy to the trapped electron, which is sufficient to emit it into vacuum [372]. This release can occur for example through thermostimulated exoelectronic emission.
Since both models do not come close to the observed delays, the authors of [321] have no explanation for these events.
In the following some alternate models will be examined.

One possibility would be large transit times from TSEs for heavy ions.
Secondary APs cannot explain the high LAP delays, since for ik no heavier ion than Cs_3Sb^+ is expected, whose second-order APs reach merely 21 µs and 32 µs for third-order.
iLk introduces delays of 3 times that of ik for a quadratic potential and for Cs_3Sb reaches only 32 µs. iaLk likely is shorter than iLk due to the large initial energy when crossing the k–d1 region and furthermore is extremely improbable to occur.
iΛLk has delays identical to iLk or shorter, not longer as for EAP effects, and is no explanation either.
iLkLk can extend up to 5 times the ik delay and 53 µs, which is still not enough, while it is uncertain whether this is observable at all. In addition, then the iLk peak should be observable with far higher size.
External ion paths from the last dynode or the anode to the cathode are also excluded as possibility, as these ions would be drawn onto the glass envelope as [395] already pointed out.
Therefore ion drift times, ion time shift effects and secondary known APs can be excluded as possible origins.

In [180] two possible AP effects are described, which occur after scintillation, i.e. after the passage of charged particles through the PMT: glass fluorescence (see 2.6.1) or excitation of metastable levels in the cathode. The mean number of afterpulses per event was found to depend on the cathode and window materials.
In Borexino the passage of a muon could in fact lead to excitation of the glass, whose fluorescence could trigger the same or other PMTs, and muon transits are also expected in the measurements of [321]; different PMT materials could explain the absence in the R5912. However, for fluorescence an exponential decay and not a peak is expected, unless additional intermediate levels play a role.
[232] also mentions that the electron impact on dynodes can emit light with delays of up to a few 100 µs, which subsequently hits the cathode, but this appears to occur only in PMTs with specific dynode compositions and for high gains – the RCA 8850 (2") in the quoted paper [394] had a GaP first dynode, and a peak near 40 µs. It might be that such dynode scintillation occurs in BeCu as well (possibly used in ET 9350, 9351), but not in Cs_3Sb (Hamamatsu R5912).

Since the decrease of dark count rate after illumination was found to have very long time scales (see p. 348), the same processes might be responsible for LAPs.
[170; pp. 449–453] states that after exposure to bright light a tail of events with initially 0.1% the count rate, which extends to over 100 µs after the stop of illumination, was observed for S20 cathodes by [392, 393]. The decay was found to vary with t^{-1}.
No explanatory approaches were listed, but fluorescence of a material such as the glass window is a natural candidate, seeing as this has been reported previously (see 2.6.1). As discussed on p. 348 the long term dark count response after exposure to light has time scales measured in seconds or hours, which is already too long to yield explanations for LAPs. Possible mechanisms are metastable excitations of the cathode [180] or persistent photoeffect, for which in semiconductors charge carriers can remain up to days after illumination [371, 399]. These processes could also play a role on a shorter µs time scale.

Of the listed processes glass fluorescence, excitation of metastable levels in the cathode and possibly persistent photoeffect are the only processes with theoretical

time scales long enough to be viable candidates for LAP production. Interestingly, glass fluorescence and metastable cathode levels are listed as possible origins of long term dark noise as well as APs following scintillation in literature. Phosphorescence, however, cannot explain a peak in the delay distribution without additional transitional levels in the deexcitation process. For the persistent photoeffect it is unclear whether this can occur in photocathode materials. This makes metastable cathode levels the most likely explanation for the LAP peaks observed so far.

2.6 Exterior background

2.6.1 Light emission

Counterintuitively, photomultipliers (PMTs) not only detect photons, but also *emit* light through various processes. This was first reported by Krall [332] and verified multiple times by different authors since then (see below).

Relevance

Light feedback to a photosensitive structure (photocathode, dynodes, aluminized glass, grid, ...; see 2.3.5) within the same device can cause photonic afterpulses in the early afterpulse (EAP, up to 100 ns) or even late afterpulse (LAP, $\mathcal{O}(100\,\mu s)$) delay range [170, 180, 332]; see 2.5.2. This is detrimental for photon counting applications and is seen as one of the principal causes for high-gain breakdown limiting the operational regime. [332] notes that for linear-focused designs internal paths can be particularly transparent (fig. 2.43c defines external and internal paths).
In addition, presence of a notable amount of photonic afterpulses (γAPs) will influence the average pulse shape (APS, see p. 343).
Furthermore, light emitted out of the PMT can trigger neighboring photosensors in a detector (crosstalk) [180] [170; p. 258].

Countermeasures

The largest amount of light is produced in the last stages through the sub-effects subsumed as electrode glow, and already in 1967 Krall remarks that all then-modern PMTs incorporate opaque dynode mounts to reduce this effect [332]. Most modern PMTs have an additional U-shaped shield placed around the last stages, which is connected to the potential of a middle dynode and serves as both light shield and ion trap (ion deexcitation is one source of dynode glow) [170; pp. 265, 266, 453].
PMTs designed for photon counting include further shuttering measures such as an extension of the deck to the glass envelope to stop light feedback via external paths. Nevertheless, for the Hamamatsu [150] R5912 (8”), which employs all these measures, in this work a photonic afterpulse rate of still no less than 13.9% was found.
Newer developments, based on a collaboration with the Cherenkov Telescope Array (CTA) project during this work, lead to the drastic reduction of external emission of dynode glow through use of extended light shields which cover all dynodes [181, 182] (see p. 429).

Recommendations

Since even for modern PMTs the emission of light out of the device through external or internal paths can be a factor, for test stands characterizing several PMTs at the same time it is important to optically isolate the PMTs from each other. This also eliminates effects from reflections off the PMT glass or cathode. For experiments the same applies where possible; placing light concentrators in front of the cathode takes care of this requirement in a beneficial way.

[150] Hamamatsu Photonics K. K.

Mechanisms

The following effects were verified or postulated to cause light emission in PMTs:

- **Photoluminescence (PL):** Photon emission after excitation through *photons* in solids or residual gas atoms or molecules with the subtypes fluorescence (relaxation from singlet to ground state) and phosphorescence (triplet to ground state, in semiconductors for excitons). This can occur in the glass, cathode, or dynode mount for either external or internal photons.
- **Cathodoluminescence (CL):** Radiative transition in a solid after excitation by electron impact. Can occur for all materials struck by electrons, most frequently for the dynodes and dynode mount.
- **Excitation and ionization of residual gas:** Light emission from deexcitation or recombination of atoms or molecules excited by electron impact. In inter-dynode space, occurrence increasing with the number of electrons in the cascade.
- **Scintillation (radioluminescence):** Radiative deexcitation in a (transparent) solid following excitation from passage of a charged particle or high-energetic photon. In the PMT glass.
- **Bremsstrahlung:** Radiative deceleration of charged particles in an electric field, more specifically of electrons on nuclei in a solid. Occurs on all structures hit by electrons.
- **Cherenkov radiation:** Emission of photons when a relativistic charged particle passes through a transparent medium with speed above the phase velocity c/n. Occurs in the PMT glass due to cosmic radiation or environmental radioactivity.
- **Transition radiation:** Emitted when a charged particle passes a change in refractive index, e.g. between materials. Postulated by [321] to occur in PMTs. While there is no lower energy threshold, at typical electron energies in a PMT the intensity of this effect should be negligible compared to the others effects.
- **Electroluminescence:** Light emission due to a current or electric field. Can play a role in the glass for negative cathode voltage.
- **Electric discharge:** For high field gradients, ionization of residual gas or field emission of electrons into vacuum can take place. The latter can create ions in the gas or hit a structure. Photons are created through deexcitation, recombination or cathodoluminescence. The subtypes depend on field gradient and gas pressure, and for PMTs corona discharge, glow discharge, arc discharge, and vacuum arcs are expected in the PMT vacuum or its base. These effects cause the dreaded “flashers”.

There is a strong connection between some of the above effects.
Photoluminescence, cathodoluminescence and scintillation differ just by particle type and energy causing the excitation [400; p. 683]; cathodoluminescence can be considered a subtype of scintillation. Therefore all three mechanisms use the same

radiative transitions after vibrational relaxation, namely fluorescence and phosphorescence, with similar time scales.
Excitation of free gas constituents is similar to cathodoluminescence for the corresponding solid material; the energy levels are similar except for the forming of bands.
Finally, electric discharge processes emit light through one of the other mechanisms.

The relevant energy scale varies between effects. For cathodoluminescence, photoluminescence and gas excitation it lies in the eV range. For gas ionization and bremsstrahlung medium energies starting at a few tens of eV are needed. And scintillation, Cherenkov radiation and transition radiation require higher energies than the PMT itself can produce and thus must be caused externally.
Electroluminescence [151] and electric discharge, on the other hand, are static effects and independent of particle interactions.

Already [401] noted that the origins of light emission can be disentangled by taking into account their emission location, the dependence on the current inside the PMT and the temporal correlation to a cascade.

Most of the above-mentioned mechanisms were verified in one setting or another; which of them are most critical depends on the use case. The great efforts undertaken by manufacturers to improve the vacuum and shutter the latter stages point to the effects causing dynode glow – that is bremsstrahlung, cathodoluminescence, gas excitation and ionization, and photoluminescence – being most crucial in the majority of applications. Scintillation and Cherenkov radiation contribute to the high-charge tail [152] of the charge distribution, but are normally too seldom to cause severe problems. Electric discharge occurs at high gains (i.e. voltages) from sharp corners inside the PMT and raises the dark count rate (DCR). But discharges can also take place for problems in the design of the PMT base or encapsulation, in which case they can turn the PMT into a flasher and potentially endanger operation of the whole detector. Electroluminescence should only matter for negative high voltage (HV) [332].

Regarding the time scale, photons are emitted instantaneously for bremsstrahlung, Cherenkov radiation, and transition radiation. Cathodoluminescence, photoluminescence, scintillation, gas excitation/ionization radiate through fluorescence or phosphorescence with delays ranging from nanoseconds up to a few 100 µs (see also 2.5.2.4) depending on the material. Electric discharges and electroluminescence are typically uncorrelated to PMT pulses and constitute random background.
When one of these effects is initiated by an incident photon, a late primary pulse (PP) is created. When light emission results from a cascade, an afterpulse (AP) can be produced. To obtain the relative time or delay of the resulting pulse(s) then also the cascade transit time has to be considered.

In the literature often terms are mixed up and e.g. scintillation is written when cathodoluminescence or electroluminescence is meant. Frequently it also is not clear whether phosphorescence is implied when the label fluorescence is used.

[151] While the radiative recombination processes are the same as in photoluminescence, the excitation occurs due to electric fields and is likely uncorrelated to PMT pulses.
[152] If the PMT glass tube (in contrast to the window in front of the cathode) is hit, photons can travel through total internal reflection to the window and cathode, but in this case the observed charge is smaller and in the 1 pe (photoelectron) range [170; p. 266].

In the following, the individual mechanisms of light emission are addressed in depth. After this, an investigation concerning the sources and properties of light emission from PMTs is presented, which was carried out in the course of this work and allows to verify theoretical expectations and implement improvements in PMT design.

Photoluminescence

Photoluminescence in PMTs was reported by [169, 170, 172, 232, 332, 391].
Fluorescence and phosphorescence will follow a characteristic exponential decay.

Glass

One source of fluorescence and phosphorescence in PMTs is from the glass envelope, especially the window region, since this directly illuminates the photocathode (cf. 2.5.2.4). All window materials except quartz fluoresce, with the intensity depending on material [170; pp. 450–451] (see p. 348). If a previously illuminated window is placed in front of a PMT, the decay of the detected rate from phosphorescence is far slower than the DCR decrease after direct exposure of the turned off PMT to light (t^{-n} with $n = 0.5$ instead of ≈ 1). This indicates that the phosphorescence component has a longer lifetime than metastable cathode levels, whose decay emits photoelectrons. [332] observed a far stronger increase of the count rate from directly exposing a PMT to UV radiation than indirectly through UV-irradiated Pyrex glass. Krall concludes from this that the dominant UV excitation mechanism for PMTs is from metastable levels in the cathode. For direct exposure [332] observed increased rates up to a day after illumination. Hamamatsu also reports photoluminescence of the glass window following UV exposure [169; p. 267]. The reports of scintillation in the glass envelope also indicate the presence of photoluminescence (see p. 414).

Cathode

In [232; p. 185] Candy interprets the LAP reports of [392, 393] as "cathode fluorescence"; however, glass phosphorescence seems more likely given the frequent reports of the latter but not of the former.

Dynode mount

In addition, light can be emitted from the dynode mount. Hamamatsu lists cathodoluminescence of the electrode support materials as a contribution to dark current [169; p. 72], which indicates the occurrence of photoluminescence as well. For Hamamatsu PMTs whose dynode mounts contain ruby, strong photoluminescence and cathodoluminescence were observed in the present work (see below and 2.6.1).

Dynodes

Finally, since CL is observed for the dynodes (see cathodoluminescence), PL should also occur.

Cathodoluminescence

This process can occur in a semiconductor or isolator when an incident electron excites an electron from the valence band (VB) into the conduction band (CB) from

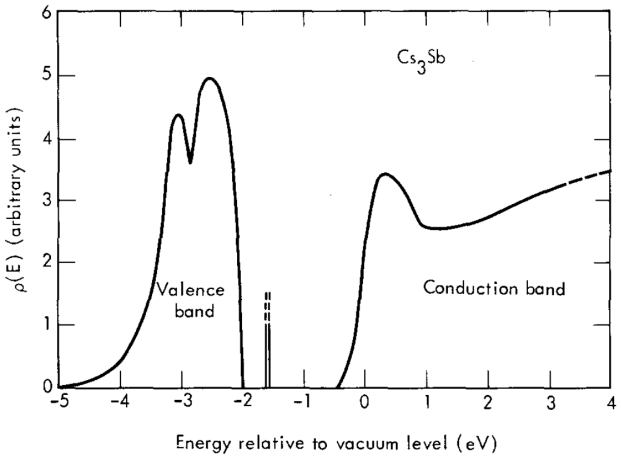


Figure 2.69: Density of states for Cs_3Sb modeled to measurements by [214]. 0 eV = vacuum level. There are vacant defect levels 0.4 eV above the valence band maximum with 1% of the states of the conduction band within −0.45 to 2 eV.

where it can recombine radiatively with a hole. A multitude of transition types are possible, which involve excitons, phonons, impurities, defects, hot electrons with energies above the bottom of the conduction band and intraband transitions [400; p. 684].

In PMTs cathodoluminescence can occur for all stages which are struck by electrons with energies of more than a few eV, i.e. the dynodes, the dynode mount (depending on material), the grid, the light shield (ironically) and the glass (see p. 414).

Cs_3Sb

The dynode material Cs_3Sb is a semiconductor with band gap 1.6 eV [402], so photon emission is expected for energies above this value or slightly below due to the formation of excitons.
The density of states (DOS) measured by [214] [153] (fig. 2.69) shows peaks in the VB at −2.6 eV and −3.1 eV relative to the vacuum level, defect levels at −1.6 eV, and a peak in the CB at 0.3 eV. The VB extends up to −2 eV, the CB down to ca. −0.45 eV (electron affinity E_A, see fig. 2.3), which results in a band gap of 1.55 eV (the newer value of 1.6 eV from [402] is used here).
Since carrier thermalization of electrons and holes is in general much faster than recombination, luminescence will be mostly produced by thermalized carriers with photon energies close to the direct band gap of 1.6 eV, corresponding to a wavelength of about 775 nm. The transition from the VB border to the DOS maxima in the CB results in photon energies of 2.15 eV and 2.65 eV, respectively 577 and 468 nm. The hot carrier transitions from the CB peak to the VB border and maxima correspond to 2.3 eV (539 nm), 2.9 eV (428 nm) and 3.4 eV (365 nm). Relaxation to the defect levels in the band gap results in 1.15 eV (1078 nm) from the CB border and 1.9 eV

[153] In [402] newer calculations of the Cs_3Sb band structure were carried out; however, the calculated band gap was 1.41 eV, which is far lower than the measured 1.6 eV, so the older, measurement-based data of [214] were used here.

(653 nm) from the CB peak. These values will be compared in 2.6.1 to the measured spectral emission from the last dynodes of a 1.5" PMT.

BeO

For BeCu dynodes, the surface is composed of BeO. This substance has a large band gap of 10.6 eV [403] corresponding to 117 nm, which is too low to play a role here. Nevertheless, [404] observed an emission band from above 400 nm down to about 200 nm from scintillation (alpha particles, X-rays), CL and PL (UV light) through fluorescence and phosphorescence, which had the same decay times in the visible and UV region. Alpha irradiation was found to cause primarily fluorescence, whereas for electron impact phosphorescence decay times of up to 400 minutes were observed. Heating the sample increased emission strength. Furthermore, stimulated emission of energy stored in the material by X-ray irradiation was possible. The stored energy showed a slow decay, and even after 30 days a large fraction was still accessible this way. The energy storage of BeO is used for dosimetry via optically stimulated luminescence [405]. [406] found emission from self-trapped excitons at 4.9 eV (253 nm) and F-type color centers at 3.4eV (365 nm). For ceramic BeO also emission bands around 3.9eV (318 nm) and 1.85eV (670 nm) from lattice defects were detected.

GaP

[232] attributes the LAPs observed up to a few 100 µs for high gains in [394] to photons from cathodoluminescence in the GaP dynodes which hit the cathode. The present author assumes this could involve the formation of triplet state excitons to prevent relaxation by other means.

Glass and dynode mount

Another source of cathodoluminescence are cascade electrons (pk or sn, see 2.3.5.1) which hit the dynode mount or the glass envelope [401] [169; p. 72]. The light intensity is proportional to the current inside the PMT (i.e. the illumination rate, the occupancy and gain). Operation with negative voltage and an external metal housing at ground potential near the cathode aggravates this effect, but this can be ameliorated by coating the glass envelope with conductive paint connected to the cathode potential (and covering it with an insulator for safety).
For some Hamamatsu PMTs the dynode mount is made from corundum with a small admixture of Cr, i.e. ruby [177]. In the present work this was found to cause a tremendous amount of emission around the ruby line at 694.3 nm due to cathodoluminescence and photoluminescence (see 2.6.1). Since the decay time for ruby lies of the order of 3 ms [407, 408], it is typically not possible to correlate an AP to the PP, because dark count pulses occur in between. Albeit this luminescence causes correlated APs, effectively it raises the DCR.

Excitation and ionization of residual gas

Photon emission after excitation/ionization likely is strongest near the anode where electron density is highest [180, 401]. According to Coates the intensity is proportional to the current inside the PMT and the gas density [401]. Wright postulates that the excitation and ionization occurs for atoms and molecules adsorbed on the surfaces, which are detached through electron impact [170; pp. 265–266]. The cathode can be

hit via external paths along the envelope, which result in photonic EAPs correlated in time to the cascade (also noted by [321]). The first evidence for recombination radiation of ions formed by electron impact was reported by [332], although only with low levels.

Scintillation

Glass

In literature scintillation of the PMT glass has been repeatedly reported [180,332,374, 401] [169; p. 265] [409]. This can be caused by beta and alpha particles and gamma rays from the PMT materials (mostly the glass) or the surroundings, and cosmic rays. Gammas and cosmic radiation can also pass the glass envelope and knock out electrons from cathode or dynodes [169; p. 265] [172; p. 4-53]. Of these interactions, glass scintillation is considered to be the dominant factor and depends on the material, as observed for glass photoluminescence (see p. 411). Afterpulses following a charged particle impact were attributed to occur from glass photoluminescence of excitation of metastable cathode levels by [180], with the number of APs depending on the material of window and cathode. [172] notes that scintillation following alphas and betas mostly produced single-photon (spe) pulses, since the scintillation efficiency is only about 10 photons/MeV and not all photons reach the cathode.
[169] reports that scintillation in the glass leads to fluorescence and phosphorescence, with an increase in dark current up to 40–60 minutes after exposure to gamma rays. [374,401] observed decay lifetimes of 20–50 µs following cosmic rays and gamma rays. Most events occurred within 100 µs and their number was proportional to the charge of the prompt Cherenkov pulse. [409] found that about half of the DCR of the Hamamatsu R12860 is caused by a continuous pulse with a burst structure. The source was identified as scintillation in the PMT glass caused by its intrinsic radioactivity through placing a glass sample which was irradiated by an alpha or beta source between two PMTs. The time distribution showed two exponential decays with decay times of about 2.5 µs and 7.8 µs.

Dynodes and dynode mount

As for PL and CL, scintillation should also occur for the dynodes and dynode mount, but this requires a gamma or cosmic ray and will only occur with small rates.

Ion-induced scintillation

[395] proposed a different mechanism, in which ions generated in the vacuum between the last stages drift onto the glass envelope where they produce scintillation light. The same should also occur for ion impact on the dynodes or cathode.

The acceleration of ionized air molecules onto a PMT with negative HV cannot produce scintillation, since the mean free path of ions in air is in the nanometer range at normal conditions and the accumulated energies are not large enough to either cause luminescence of air constituents or scintillation of the PMT glass. However, these ions will become adsorbed to the glass window where they can recombine with electrons provided by the cathode current, and this return to the ground state can emit light which is caught by the cathode. This mechanism might be one reason for the higher DCR of PMTs operated with negative voltage.

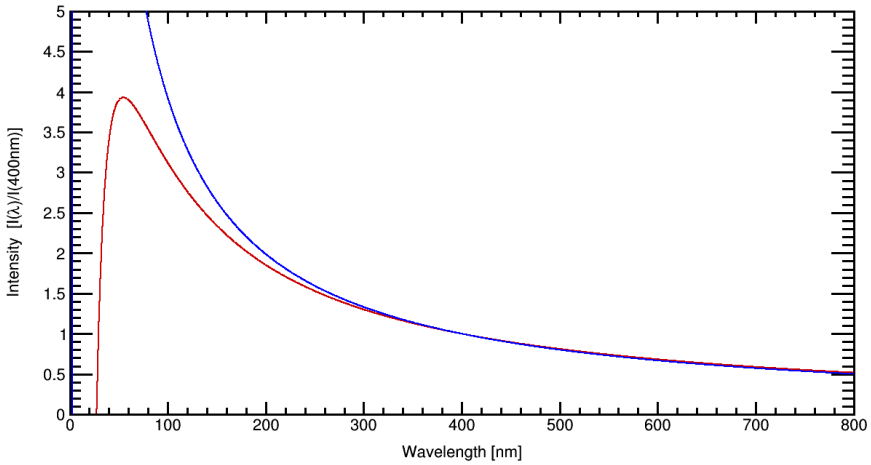


Figure 2.70: Bremsstrahlung spectral intensity for impact on thick target and typical electron energies in the Hamamatsu R5912 (8"): 514 eV (blue, U_{k1}) and 45.4 eV (red, U_{56} to U_{10a}). Graphs normalized to the values at 400 nm. Discussion see text.

Bremsstrahlung

For electron impact on a target, a continuous bremsstrahlung spectrum superimposed by characteristic X-ray peaks is produced (see also pp. 75 and 324).
According to [410; p. 94] for a thick target the minimum observable wavelength is

$$\lambda_{\min} = \frac{h\,c}{e\,U} = \frac{1.240 \cdot 10^{-6}\,\mathrm{V\,m}}{U} \tag{2.222}$$

where U is the acceleration voltage of the electron. The spectral intensity follows [410; p. 94]

$$I(\lambda)\,d\lambda = \frac{KiZ}{\lambda^2}\left(\frac{\lambda}{\lambda_{\min}} - 1\right) = \frac{KiZ}{\lambda^2}\left(\frac{\lambda\,U}{1.240 \cdot 10^{-6}\,\mathrm{V\,m}} - 1\right) \tag{2.223}$$

with K the Kramer constant, i the impact current, and Z the atomic number of the target material. The number of emitted photons is proportional to the electron energy E, whereas the intensity is proportional to E^2 (p. 75).
For electron energies present in the PMT, the spectrum shows a peak in the far UV between about 2.5 nm (1000 eV, first dynodes after forward scattering) and 62 nm (40 eV, last stages). The intensity falls with wavelength and emission occurs mostly in the UV and blue. This is shown for the voltage range present in the Hamamatsu R5912 in fig. 2.70. For 514 eV the peak lies at 4.8 nm, for 45.4 eV at 54.6 nm. For wavelengths below 300 nm increasing differences show, but above this wavelength the spectral shape is nearly identical. A bialkali cathode has notable sensitivity between about 115 and 600nm (see fig. 2.5b) to the bremsstrahlung produced in its PMT – here the UV cutoff from the window material does not apply, since the light hits the cathode from the *inside*. This is different for adjacent sensors, so the absorbance in the glass window of the emitting PMT reduces crosstalk from light emission.

Cherenkov radiation

Energetic radiation such as gamma rays, electrons, muons or other charged particles crossing the glass envelope can create a large number of photons through Cherenkov radiation [170; pp. 265–266, 276–279, 603–609] [172; p. 4-53] [332] [169; p. 72].

Gamma rays produce light indirectly through the creation of electrons, which can emit Cherenkov radiation if their energy is high enough. The amplitude of the light from Cherenkov radiation is proportional to the glass thickness and amounts to ≈ 500 photons/cm between 300 and 600 nm [172]. Therefore input windows should be as thin as possible. Due to the limited UV transmission, glass transmits less photons to the cathode than quartz. Pointing the window downwards reduces the effect from cosmic rays, since the photons are then emitted outwards.

Electroluminescence

Electroluminescence of the glass was found to create light when a voltage as low as 50 V was applied to a glass sample placed in front of a PMT [332]. For PMTs operated with the cathode at ground, the window is not subject to a significant electric field, and no substantial light emission is expected. However, electroluminescence poses a problem when negative voltages are applied to the cathode. The latter is confirmed by [169; pp. 267–268]: Since the inner glass surface close to the cathode is Al-coated and at cathode potential, a large potential differences to the outside will create electroluminescence which is reflected onto the cathode. For this reason a minimum distance of 4 mm to a grounded case around the PMT is recommended when using negative HV [169] [170; pp. 459–460].

Electric discharge

Gas

In a gas, electric discharges cause photon emission through the excitation and deexcitation of gas atoms or molecules, including ionization. [411]
Corona discharges occur in regions of high fields close to sharp points for high voltages near the end of the (dark) Townsend regime.
If the voltage is higher and exceeds the breakdown voltage, a glow discharge occurs. The avalanche sustains itself through production of ions in the gas, which themselves strike the electrode and produce secondary electrons. For continuation of the discharge now a smaller voltage than the breakdown voltage is sufficient. At the same time the current is much higher than for a Corona discharge. If the resistance of the current supply is too high, the current is limited and the gas will remain in the Corona regime with the discharge occurring only near the electrodes.
When the currents increase enough, the electrode becomes sufficiently hot for thermionic electron emission, and if the current supply has a low enough resistance an arc discharge is started. This shows very high currents and a very low voltage is sufficient to sustain the arc.

Vacuum

In a vacuum, field emission or thermionic emission of electrons can lead to a vacuum arc. [411] Photon emission occurs through incandescence of the emission spot and cathodoluminescence on impact, for high currents also through incandescence of the impact spot.

Discharges in photomultipliers

In PMTs, electric discharges *usually* (see below) occur in form of field emission from sharp corners and points of the metallic structures, especially for new PMTs, which benefit from a period of "burning in" [180], and for high voltages [169; p. 72].

This should lead to Corona discharges of the residual gas and vacuum arcs. [401] expects microarcs or discharges at points or across insulators, which are uncorrelated on short time scales and not proportional to the current inside the PMT. [232; pp. 185–186] also expects formation of microarcs at high voltages. [170; pp. 265–266] used an image intensifier to observe light emission from PMTs at usual gains and found that this occurred from the last dynodes, patches of alkali metals on the glass window, and in fact also from sharp metallic points.

Flashers

Flashers are PMTs emitting strong bursts of light, which illuminate neighboring sensors, often accompanied by large pulses in the source PMT. This has been observed in many experiments (among others IMB [154], Super-K [155], SNO [156], Borexino, Daya Bay, KamLAND-Zen,[157] [185, 412–416]), but the behavior varies greatly and is likely caused by different effects.
In SNO the flashing was observed for all PMTs but with a low rate of roughly one flash/week per PMT and otherwise normal PMT operation. Initially this occurred intermittently and often periodic (period 15 min to hours) [185, 412].
Super-K observes PMTs becoming active and then flashing worsens with time, but switching them off for a while restores normal behavior [185].
In Daya Bay some PMTs are more likely to cause flashing [415].
For Borexino flashing appears for single PMTs which then have to be disconnected [413].
KamlandZen observed light emission of one PMT detected by another, which was optically coupled face-to-face to it, for the types Hamamatsu R6091 (3”), R11065-20 (3”), R13444X (4”), and R1250 (5”). This was present in form of intense coincident noise pulses of about 10 ns width, which remained unaffected by the exchange of acquisition hardware and moving the setup location (above and below ground), and was attributed to flashing.

Usually, analysis cuts can be developed to exclude flasher events without affecting the efficiency of physics analyses much.

Possible origins for flashing are:

- Gas or vapor entering the PMT bulb, leading to very intense continuous flashes due to internal arcing and signaling imminent PMT failure [412] (SNO). Borexino suspects water vapor entering the vacuum tube via ruptures and inducing continuous discharges [414].
- Electrical discharges in the base [415] (Daya Bay).
- Corona discharges of trapped air or vaporous epoxy components at the edge of the cathode pin in the epoxy covering the base [417] (Double Chooz). The effect occurred as fast $\approx$ 10–100 ns light flashes or trains within $\approx$ 1–10 µs accompanied by large pulses in the source PMT. The light emission showed a clear increase with voltage and temperature and was reproduced with epoxy samples in air.

[154] Irvine-Michigan-Brookhaven detector
[155] Super-Kamiokande
[156] Sudbury Neutrino Observatory
[157] KamLAND Zero-Neutrino Double-Beta Decay

All explanations involve electric discharges either in the base or the PMT itself, and likely some form of electric discharge underlies all observed occurrences of flashers. It is therefore recommendable to only use tested base materials, optically shield the base in some manner (e.g. by using intransparent casting slip), use optical modules to protect the PMT, and prevent water from entering the PMT base through cables and seals.

Electrode glow

Electrode glow (also called anode glow or dynode glow) is a phenomenological term for light emission observed from the last stages of a PMT (see 2.5.2.2) due to electron bombardment. The emission times are correlated to the transit of the pulse and the intensity increases with the number of electrons in the cascade. All mechanisms producing light in this region contribute to electrode glow, namely bremsstrahlung, cathodoluminescence, photoluminescence and gas excitation/ionization (either in the vacuum, from gas constituents adsorbed on the dynode surfaces or both); [401] also proposed surface polarization as a mechanism. The extent of the individual effects is not completely clear. This is the origin of the γnk and γn1 effects, which were discussed at length in 2.3.5, 2.3.6 and 2.5.2.2.
The intensity of the electrode glow is sufficient to be photographed and seen by the dark adjusted eye [170, 180, 332] and in photon counting applications can activate an adjacent PMT unless shielded [180, 332]. In [332] a monitor PMT was directed at the last dynodes of a test PMT through a blue filter, which saw peaked light emission corresponding to γnk of the last stages and even γnk-γnk. The test PMT showed γnk. The emission displayed a peak around 400 nm with a width of $\approx$ 100 nm, which points to a prevalence of bremsstrahlung as origin.
Electrode glow is the primary reason of breakdown for high voltages and limits PMT gains to a maximum of $1 \cdot 10^9$ [173, 180, 332]. As a consequence, it is necessary to shield the cathode from electrode glow through comprehensive shuttering of the dynodes, which was state of the art as soon as 1967 [332].

Investigation

In the course of this work, the origins and interconnections of light emission and early afterpulses in PMTs were studied based on previous analyses at the MPP [158] [418,419] in collaboration with Max Ahnen (formerly Knötig), Razmik Mirzoyan, Uta Menzel, and Jürgen Hose of the MAGIC [159]/CTA group at the MPP, since these effects were considered potentially problematic for both CTA and LENA [160]. Measurements were carried out for the 1.5" Hamamatsu R11920-100 prototype – a candidate PMT for CTA with 8 linear-focused dynodes, matt photocathode and gain $g = 4 \cdot 10^4$ [419] – and where possible for large LENA candidates (Hamamatsu 8" R5912 and 5" R6594, both with 10 box-and-line dynodes, transparent cathodes and $g = 10^7$). The results were in large part already reported in [420] and [397] for small PMTs and will be repeated, reanalyzed and complemented with the findings for large PMTs here. After this study, the origins of EAPs were analyzed further based on pulse shape analysis,

[158] Max-Planck-Institut für Physik
[159] Major Atmospheric Gamma Imaging Cherenkov Telescopes
[160] Low Energy Neutrino Astronomy

which led to the development of time shift theory (TST), in which all findings are summarized (see 2.3.5, 2.3.6 and 2.5.2.2).

PMT light emission was investigated regarding its location, spectrum, spectrally resolved location, self-feedback, time, and temporally resolved location. At the end of this section, the light shield improvements made by Hamamatsu as a consequence of these studies will be shortly addressed.
The PMTs were illuminated through a fiber coupled to a ps-laser with 405 nm wavelength (PicoQuant PDL 800-B) with occupancies between single photon level and several 100 pe.

Emission location

In order to pinpoint the origin of PMT light emission more accurately, photos of the emission were taken by pointing a charge coupled device (CCD) camera at the dynode chain and cathode of several small and large PMTs. Images with identical takes were recorded a) in ambient light with the device turned off, and b) in dark with applied HV and the PMT illuminated by a laser with high occupancy and repetition rate. As camera an Andor Clara Interline with an f/0.95 Navitar DO-1795 lens, cooled to −55°C to reduce noise, and operated in extended near-infrared mode with sensitivity between 370 and 1000 nm was used. The results are shown in fig. 2.71.

For the 1.5" R11920-100 shots of the dynode chain were taken with a 600 nm filter placed in front of the camera (see p. 421). The image (fig. 2.71a) clearly shows emission from d3 [161] (right bar) and d5 (strongest bar, in the middle) through the gaps of the opposing dynodes. The intensity is larger for d5 as expected from the higher electron number in the cascade at this point. Shadowing by wires in front of the dynode stack is visible, which verifies that the light comes from the inside. The gap is larger between d2 and d4 than between d4 and d6 and very narrow between d6 and d8, but a few photons from d7 can still escape through the latter (weak line on the left). On the other side of the dynode stack a light shield covers the region from d5 to the anode. When looking closely, for d3 and d5 the intensity shows peaks at the dynode brink, which will be discussed on p. 425.
Since the R11920-100 had a matt cathode, no image with front view was recorded. Emission through the aperture to d1 and the cathode was, however, observed in a previous work [418; p. 51].

For the large PMTs shots of the aperture to d1 and the dynode chain through the semitransparent cathode could be taken and are shown in fig. 2.71b for the 5" R6594 and in fig. 2.71c for the 8" R5912. In the aperture of the deck (dark circle/rectangle in center) d2 is visible at the far left and d1 covers everything right of it. Both PMTs show emission from internal paths in the dynode chain leaving the aperture through reflections (vertical bars) and gaps (horizontal lines). For the R6594 the reflection/emission from d2 (bar at left) is strongest, whereas a reflection of d1 dominates for the R5912. The visibility of the reflections off d2 and d1 is expected to strongly depend on the viewing angle, which likely is the explanation for this difference. It is difficult to assess and compare the absolute amount of

[161] dn denotes the n-th dynode.

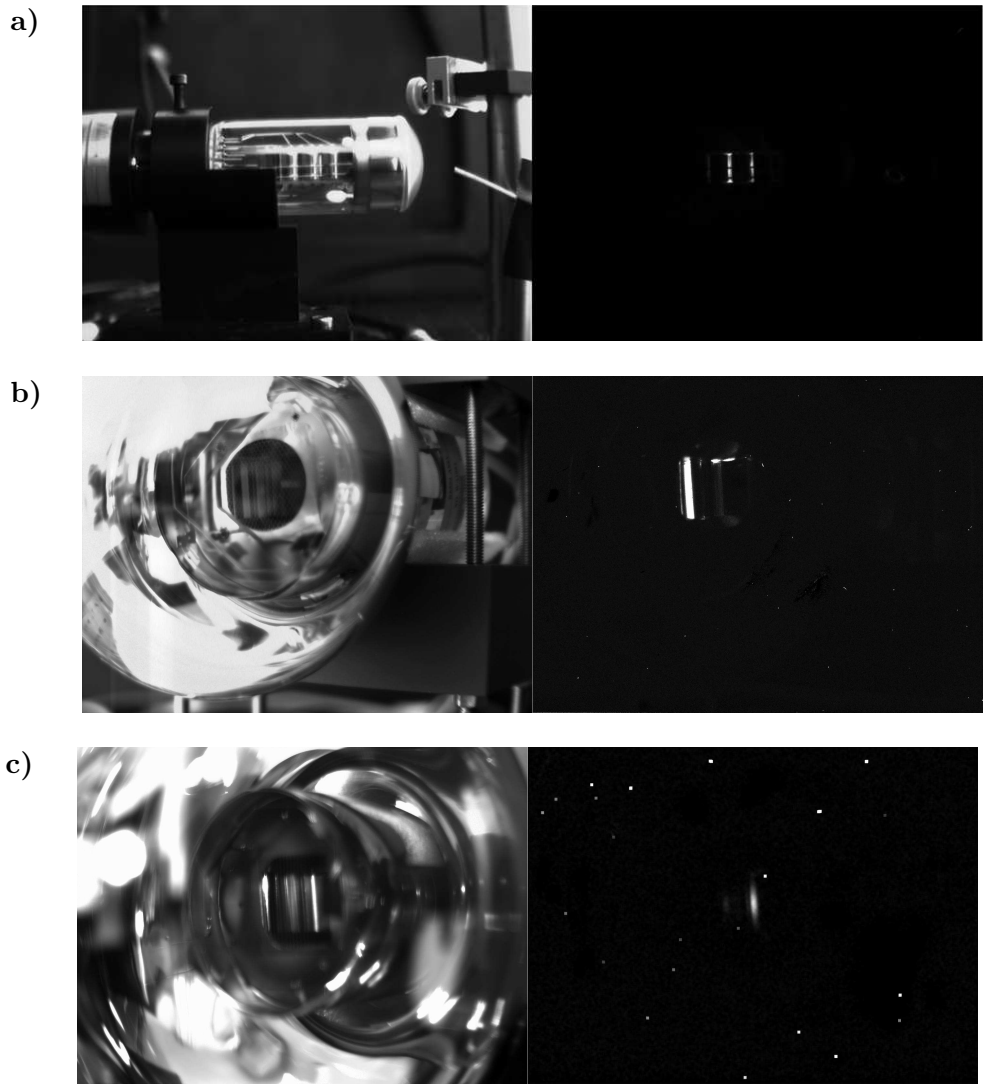


Figure 2.71: Light emission from PMTs: Images of PMTs recorded with a charge coupled device (CCD) camera (Andor Clara Interline), on left with open dark box and room illumination, on right in dark with PMT turned on and cathode illuminated by a 405 nm ps-laser through a fiber with occupancy > 100 pe and high repetition rate. Discussion see text.
a) Hamamatsu R11920-100 (1.5”), view on dynode chain through 600 nm filter, the cathode points to the right, fiber visible on left picture, from [397] (M. Ahnen, U. Menzel). Light emission from gaps in the dynode chain (external paths) is clearly visible.
b) Hamamatsu R6594 (5”), front view on deck, d1 and d2 through the cathode. Emission from the dynodes via reflections on d2 and gaps in the dynode stack (internal paths) are apparent.
c) Hamamatsu R5912 (8”), front view as for R6594. Light emission through internal paths is visible, mostly from reflections on d1. Bright pixels are from muon hits of the CCD (subtracted in a)).

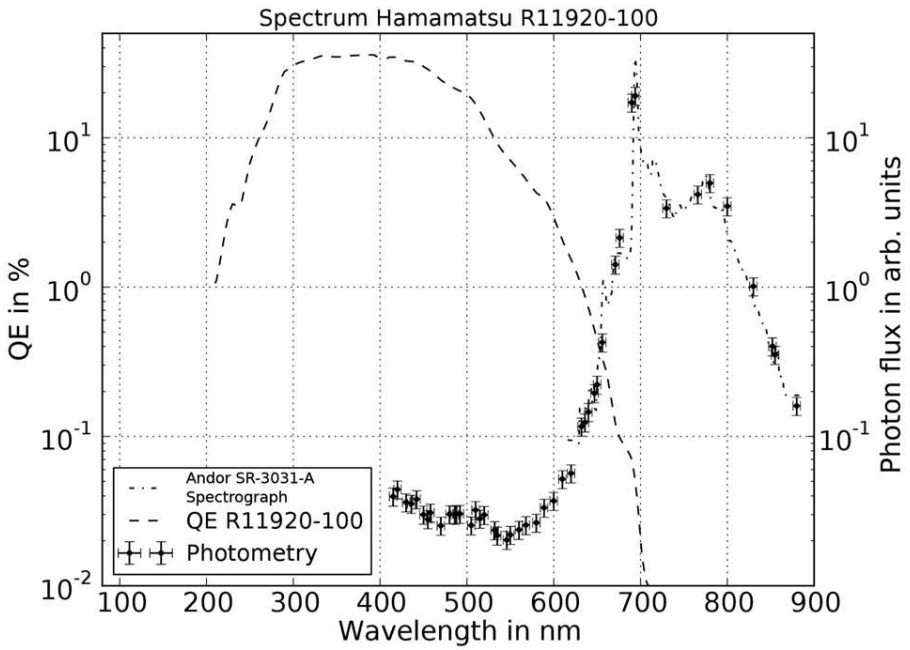


Figure 2.72: Emission spectrum from an R11920-100 from aperture photometry using wavelength filters (dots with error bars) and a spectrograph (dash-and-dot line), both in arbitrary units, versus the quantum efficiency (QE) curve of the PMT (dashed line); from [397]. Discussion and details in text.

light emission for the two sensors. Both PMTs were operated with gains around 10^7 (R6594 +1670 V, R5912 +1425 V), the laser was triggered with 100 kHz rate and the PMTs were in saturation from an occupancy of at least 100 pe. The integration time of the CCD was 30 minutes for the R5912 and 60 minutes for the R6594, which could be one cause for the reduced emission intensity which is observed for the R5912. However, the dependence on viewing angle should prevail and prevents direct comparison (in addition to the saturation).
Additionally, for the R5912 it was attempted to image the light emission from the dynode chain. The dynode stack is completely covered in light shields, and in addition a large-diameter cylinder attached below the deck prevents photons from reaching the cathode (confer fig. 2.39a). An image was recorded from the back at an angle – the only viewing position from which the dynode stack was visible – at the maximum allowed HV of +1800 V with 16 h integration time and otherwise identical setup conditions as the front view. Nevertheless, no light notably above background levels could be detected. This shows that the dynode shields are effective in suppressing external paths and that internal paths dominate for the R5912.

Emission spectrum

To investigate the processes underlying light emission, the emission spectrum of the dynode chain of an R11920-100 was recorded with two methods (fig. 2.72) [397].
In the first procedure (data from M. Ahnen, U. Menzel) a filter wheel with 10 nm wide interference wavelength filters between 415 and 905 nm was placed in front of the Andor Clara Interline CCD camera operated as described above. This allows to identify the emission regions with wavelength dependence, which is discussed in the following section, and perform aperture photometry on the acquired images by

integrating the pixel counts within the area covered by the dynodes after background subtraction. The integration time was chosen short enough for each filter to prevent saturation. The sensitivity curve of this setup was calculated from the transmission through PMT glass and filters (measured with an Anthelie 2 Advanced Spectrophotometer), the QE spectrum of the CCD camera and the transmission curve of the lens. The spectral intensity is then obtained from the sensitivity, the CCD counts and the integration time.
The second method (data from J. Hose) employed an Andor SR-3031-A spectrograph containing an iDus DV420A-OE CCD camera, which had a higher resolution than the measurement with filters, but was sensitive between 650 and 850 nm only.
For both measurements, the laser was pulsed with 100 kHz rate, the occupancy was around 100 pe and the PMT was operated at about 1000 V.

The results of both measures are in good agreement within error bars. Between 415 and 550 nm a continuous emission decreasing towards the red is found, to which one or two small broad peaks around 500 nm contribute. After 550 nm the emission shows an initially slow and then steep rise to a dominant sharp peak slightly below 700 nm, followed by a strong broad peak around 775 nm, which extends to at least 880 nm. Smaller peaks are encountered at about 715, 675 and 655 nm. The dominant contribution comes from the strong lines around 700 nm.
In raw data three outliers were present at 467, 500 and 508 nm. When the CCD is hit by a muon during acquisition, this produces a spike in the spectrum. For this reason these values were not included in the published data; however, considering the theoretical treatment of CL for Cs_3Sb (p. 412) or emission lines from the excitations of gas constituents, these values might have origins in PMT physics (discussed below).

Bremsstrahlung

The continuous slope in the blue spectrum can be attributed to bremsstrahlung of the cascade electrons on the dynodes (and dynode mount). Above 400 nm the calculated bremsstrahlung spectra in fig. 2.70 show virtually no difference between an impact of 45 and 514 eV. In the R11920-100 U_{k1} is fixed at 300 V, and the resistance between other stages is $2R$ for U_{23} and U_{78} and $1R$ for all others [419]. At 1000 V total voltage this leads to interstage voltages of 70, 140 and 300 V, which are within the range shown in the figure. The shape of the bremsstrahlung spectra matches that of the emission spectrum well between 415 and 550 nm outside the peak(s) around 500 nm.

Cathodoluminescence of dynodes

The cathodoluminescence of the dynodes is suspected to cause many of the broad peaks. Since it is unknown, whether Cs_3Sb or BeCu (at surface BeO) is used as dynode material, expected features of both will be briefly compared to the spectrum (see pp. 412 and 413).
For Cs_3Sb the strongest emission line should occur for the band gap of 1.6 eV with 775 nm, which perfectly matches a very strong broad peak. Transitions from the VB border to DOS peaks predict lines at 577 and 468 nm, where the former agrees with a shoulder on the left flank of the largest peak. The wavelength of the latter

corresponds to one of the outlier spikes at 467 nm (not shown, see above), but for this transition a broad peak would be expected. Therefore, this spike is likely from cosmic background and this transition is not observed, in accordance with the expectation that such deep CB levels are likely filled and unavailable for recombination. Hot carrier transitions from the VB maximum to the CB brink and uppermost peak could occur at 539 and 428 nm. The first might correspond to a small excess above the bremsstrahlung, the second is not observed. These wavelength do not match the broad structure around 500 nm; however, it should be noted that the data on which the calculations were based [214] have small uncertainties and already a discrepancy of 0.1 eV is sufficient to match the 467 and 539 nm lines to the left and right part of the 500 nm structure. This is within the expected accuracy of the data shown in fig. 2.69.
For BeO all emission lines are expected below 400 nm with only broad features possibly extending into the measured spectrum.
Due to the good agreement of Cs_3Sb with the observed features, it is suspected that Cs_3Sb dynodes were deployed in the R11920-100.

Cathodoluminescence of dynode mount

The tremendous sharp line near 700 nm was identified to originate from the dynode mount (see p. 425). Inquiries about the composition of its material submitted to Hamamatsu revealed that the dynode mount mainly consists of corundum (Al_2O_3) and a few percent of other elements, namely C, Si, Cr and Mg. Cr impurities in corundum, however, turn the material into *ruby*, which has strong fluorescence lines at 694.3 and 692.9 nm and was used to construct the first laser [421]. This also explains the purple color of the dynode mount in this PMT.
In [422] the temperature dependence of the ruby emission lines were studied. At room temperature, lines at 670, 675, 705, 715, 730 and 740 nm were observed. This matches the features found in the emission spectrum. At 663 K these side-lines were dwarfed by a broad peak around 705 nm extending from below 600 to above 900 nm. If the dynodes are heated up from continuous electron bombardment, this would provide an explanation for the broad peak from 600 to above 880 nm with the Cs_3Sb band gap transition at 775 nm sitting on top of it.
An excitation of the dynode mount of this order of magnitude can only occur from CL. While photoluminescence will also cause the observed emission structures, the photons required for it would have to originate from other processes; PL is merely secondary light emission. Since the contributions to the spectrum from bremsstrahlung, dynode CL and gas excitation/ionization (if present, see below) are far smaller than the ruby peaks, they must be caused mainly by CL.

Gas excitation and ionization

The excitation and ionization of residual gas constituents could be visible in form of characteristic emission lines in the spectrum. In the ionic afterpulse (iAP) analysis of the R5912 a strong presence of Hg, Cs, N, N^2, Ar and to a lesser extent of He and H was found (see p. 392), but of course for the R11920-100 the composition could differ.

The following list compiles the strongest emission lines for the neutral and single-ionized atomic elements between 400 and 880 nm according to [423]; the unit is nm,

close lines were merged, and relative intensities are given in brackets and are not necessarily comparable for different measurement conditions [423]:

- **Hg**: 436 (1000), 543–6 (580), 405 (400)
- **Hg$^+$**: 615 (265) and 794 (250)
- **Cs**: 460 (1000)
- **Cs$^+$**: 523–5 (1100), 852 (1000), 696–8 (645), 593 (500), 495–7 (480), 556 (400)
- **N**: 857–872 (1330), 742–7 (550), 575 (150)
- **Ar**: 806–810 (280)
- **Ar$^+$**: 671–672 (1300), 723–726 (1150), 662 (150)
- **He**: 584 (1000), 537 (400), 506–22 (276)
- **H**: 656 (300), 486 (100), 434 (30)

The wavelength corresponding to the energy of the first ionization ($e^- + \mathrm{A} \rightarrow \mathrm{A}^+ + 2\,e^-$) lies in the UV range for all listed elements and outside the measured range: Hg 119 nm, Cs 319, N 85, N^2 80, Ar 79, He 50, H 91, H^2 80 [423, 424].

In general, sharp lines are expected unless several lines are very close. If one line matches a feature in data, the others should also be present with about the corresponding relative intensity.
No notable Hg peak is observed, but Hg^+ peaks could match low intensity features in data.
Cs would lie at or beyond the border of the bandpass filter centered at 467 nm which showed one discarded spike (see p. 422). The 524 nm Cs^+ peak could contribute to the right part of the 500 nm ridge, 593 nm and 556 nm might fit bumps in the left ruby flank, and the peaks above 600 nm are dwarfed by the ruby peaks. 496 nm would match the spike at 500 nm but is too weak compared to the other lines to explain it and could rather contribute to the left part of the 500 nm bump.
The 575 nm N peak would match a shoulder, the other peaks would be covered by ruby.
Ar^+ is indistinguishable from ruby peaks, while Ar produces no discernible peak.
He and H might contribute weakly to other peaks.

Except for the tenuous match of the 460 nm Cs excitation none of the sharp outliers are represented by emission lines, so likely all spikes are caused by cosmic background or errors in the calibration and are rightfully discarded.

Many of the transitions predicted from gas excitation/ionization overlap with broad structures, which can be attributed to other effects. The gas-based emissions, however, do not produce prominent lines and cannot explain the broad features by themselves. In conclusion, light emission from residual gas is certainly weaker than the ruby lines but can contribute to many of the observed smaller features. Its intensity appears to be lower than the level of bremsstrahlung and dynode CL in the region not covered by ruby CL.

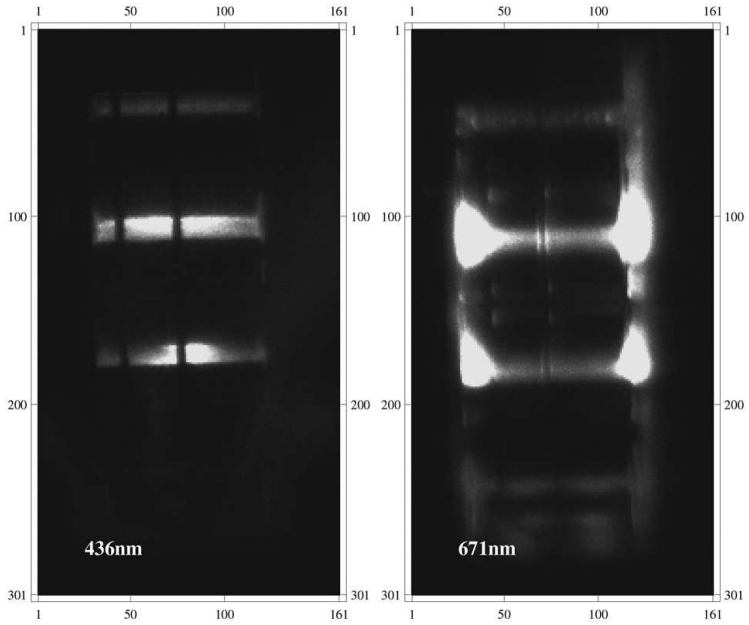


Figure 2.73: Close-up of the R11920-100 dynode chain with wavelength filters attached to the CCD camera; from [397]. Cathode at bottom. Discussion in text.

Comparison

If another dynode holder material is used – as is the case for the R5912 whose dynode mount is green and not reddish – the ruby peaks will disappear, which likely leaves bremsstrahlung as the strongest effect, with notable contributions from dynode CL and/or gas constituents. A study of such a PMT would be a worthwhile endeavor fur future investigations.

In fig. 2.72 also the QE curve of the R11920-100 is plotted. This shows a peak of 35% near 390 nm and the QE is three orders of magnitude lower at the position of the ruby CL. Since this is roughly the factor by which the emission from ruby CL is larger than that of the bremsstrahlung, the emission in the blue has the same peak detection probability as the ruby emission and an even higher one when integrated. Thanks to the rapidly dropping QE thus not the ruby peak but *bremsstrahlung* is the dominant light origin of photonic afterpulses. Furthermore, because the emission lifetime of ruby CL lies around 3 ms [407, 408], its APs cannot typically be attributed to this effect and effectively raise the DCR.

In conclusion, cathodoluminescence of the dynode mount and likely the dynodes, as well as bremsstrahlung could be verified as main origins of electrode glow, while gas ionization/excitation and photoluminescence might also occur but play a smaller role.

Location–wavelength dependence

In fig. 2.73 two CCD images recorded as part of the wavelength-resolved aperture photometry (M. Ahnen, U. Menzel) are shown, with central filter wavelengths of 436 and 671 nm.
At 436 nm, which lies within the range dominated by bremsstrahlung, as expected the emission is strongest at the center of d3 and d5 where the cascade is focused by electron optics.

Within the ruby peak at 671 nm, however, the light originates from the dynode mount at the brink of d3 and d5. The intensity is strong enough to illuminate the whole dynode stack through reflections and diffuse scattering, which highlights all gaps between dynodes and dynode holder.

Emission time and self-feedback

Using the technique introduced by [332], a monitor PMT was directed at the dynodes of a close-by R11920-100 illuminated by the ps-laser with spe occupancy to observe the time-resolved light emission (fig. 2.74).

Improved transit time calculations compared to [397] are now possible using TST. Based on the dimensions and voltage divider ratios given in [419] and [397], assuming 2.6 cm distance between cathode k and d1, 1 cm between all dynodes and 0.15 cm between d8 and anode a, as well as the voltages already stated above (1000 V total) the following transit times were computed:

- Total electron transit times [ns]: From k to {1,2,3,4,5,6,7,8,a} = {6.8, 9.0, 11.1, 13.7, 16.4, 19.0, 21.1, 23.3, 23.8}, with tk1 the mean of linear and quadratic field, all others interstage transit times for linear field using the potential of the dynode *after next* (see p. 180) except for t8a, pk and sn emission energy assumed at 1.1 and 5 eV.
- Delays of γnk: n = {1,2,3,4,5,6,7,8,a} → {6.9, 9.1, 11.3, 13.9, 16.6, 19.2, 21.4, 23.6, 24.1}
- L1 relative transit time T_r: linear fields 9.5, quadratic 17.8, mean of both 13.6

The calculated total transit time of 23.8 ns is in good agreement with the measured value of 23 ns [419].

These values can be compared to the features seen in test PMT and monitor PMT (fig. 2.74). Photon flight times between the dynodes of the test PMT and the monitor should be less than 0.2 ns and can be neglected.

The test PMT showed a peak at ≈ 7.1 ns delay after the γk peak from the laser, which matches γ1k (6.9 ns). The second peak at 10.4 ns does not closely match any feature and most likely stems from L1 (13.6 ns) whose T_r is shifted forward by either oscillations of the APS (visible in [182]) or a more complicated field layout between k and d1, which results in a value closer to that for linear fields (9.5 ns). This would require further measurements.

The important result is that the monitor PMT detects light emitted from the dynodes of the test PMT with a prominent peak 21 ns after the laser peak (γk) of the monitor, with some previous events as early as 12 ns and a trailing flank extending up to 32 ns. The strong peak can *not* be from time shift effects (TSEs) of the monitor PMT, since these would be smaller in intensity than the laser peak. This shows that PMTs are sensitive to their own emitted light – the QE curve of the ETE[162] 9111 is similar to that of the Hamamatsu R11920-100.

[162] Electron Tubes Enterprises Ltd.

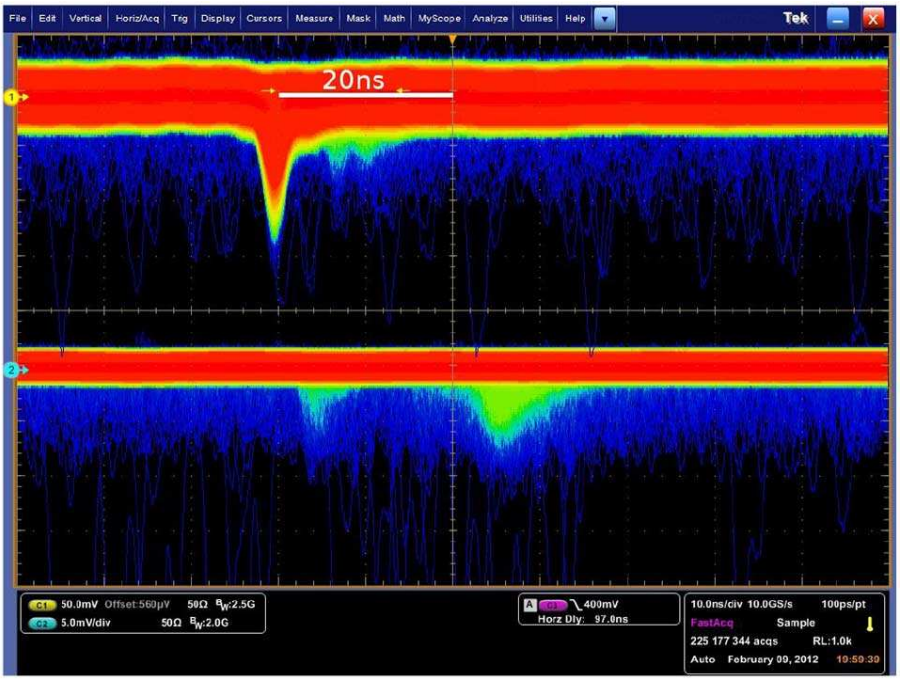


Figure 2.74: Output of a test PMT (R11920-100) illuminated by laser light (top) and a monitor PMT (ET Enterprises 9111, bottom) directed at the dynode chain of the former, recorded with a digital oscilloscope (Tektronix DPO7254-C); from [397]. In the test PMT γ1k and L1 are visible at delays of about 7.1 and 10.4 ns. In the monitor PMT the first peak is from reflected/scattered laser light (γk), the second peak with a delay of $\approx$ 21 ns results from light emission from the dynode d7 of the test PMT through the gap between d4 and d6 after internal reflections. On closer examination also events from earlier dynodes are visible at 12 ns (d3) and 16.4 ns delay (d5). The trailing flank after 21 ns could contain light emission from d8 and the anode and dynode cathodoluminescence.

In [397] the strong peak at 21 ns delay was attributed to emission from d5; however, with the more accurate calculations available here, interestingly, the transit time of the cascade to d7 gives a better match (21.1 ns) than d5 (16.4 ns). This indicates that the majority of the light emitted from d5 (see fig. 2.71a) in fact originates from d7 (which itself is covered) and is reflected or scattered to d5.
At 16.4 ns delay a small peak is present, which corresponds to emission from d5; the events at $\approx$ 12 ns would match emission from d3 (11.3 ns). As expected, the number of emitted photons increases with stage.
Since the dynode geometry is identical after d2, the emission from d3 visible in fig. 2.71a might also mostly originate from later dynodes. This is corroborated by the weakness of the feature matching d3 compared to the peak of d7 in fig. 2.74.
The d7 peak likely also contains emission from d8 and the anode at later times with reduced intensity. The long trailing flank possibly stems from fluorescent CL of the dynodes (500 nm ridge).

Location–time dependence

The dependence of emission location on time was studied by recording an emission video with a gated camera consisting of a Hamamatsu C9546 gated image intensifier in front of a cooled SBIG CCD camera (fig. 2.75). The image intensifier had a peak sensitivity at 400 nm, the minimal gate width of 3 ns was used. The measurement

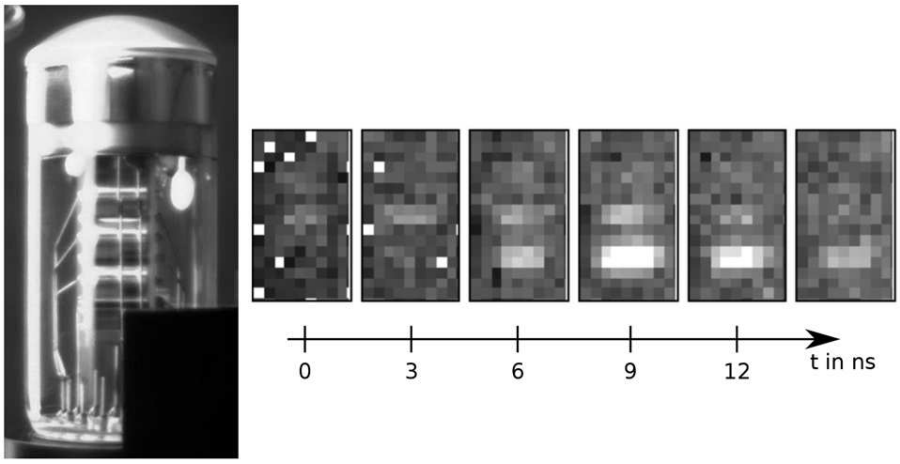


Figure 2.75: Video of the light emission from the dynode chain of a R11920-100, recorded with a CCD camera (SBIG) connected to a fast gated image intensifier (Hamamatsu C9546) whose delay was adjusted in steps of 3 ns (arbitrary offset); from [397]. A scaled and aligned picture of the PMT is shown on the left, the cathode is at the top. Light emission is visible first from the gap in the dynode chain between d2 and d4 and then also from the gap between d4 and d6.

(M. Ahnen) was performed through acquisition of a series of images with exposure of 1 h, 30 kHz laser rate and occupancy ≈ 500 pe each, while adjusting the delay of the gate, and was controlled through LabVIEW [425].

The resulting images clearly show emission from d3 and a stronger and later emission from d5 through the gaps between dynodes. In view of the discussion for the measurement with the monitor PMT, it appears that the emission from later dynodes is scattered onto d3 and d5 and can leave the respective gap. This explains the large time range over which emission from the d3 gap (12–13.5 ns) and the d5 gap (at least 9 ns) is observed. If the cascade arrival at these dynodes were stretched over such large times, the pulse width would be far larger than observed (3.0 ns [419]), as would the transit time spread (1.3 ns FWHM [163] [419]). With this in mind, the offset of the time axis can be estimated to approximately coincide with arrival of the cascade at d3. d4 is then hit around 2.6 ns on the time axis in the figure, d5 at about 5.2 ns, d6 at 7.8 ns, d7 at 10.0 ns, d8 at 12.2 ns and the anode at 12.7 ns. This offset matches the first frames in which d3 and d5 are visible. The peak emission for both d3 and d5 then is from d6 or d7. The different shape of d8 could explain why the d5 gap peak is not from d8 or a. The collection of the cascade at the anode coincides with the decreasing light from d5 in the last frame. Dynode CL fluorescence could contribute to a trailing emission from both gaps. As found before, the emission intensity is higher for later stages, with the exceptions listed above.

In [397] the interstage transit time was estimated from the peak of both emissions to ≈ 1.9 ns, which has the same magnitude as the calculated values of $t34 = t45 = 2.6$ ns, but is significantly lower. The superposition of reflected light from later dynodes with the weaker emission from d3 and d5 itself could shift the peak times and explain the discrepancy.

[163] Full width at half maximum

The images show emission from the dynodes but not the dynode holder. The ruby lifetimes of $\approx 3\,\mathrm{ms}$ would not be visible on the ns scale.
Since the image intensifier is more sensitive in the blue, this confirms the origin of the bremsstrahlung from the dynodes, but cannot exclude emission in the red on short time scales.

External versus internal paths

For the R5912 no light emission from the dynodes was visible, which is in all likelihood due to the thorough shielding of its dynode stack. In addition, the dynode mount had a different color and was green instead of purple, which indicates that a material with different composition was used here. Thus no ruby CL is expected, which greatly reduces the amount of detectable light and probably is the reason why such large illumination times (30–60 minutes) and intensities (see p. 421) were required to detect light emission through the cathode. Since emission through the cathode was observed, internal paths to the cathode appear to dominate.

For the R11920-100 shielding the complete dynode stack effectively eliminated light emission from the dynodes [182] and reduced the number of γAPs [181]. This shows that external paths to the cathode are present if not blocked by shields. The geometry of linear focused designs shows a gap between dynodes which extends to d1 as in the box-and-line layout (see fig. 2.15), but likely reduces the transport factor from reflections inside the dynode chain. Therefore a reduced contribution of internal paths is expected compared to box-and-line designs. After shuttering the dynodes, no increase of the EAP rate with voltage was observed in [181], unlike if without shields (increase up to $3 \cdot 10^{-3}\,\%$ of DCR, no absolute rate quoted). However, γ1k is independent of voltage (U_{k1} is fixed) and was visible in the PMT waveforms (fig. 2.74), whereas γ2k and later did not appear. Thus the light suppression for internal paths from d2+ is effective, but d1 must be visible from the cathode for the device to function and its light emission can never be prevented. The γ1k peak had a comparable intensity to the L1 peak, which typically has a probability of about 2–3% and likely is a bit weaker here due to the rather small U_{k1}.
In conclusion, for the R11920-100 both internal and external paths create γnk, but with the present data it can not be concluded which dominates.

These results highlight the importance of a) proper dynode shielding and b) designs suppressing internal light feedback for box-and-line layouts.

Improvements

One possible countermeasure against light emission is to completely shield the dynode stack. A reduction of the aperture to d1 would reduce internal paths but also lower the collection efficiency (CE), which is not an option. The ruby emission could be eliminated by use of a different dynode mount material as for the R5912, or by admixture of iron to the corundum [397]. Since the emission scales with the current in the PMT, its severity can also be reduced by operation at low gain.

In fact, as a consequence of the results of these measurements, Hamamatsu produced prototypes with two improved shield variants which both completely enclosed the dynode stack [181]. Both reduced the γAPs from external paths which

increase with voltage to non-measurable values [181] and practically eliminated the light emission seen by a monitor PMT [182]. In the selected variant the shields are fixed directly to the dynode mount (see fig. 8 in [182]).

Measurements by KM3NeT [164]

The KM3NeT [165] collaboration uses a digital optical module including 24 Hamamatsu R12199-02 (3") PMTs packed close together in a 24" glass sphere. When turning on all PMTs an enhanced DCR was observed which was assumed to originate from light emission from either the PMTs or discharges in the bases.
Based on the procedures discussed before, measurements were carried out to identify the origin. A monitor PMT was placed facing the dynode chain of a test PMT and observed increased rates when powering the test PMT. Placing a paper between the both eliminated the increase and proved that it resulted from photons and not electrons or an electromagnetic field. Applying a black coating to the envelope around the dynodes eliminated the emission and showed that it did not originate from the base. Discharges on the glass were refuted as cause of photon emission by measuring a PMT coated with an insulating but transparent layer. Photon emission was shown to be mostly from single photons, which substantiates that discharges were not the origin. A timing measurement with the monitor PMT facing the cathode showed emission from d1. CCD images of the dynode chain were acquired to locate the exact source but could not find light emission; images of the cathode were not taken. Due to these results the PMTs will be coated in the final digital optical module.
These findings support the ones presented here.

2.6.2 Radiopurity [166]

The most critical radioactive contaminants which are contained in traces in PMT materials are ^{40}K, ^{232}Th and ^{238}U, where the latter two are the starting point of decay chains (e.g. [183; p. 10]). Most critical for operation in a large neutrino detector are the produced gamma rays, since these can penetrate meters into the active volume where they generate fake events, which requires to limit the detector volume with fiducial volume cuts.

The PMT consists of glass (window, tube and stem), ceramic (dynode mount) and metal (dynodes, focusing electrodes, anode, wires, grids, Al coating). Glass and ceramic are typically produced from natural sources, which is why they introduce the most radioactivity. The largest contribution comes from the glass, more specifically its window.

Glass

The main component of glass is sand, some of whose sources have lower activity than others, which nowadays are known to all PMT manufacturers. Potassium is added

[164] Summary of [426]
[165] Cubic Kilometre Neutrino Telescope
[166] Content based on [170; pp. 274–276]

to the glass to achieve desirable physical properties; leaving it out causes formation of striations, cords and bubbles in the glass, which according to Wright do not affect performance or longevity [167]. However, natural potassium contains 0.0118% of ^{40}K, which requires to minimize its content for low-background applications.

Glass variants

Today, PMT glass types with low and ultra-low background from radioactive elements are offered by the manufacturers [189, 201] (see p. 51). Even then PMTs are often one of the largest sources of radioactive background in a sensitive detector such as JUNO [168] or LENA. For ETE PMTs, standard glass contains < 60000 ppm K_{nat}, < 1000 ppb ^{232}Th and < 1000 ppb ^{238}U, low background glass 300 ppm K_{nat}, 250 ppb ^{232}Th and 100 ppb ^{238}U, and ultra-low background glass (B53) 60 ppm K_{nat}, 20 ppb ^{232}Th and 10 ppb ^{238}U. For example, the complete envelope of the 8" 9354 is made from B53 [263].
To fulfill the extreme demands of the Borexino and SNO experiments, a special charge of Schott 8246 glass was produced [169], which had [170] $\approx$ 30–33 ppm K_{nat}, $\approx$ 10–25 ppb ^{232}Th and $\approx$ 18–34 ppb ^{238}U [170] and was used for the ET [171] 9351 and Hamamatsu R1408 PMTs [80, 162]. Comparison to table 1.1 shows that similar glass will be required for LENA.
Where even lower activities are needed, fused silica – synthetic quartz made from silane gas (SiH_4) – can be used, which is essentially free of radioactive background (< 5 ppm K_{nat}, < 5 ppb ^{232}Th, < 5 ppb ^{238}U) and offers window transmission down to 160 nm [170, 201].

Window

The largest part of the activity is introduced by the window, which contains the largest mass and is closest to detector and photocathode. Most PMTs thus have low background glass only in the window [427]. ETE offers three configurations where either the window, window and bulb, or the whole envelope including the stem are made from low background glass.
A PMT made entirely from quartz is not possible since the metal pins in the stem (Hamamatsu uses Kovar [169; p. 38]) have a significantly different expansion coefficient to fused silica. In this case graded seals – fused glass rings with slightly different expansion coefficients – can be used to couple a glass window and bulb to the PMT base made of borosilicate glass (also [169] [189; p. 6]). However, the graded seals themselves have rather high activity, typically 7200 ppm K_{nat}, 1030 ppb ^{232}Th, 420 ppb ^{238}U according to Wright, which is too high for neutrino detectors. In addition they are vulnerable to shock [169].
ETE developed an all-quartz PMT with a band of Mo close to the pin connections; Hamamatsu offers a PMT with quartz window and aluminum envelope with very low background. But such devices are rare and very expensive.

[167] While the influence on optical properties may be minor, the influence of at least bubbles on pressure compatibility might be a serious issue.
[168] Jiangmen Underground Neutrino Observatory
[169] According to Wright possibly the glass with the lowest activity ever manufactured.
[170] It is difficult to measure such low activities.
[171] Electron Tubes Ltd.

Ceramic

The ceramic rods and side plates of the dynode mount can notably contribute to the overall radioactive background if the materials are very impure. Ceramic material with radiopurity comparable to ultra-low background glass is available.

Metal

The metal parts of the device are largely free of activity.

Base

Finally, the materials of the base or optical module may not be neglected. Lists of the measured radiopurity of detector materials are presented in [428,429]. Some electronic components of the voltage divider can contain rather high radioactive backgrounds. For sensitive experiments careful screening of all materials and components is required before installation.

2.7 Environmental factors

2.7.1 Magnetic fields

Photomultipliers are very sensitive to even weak external magnetic fields such as the geomagnetic field whose total intensity amount to about 24–66 µT [166]. The reason for this is that the Lorentz force can deflect the electrons from their path so that they miss the next amplification stage altogether, which lowers the collection efficiency (CE), or hit it with suboptimal position or angle, which reduces gain [195; pp. 195–196] [169; pp. 252–254].
This is most severe for the k–d1 region due to the long trajectories and even more so for large PMTs [170; pp. 456–458, 461–463] [169]; on p. 179 it was shown that even without magnetic fields for large PMTs the spot size on d1 due to lateral pk momentum is larger than d1, which limits CE.
Magnetic fields, however, also have an influence on the movements in interdynode space.

The Lorentz force acts perpendicular to the movement of a charged a particle and for an electron it reads [363; p. 36]

$$F = e\,v\,B\sin(\theta)$$

with v the velocity, e the elementary charge, B the magnetic field strength, and θ the angle between the direction of movement and the magnetic field.
The deflection for a linear field with voltage U over a distance x and the electron mass m amounts to [363; p. 37]

$$z = \sqrt{\frac{2\,e}{9\,m}}\,\frac{B\,x^2}{\sqrt{U}} \tag{2.224}$$

For a 20” Hamamatsu R12860 k–d1 path with $B = 44\,\mu\text{T}$, $x = 35\,\text{cm}$ and $U = 700\,\text{V}$ the deflection is no less than 4.0 cm [363; p. 65].

Affected properties

This affects a whole series of crucial sensor properties [430] [171; p. 19] [169; pp. 252–254] [195; pp. 195–196].
The CE is reduced through missing pk or sn, which lowers the photodetection efficiency (PDE). A non-ideal hit produces less secondary electrons, which decreases gain and charge.
Often measurements are not made separately for CE and gain, but instead the output current is studied, which is proportional to the product of both. The output generally decreases with increasing magnetic field [195], see fig. 2.76, and for most PMTs drops to 50% at external fields of less than 0.1 mT up to a few mT [169]. Interestingly, the overall CE or gain not necessarily has its maximum at zero field, and often side maxima are observed (e.g. [171, 190, 195]). This apparent improvement, however, in all likelihood deteriorates other properties.
Since the 1 pe charge is affected, so are the peak-to-valley ratio (P/V) and the charge resolution.

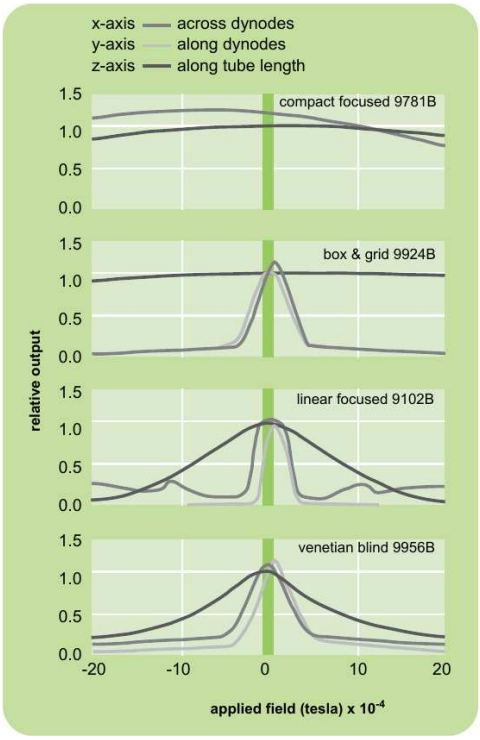


Figure 2.76: Relative output (proportional to collection efficiency times gain) vs. magnetic field for ETE 1–2" small PMTs with different dynode types, from [171; p. 19].

In addition, due to the curved trajectory the total transit time and the transit time spread are altered. The occurrence of the time shift effects (TSEs) and early afterpulse (EAP) types should also depend on magnetic fields; [430] did not find a large dependence for the R5912 with high and standard QE and the R7081, but the measurement conditions (0.33 pe threshold, 10–80 ns) only could detect BL1 and γnk.

Dependence on orientation

The influence of a magnetic field strongly depends on the orientation of the PMT towards it as well as the degree of symmetry of the PMT layout (see fig. 2.76) [195] [170; pp. 456–458, 461–463] [169]. While the cathode is rotationally symmetric, the dynodes have a mirror symmetry, which becomes visible in the field dependence. Magnetic fields along the rotational axis of the cathode have the least influence, as only the movement components vertical to it cause deflections. The device, however, is very vulnerable against fields perpendicular to this – either along or across the dynodes – since the main component of the electron movement vector contributes to the Lorentz force.

Dependence on layout and voltage

Magnetic fields are more critical for long electron trajectories and small apertures to d1, see (2.224). For higher voltages the effect of magnetic fields is reduced [169] and a front-tapered voltage divider therefore can also improve the magnetic tolerance [213].

Dependence on dynode type

Lastly, the influence depends on the dynode type (fig. 2.76). Proximity focusing dynodes such as metal-channel and fine-mesh are much less sensitive to magnetic fields than fountain focusing types (e.g. linear-focused and box-and-line) [195]. Small PMTs with linear-focused, box-and-grid and venetian blind dynodes are similarly sensitive to fields perpendicular to their rotational axis and their output is halved at about 0.2 mT [170]. For metal channel PMTs this is the case at 10 mT due to short interstage distances and shielding effects of the metal housing [169, 170]. Fine-mesh dynodes only lose sensitivity around 1 T, which makes them suitable for high-energy physics. The typical field resistance along the tube axis is 0.1 mT for linear-focused, circular-cage, box-and-grid and venetian blind, 5 mT for metal channel, 500–1500 mT for mesh and 1500 mT for MCP-PMTs[172] (if this is considered a PMT subtype) [169; p. 47].
Many hybrid photosensors are immune to the terrestrial magnetic field, since they are either proximity focused or use very high voltages between the cathode and the first amplification stage (see 4.3.1.1 and 4.3.1.2).

Sources of magnetic fields

For this work only fields in the same order as the geomagnetic field (GMF) are relevant, that is, around 0.1 mT.
The GMF itself varies between about 24–66 µT total intensity across the surface of the earth. One can use a field model such as the International Geomagnetic Reference Field [166] to estimate the local field and its vector. However, before construction of a detector measurements of the exact fields, which also depends on local anomalies and laboratory sources of currents and magnetism, are required.
Proximity of other sources of magnetism should be avoided, which includes magnets, transformers, vacuum pumps and electromagnetic relays [169, 170]. Furthermore, exposing the PMT to magnetic fields must be avoided *in general*, since the dynode substrates can contain nickel and the leads and electrodes can be magnetized [169, 195]. Weak field such as the GMF pose no problems, but higher fields cause residual magnetization, which permanently changes the gain, etc. The device can be demagnetized with a gradually decreasing alternating current magnetic field.

Examples

Large area PMTs often show tremendous drops in performance through the geomagnetic field alone.
For the 12" Hamamatsu R11780 for transverse fields between ±45 µT the output varied between 150% and 0% of the value without field [190], which makes it unusable without shielding. Within ±50 µT transverse field the 10" Hamamatsu R7081 showed a P/V decrease from 4.2 to 1.2, the gain dropped by 27% or improved by 7%, and the CE decreased by 50% [431]. Both PMTs employ 10 box-and-lines dynodes.
ETE PMTs seem to be more field tolerant: The output of the 11" D784 varied by $< 10\%$ up to 45 µT transverse fields [306], the 8" 9354 shows an output decrease by 50% only at 400 µT [263]; both PMTs have 12 linear-focused dynodes.
Although several models show reduced sensitivities, for all large PMTs magnetic compensation in one form or another is required to achieve the targeted performance without losses.

[172] Micro-channel plate PMT

Field compensation

The influence of magnetic fields can be reduced by either passive magnetic shielding (placing the device inside a material with high permeability), or active compensation of the field by running a current through coils placed around the experiment.

Shielding

The shielding factor $S = \frac{H_e}{H_i}$ of a magnetic shield is the ratio of the unshielded exterior field H_e to the remaining internal field H_i. For a thin, long cylindrical shield in a transverse field this can be approximated with [432, 433]

$$S_t = \mu_r \frac{d}{D} + 1 \tag{2.225}$$

where μ_r is the relative permeability of the shield, d its wall thickness and D its average diameter. Eq. (11) in [432] provides a formula for a shield with multiple layers.
Towards both ends of the cylinder the shielding decreases, and the transition region extends inside the shield along the rotational axis by roughly half the diameter of the cylinder [189; p. 9]. [434] gives a more accurate formula for a cylinder with finite length which accounts for edge effects, but is valid only on the rotational axis. Analytical treatment is limited to a few special cases, so for the exact field distribution inside the shield numerical simulations are required.
It is most important to cover the k–d1 region. However, it is recommendable to cover the complete dynode stack, and from the edge effect follows that it is best to extend the shield by half its diameter beyond the cathode [195; p. 195] [171; p. 19] [170; p. 457]. However, this is seldom possible since it occults the field of view of the PMT.

The commonly employed solution to the sensitivity to magnetic fields is to wrap the PMT in one layer of mu-metal – a general term for many proprietary Ni–Fe alloys, all of which have very high relative permeability μ_r up to about $8{\cdot}10^4$–$1{\cdot}10^5$ [363, 434]. This material has to be annealed under hydrogen atmosphere in its final shape for optimum permeability and is very sensitive to shock or bending, which complicates handling and processing. Water jet cutting and gluing or possibly spot welding have to be used for mu-metal foil if another post-processing annealing step is to be avoided. Ready-to-use magnetic shields are offered by PMT manufacturers up to 5" [189] resp. 8" diameter [201]; for larger diameters home- or custom-made solutions are necessary. Wrapping one turn of 0.125 mm thick mu-metal foil around the PMT is usually sufficient for shielding; further turns or using a thicker formed shield increases screening [170; pp. 456–459, 461–463].

In the last years, other high-μ_r materials have emerged on the market.
FINEMET is an amorphous metallic alloy consisting of nanocrystalline grains of Fe, Si, B, and small amounts of Cu and Nb [435]. A 0.12 mm thick sheet has a maximum μ_r of $7 \cdot 10^4$ and contained low amounts of radioactive impurities, for the ^{238}U series 18 ± 5 ppb, the ^{232}Th series 15 ± 5 ppb, and K_{nat} 4 ± 2 ppm. Rn daughters can, however, slowly accumulate on the surfaces if not stored in ventilated areas. Accelerating aging by immersion in mineral oil at elevated temperatures and monitoring the absorbance of the liquid (as in 5.3.2) showed no notable increase

between 300 and 650 nm after 28 days at 40°C. In purified water at 70°C, however, the absorbance increased below 480 nm, with a rapid initial rise down to 300 nm, which leveled off after two weeks. Below 300 nm the absorbance continued to rise during the whole exposure time of 26 days. The integrity of FINEMET did not change in both liquids.
Supermimphy is a Ni–Fe alloy whose permeability with $3{\cdot}10^5$–$5.2{\cdot}10^5$ is even higher than that of mu-metal [434, 436].

Soft iron (high purity iron) also can be used for shielding. While its μ_r with about 500 is far lower than mu-metal [433], it is insensitive to shock. For strong fields [195] recommends to also place a soft iron layer around a mu-metal shield.
Mild steel (low carbon content) has a μ_r between 800 and 2000 [363].
Better than both is silicon iron, which consists of typically 96% Fe and 4% Si, is ferromagnetic and has a high relative permeability of up to 7000 when annealed [363].

A magnetic shield can also be used for electrostatic shielding by connecting it to cathode potential [170].
High fields which are part of an experiment cannot be shielded since this would affect the field distribution in the detector [170]. In this case sensors compatible with high fields have to be used (see above).

Active compensation

One can also compensate the magnetic fields by placing large current-carrying coils around the measuring setup resp. detector. Usually rectangular coils are used for simplicity. Sometimes only the main component of the terrestrial field is compensated; good setups include coil pairs for all three axes [190]. Even then it is difficult to achieve good uniformity in the complete PMT or instrumented space, field simulations and field measurements are required to adapt layout and currents, and the power consumption introduces running costs, which can add up for large detectors [185; p. 2-79]. It is often easier and sufficient to use passive shielding.
For optimum PMT performance, passive and active compensation can be combined [428].

2.7.2 Pressure

The Super-Kamiokande implosion

Following 4.6 years of operation of the Super-Kamiokande detector (Super-K) without failures maintenance was performed by draining the detector and replacing some PMTs. On 12 November 2001 during refilling at about ⅔ height (total height 36 m) and 0.400 MPa pressure, a single PMT on the detector floor imploded [41, 185, 313]. The resulting pressure pulse caused 6779 of 11146 PMTs (61%) to implode in a chain reaction, destroying most PMTs > 5 m below the water surface, while all within 5 m survived. During operation the detector was filled with ultrapure water and 0.480 MPa maximum pressure at the tank bottom where about 1800 PMTs are located. Hamamatsu R3600 PMTs (20") were installed, which had ≈ 4 mm average glass thickness and actually were rated for 0.608 MPa, well beyond the pressure in the

detector. Reconstruction of the event lead to the conclusion that the triggering PMT was likely damaged by impact during maintenance and imploded at the neck. This created a shock-wave with 10 ms delay (due to the large size of PMT), which produced a peak pressure of 10 MPa over 50 µs at the neighboring PMTs and destroyed them. As a consequence, during reconstruction Super-K installed acrylic PMT housings that slow the in-flow of water to prevent or attenuate a shock wave in case of implosion. In addition, refitted PMTs were tested with 0.65 MPa pressure before installation, with a rejection rate of 0.3% [41; p. 115].

This alarming example shows the importance of the compressive strength of the PMT glass envelope. If the ambient pressure exceeds the one listed by manufacturers or found in pressure tests, it is of paramount importance to enclose the PMT in a pressure housing (see 5.4). Even if the surrounding pressure is well within the pressure capability, it is recommendable to at least use a protective cover, which mitigates the formation of a shock wave and at the same time shields adjacent devices.

After the Super-K incident, even greater attention was directed on the pressure compatibility of PMTs in the construction of large detectors (e.g. [185, 313]) and many tests and improvements were carried out (see below).
SNO developed a conservative accelerated testing protocol based on data of fracture stress versus time, which states that testing for one hour at the double pressure is equivalent to 20 years at normal pressure; the Long-Baseline Neutrino Experiment Water Cherenkov Detector (LBNE-WCD) planned to employ the same protocol [185; p. 3-106]. Also the Hyper-Kamiokande (Hyper-K) experiment intends to place all PMTs in a high pressure tank for a few minutes to be qualified [41; p. 115].

In late 2020 GERDA emptied the water muon veto tank in preparation for the next phase of operation. In doing so, they found that several of the PMTs (ET 9350KB and 9354KB encapsulated in optical modules (OMs) filled with mineral oil [437]; design based on the Borexino outer detector OM) had imploded during the 10 years since their installation [438], although the pressure from 8.5 m maximum water column was well below the rated pressure capability of 0.2 MPa [437]. One possible cause for a decrease of the pressure resistance with time is glass corrosion (discussed on p. 441). The matter is presently being investigated.

Dependences

The failure pressure depends on the shape of window, shaft and stem, the PMT diameter, the glass thickness and its material [169; p. 266].
Plano–plano windows are most sensitive and thus require high thicknesses but still show slight deformations in the window center already at atmospheric pressure [170; p. 84]. Plano–concave and convex–concave windows are more stable, and hemispherical windows are particularly stable due to their shape [169] [170; p. 59].
For large PMTs the hemispherical windows itself typically is very stable and seldom breaks. The weak points of the glass bulb are the points of largest curvature, which are located in the rather abrupt transition regions between a) the elliptical window and the shaft (neck), b) between shaft and stem, and c) the 90° curve from the stem to the flat back where also the pin feed-throughs increase stresses. In pressure tests of [185; pp. 3-107–3-108] the Hamamatsu R7081 (10”) broke at the base surrounding

the pins, the ETE 9354 (8") and the Photonis XP1807 (12") at the neck. During development of the Hamamatsu R11780 (12"), R12860 (20"), R12850 (20")[173] and the ETE D784 (11") stress analyses were carried out to locate the critical regions and successively improve the shape, combined with pressure tests to find the regions of failure [185; pp. 3-101–3-125] [165, 186] [41; pp. 113–124]. This lead to the implementation of smoother curvatures and thicker glass at critical points. For the R12860 after improvements the most critical point was at the stem around the pins, which was not optimized but rather is foreseen to be covered by a hemispherical waterproof polyphenylene sulfide guard cover for protection in Hyper-K.

Smaller PMTs can usually withstand higher pressures since the glass thickness is similar independent of size. [169; p. 266]

Failure pressure was found to correlate with the thickness at the thinnest point [41; pp. 113–117]. Hyper-K thus plans to subject glass thickness to quality control and reject devices for too low values.

PMTs with quartz windows have less pressure resistance due to the graded seals [169].

The pressure at failure is expected to be lower for smaller loading rates, i.e. the rate of applying the pressure [185; p. 3-106]. Therefore it is recommended in pressure tests to determine the pressure at failure for varying loading rates to predict behavior in the detector.

Pressure capability values

In recent years, in response to the growing size of neutrino detectors the manufacturers have taken measures to increase the compressive strength of new large area PMTs, as discussed above.
Older series can tolerate pressures up to 0.2–0.7 MPa, while newer types reach 0.8 or even 1.25 MPa.
The older 8" ETE 9354 can withstand 0.202 MPa [263], whereas the newly developed 11" ETE D784 reaches 0.808 MPa [262].
The older Hamamatsu PMTs R3600-02 (20", venetian blind dynodes) and R7250 (20", box-and-line dynodes) can both tolerate 0.6 MPa, the 10" R7081 and the 8" R5912 withstand 0.7 MPa [212]. In [185; p. 3-106] for the R7081 the pressure rating given by Hamamatsu is quoted as ≈ 0.8 MPa and was verified to be essentially correct by pressure tests: Of 6 tubes none imploded below ≈ 1.0 MPa and several other PMTs were cycled up to ≈ 0.95 MPa several times without failure.
For newer Hamamatsu photosensors the 12" R11780 reaches 0.8 MPa [439], the 8" Hamamatsu HPD R12112 (p. 686) withstands 0.9 MPa [165] and the new 20" R12860 even 1.25 MPa [41] (≈ 0.9 MPa were estimated from simulations [326]). The 20" HPD R12850 is estimated to withstand even larger pressures [165; p. 28]. The R12860 uses borosilicate glass with ≈ 3 mm thickness and none of 50 samples subjected to 1.25 MPa suffered damages [440], which leads to a safety factor of 1.9 for a water depth of 60 m and 1.5 (as was used for Super-K) for 80 m [165]. Hyper-K still plans to house these PMTs in shock-damping covers, and for larger hydrostatic heads than in Hyper-K (60 m water) use of pressure encapsulations seems prudent, if only to reach a higher safety factor than the one used for Super-K.

[173] This is actually an hybrid avalanche photodiode; see p. 684.

2.7.3 Failure rate

Failure modes

Failure modes for PMTs are roughly separable into gradual failure and breakdown failure [169; p. 273], as well as electric failure of the external components necessary for operation (voltage divider, cabling, electronics).

Gradual failure

Gradual failure is the most common cause and occurs over time due to wear of the components from operation (even without signals, dark count rate (DCR) and dark current cause wear) or environmental conditions. This comprises loss of cathode sensitivity, gain reduction, dark current increase, and resistance decrease. In this case the PMT can still be operated, but with deteriorated properties, which possibly do not meet the requirements of an experiment anymore. Most often this is due to an increase of dark current or gain loss/variation [169]. Environmental conditions, such as exposure to excessive amounts of light during operation, storage/operation in high temperatures and humidities, or traces of He in the surrounding atmosphere, accelerate the deterioration.

Cathode sensitivity degrades due to the effects listed on p. 60.

Gain loss occurs due to a decreased $\delta 1$ from ion impact and contamination of the dynode surface [184; p. 231]. As a consequence, the high voltage (HV) has to be increased to maintain the desired gain. Usually the PMT operating life is measured by the accumulated anode charge after which the gain drops by a factor of two, which lies in the order of 200 C according to [170; p. 573] and between 300 and 1000 C in [172]. When this is reached depends on the accumulated photon flux and the DCR. For anode currents below a few µA, the total output charge is not the decisive factor for PMT lifetime and He migration and internal diffusion dominate [172].

Helium can permeate the PMT glass and increase the gas pressure and the ionic afterpulse (iAP) rate (see p. 444). For pressures exceeding 10^{-3} torr electrical breakdown is likely, rendering the device inoperable [171; p. 22]. The He content in the atmosphere is sufficient to reduce the lifetime of PMTs with quartz windows to a few years; exposure to higher concentrations near He dewars – even during storage of the PMT – can cause this far faster for any PMT [170; p. 448].

In internal reliability tests Hamamatsu uses as failure criteria a variation of the gain by more than 50%, a change of the cathode sensitivity by more than 25%, and an increase of the dark current by a factor of 10, among others [169; p. 275].

Breakdown failure

This can occur due to applying too high voltages (breakdown regime), exposure to high light levels during operation, loss of vacuum through microscopic cracks in the glass or the pin feed-throughs, and implosion [169; p. 473] [170; p. 535]. This type of failure makes the PMT permanently unusable – and in case of implosion possibly many others around it (see 2.7.2).

The failure criteria of internal reliability test of Hamamatsu comprise i.a. discharges and cracks [169].

Cracks in the glass can be caused by strong shocks (dropping, impact), vibration, pressure (see 2.7.2) or low temperatures [169]. Leakage of gas or vapor (see p. 417) then can destroy the device. Excessive stress on the pins can also cause leakage [169] and the pins are not watertight under high pressure [185; p. 3-101]. Excessive shock or vibration can also mechanically damage internal structures, and lose material produced hereby can cause short circuits [171; p. 20].

Ultrapure liquids, such as ultrapure deionized water or scintillator after multiple purifications, are very corrosive due to their low ion concentration [185; pp. 3-101–3-106]. Furthermore, glass is subject to stress-induced corrosion under pressure, where the main mechanism is considered to be based on small surface cracks, which concentrate stress and deepen with time, until the concentrated stress exceeds material strength and the glass envelope breaks.
Thus the hydrolytic properties of the PMT glass (or the materials of an enclosure surrounding it) over the foreseen operation time of an experiment need to be understood. In the BaBar experiment, an incorrect composition of the used ultra-low background glass B53 caused the PMT windows to become milky or frosty after several years of immersion due to dissolution in the water [185,441]. While this did not impact on-site PDE due to the good optical coupling of water and no failure occurred due to the small pressure, it shows that great care has to be taken when selecting the PMT glass, and material compatibility tests are advisable. The LBNE-WCD group has identified specific industrial tests (ISO-719 and ISO-720) for the hydrolytic capabilities of PMT glass [185]. For Schott glass these are well-known and can be calculated from existing data, from which follows that Hamamatsu and ETE glasses according to ISO-719 belong to the hydrolytic class 1, which has the lowest dissolution rate.

Exposure to excessive light during operation, usually due to human error, can deteriorate performance (see above) or destroy the PMT in more severe cases (see pp. 60 and 349).

Voltages exceeding the operational range can also destroy PMT (see pp. 408 and 418) or voltage divider. Visible light emission from the anode region can warn of the beginning of breakdown if the user is in a position to observe it [170; p. 265].

Electric failure

Apart from the PMT, its base, which contains the voltage divider, the cables from the base or the electronic channel (amplifier, discriminator, readout electronics) can fail [185; p. 3-119]. The processing electronics is usually placed outside of a detector and can be replaced, but the former two are often not accessible without interrupting operation and performing maintenance work.

Definitions [174]

The failure rate denotes the probability of failure per time. Field data for Hamamatsu PMTs show failure rates between $2 \cdot 10^{-6}$ and $2 \cdot 10^{-7}$ failures/hour at room temperature, regular voltage and for 100 nA output current; screening can lower this to $5 \cdot 10^{-7}$ failures/hour.

[174] Content based on [169; pp. 273–275]

The mean life (also called mean time between failures or mean time to failure) is the average time until failure occurs. For an exponential failure distribution, the mean life τ is connected to the failure rate λ through

$$\tau = \frac{1}{\lambda} \tag{2.226}$$

For the above numbers this results in $5 \cdot 10^5$ to $5 \cdot 10^6$ h (57–571 a) and for tested PMTs $2 \cdot 10^6$ h (228 a).
The Weibull distribution is often used to describe mean life, and contains the exponential decay as a special case.

The reliability R shows the fraction of operable devices after the operation time t and is defined as

$$R(t) = \exp(-t\,\lambda) \tag{2.227}$$

This formula can be used to estimate the reliability level of PMTs. Using the values from above after 1 year 98.3–99.8% of PMTs remain functional, after 10 years 83.9–98.3%, and after 30 years 59.1–94.9%. The great variation for the later two time scales shows the importance of minimizing failure rate for long term experiments.
Conversely, one can deduce the required λ from(2.227) using a given detector operation time t and a target reliability R. For $t = 30$ a and $R = 0.9$ follows $\lambda = 4.01 \cdot 10^{-7}$ failures per hour. This is close to the value for screened PMTs, which would lead to $R = 87.7\%$ after 30 a.

Early failure rate (EFR) denotes the fraction of devices which fail either before installation or within a short period afterwards. If this occurs still during construction of the detector, it allows to replace the PMT before operation and thus is preferable to later failures.

Failure rate in experiments

For the ET 9351, Borexino observed an EFR of 268 PMTs (2212 mounted in detector) in the inner detector (ID, filled with pseudocumene (PC)) in January 2007 [442, 443]. The subsequent number of failures follows roughly a linear curve until 2010-04 with about 1.7 failures/month, but then the slope *increased* to ca. 6.3 failures/month in the period from 2011-03 to 2019-12, with 1226 PMTs remaining functional at the end after repair of several channels [444]. From (2.227) the total number of failures would be expected to follow the function $1 - \exp(-t\lambda)$. One cause of failure was suspected to originate from water entering the base and voltage divider [414, 445]. A delayed onset of failures due to processes such as diffusion can change the curve shape. In the Borexino ID, 767 failures (34.7%) were observed in 13.5 a plus 12.1% EFR, which results in 46.8% total failures (reliability 53.2%) and leads to an initial (linear) failure rate of $1.04 \cdot 10^{-6}$ failures/h and a raised, later mortality of $3.89 \cdot 10^{-6}$ failures/h. The first value falls within the range quoted by Hamamatsu, which likely is predominantly determined by a subset of gradual failure modes, but the latter lies above it. This points to additional causes of failure (gradual, breakdown, electric) occurring with delayed onset, as suspected by Borexino and visible from the deviations from the theoretical curve. This has to be taken into account when planning an experiment by eliminating all known and conceivable problems.

In the outer detector (OD, filled with water, PMTs placed in OMs filled with mineral oil), an EFR of 19 out of 208 PMTs was found (9.1%) and 21 additional failures (10.1%) occurred up to 2019-12, which results in 19.2% total failures and a reliability of 80.8% – significantly higher than for the ID. The PMTs used for ID and OD are from the same manufacturer, series and production batch and should be comparable in every aspect. In the OD the initial mortality was found to again follow a linear curve up to including 2015 with about 1.1 failures/a, which is followed by an increased linear or exponential failure rate of ca. 3.0 failures/a (assuming linear curve). This results in an initial failure rate of $6.17 \cdot 10^{-7}$ failures/h and a later rate of $1.65 \cdot 10^{-6}$ failures/h. Despite the more critical environment (water) both values are considerably better than in the ID, the increase of slope is less severe, and the onset of the raised mortality is later.
Using eq. (2.226) the mean life for long term operation with and without OMs can be estimated by plugging in the final failure rates and neglecting the initial lower mortality. This results in 29.3 a without and 69.2 a with OMs. According to eq. (2.227) and using the same approximation, after 30 years a fraction of 36.0% of unprotected PMTs and 64.8% of housed PMTs survive.
All of these observations clearly demonstrate that the use of OMs can help to reduce the failure rate and is indispensable for long operation times.

Daya Bay "burned in" the ca. 2500 Hamamatsu R5912 PMTs after they had passed visual inspection for 3 days through operation at a gain of 10^7 to reduce EFR. 0.5% of the PMTs failed this test and an additional 1% did not meet requirement in the subsequent electronic tests.
All experiments involving large detectors screen their photosensors before installation anyway, so if a burn-in phase before testing can help to raise EFR before installation and reduce the number of later failures, this procedure is highly recommendable.

2.7.4 Other environmental factors

There are several other environmental factors affecting PMT performance (see [169–172]). Most relevant here are the influence of temperature, intense light, external radioactivity, the physical and chemical reactions with a surrounding liquid, and helium in the atmosphere. Their impact on sensor performance was addressed in the sections covering the respective properties (see summary below).

Temperature

Temperature affects dark count rate and dark current (pp. 56, 348 and 114), quantum efficiency (p. 59), gain (p. 139), linearity (at low temperatures, p. 143), light emission (ruby emission p. 423, Corona discharges p. 417), failure rate (gradual failure p. 440; for low temperature also breakdown failure, p. 440), and in principle also timing, time shift effects and afterpulses through their dependence on gain, quantum efficiency and light emission. In addition, the voltage divider can raise the PMT temperature (p. 146), and for characterization the temperature stability is important for light source and electronics (p. 334).
Its far-reaching influences make temperature one of the most important environmental factors. For large neutrino experiments physical reasons (convection) and

cost constraints (cooling the whole detector) often prohibit setting a desirably low temperature for the PMTs (dark count rate, gain), but in lieu of that temperature stabilization is highly recommendable.

Light

Illumination with intense light (e.g. ambient light) during operation can destroy the PMT or deteriorate performance (section 2.7.3, pp. 60, 349, 408 and 418).

Radioactivity

External radioactivity (see also 2.6.2) affects dark count rate (2.5.1, p. 112), the charge distribution (pp. 98 and 115), light emission (2.6.1) and cathode lifetime (p. 60).

Chemical compatibility

The compatibility with surrounding liquids was discussed in 2.7.3.

Helium

Ingress of helium through the PMT glass affects the ionic afterpulse rate (pp. 403 and 386), light emission (p. 423) and gradual failure rate (2.7.3, p. 60).

3 Photomultiplier characterization and selection

In the preceding chapters, first the photodetection requirements of a large liquid scintillator neutrino detector (LSND) were established and listed for the example of LENA [1] in 1.4; they are similar for other planned LSNDs. Photomultipliers (PMTs) can fulfill these requirements, have many favorable properties, and have been used in all large low energy neutrino detectors so far, which is why they were selected as baseline photosensors. Therefore, in ch. 2 the theoretical and practical background knowledge necessary to understand and measure crucial photosensor properties was discussed and extended with a focus on PMTs. The contents and results are in large part applicable to other photosensors as well.

The present chapter covers the selection of the best-suited PMT series for a future large LSND based on the criteria listed in 1.4.2.
First, in 3.1 promising candidate series are identified by means of an extensive literature research.
For some of these series crucial properties were not known, which required to carry out measurements which are described in 3.2. The setups and methods employed for this purpose are detailed in 3.2.1. Initially, an existing photosensor measurement setup of Borexino was used to characterize PMT series (see p. 457) besides a newly constructed setup at the Technische Universität München (TUM). In the course of these studies, the importance of a better understanding of the timing, early afterpulse (EAP) and light emission properties of PMTs became increasingly clear. This demanded the further development of the TUM experiment to an improved shielded photosensor analysis setup (SPAX, see p. 460), which was used in the later part of this work for photosensor research. The findings which are of general interest to the theoretical understanding of PMT behavior were in large part already discussed in ch. 2. The results of the series characterization are presented in 3.2.2, which also includes the effects of varying voltage on PMT properties and the variations between different series.
Finally, based on the results of literature analysis and measurements, in 3.3 the PMT series which are most qualified for future LSNDs are selected.

3.1 Preselection of promising candidates

Criteria

Combining the photosensor requirements of LENA listed in table 1.1 and the in-depth discussion of PMT properties in ch. 2, the following desirable properties for PMT series can be identified:

[1] Low Energy Neutrino Astronomy

- **Cathode material:** A bialkali photocathode should be used. Rb-bialkali has a four times higher dark count rate (DCR) and GaAsP is not available for large cathodes, degrades over time and has enormous DCRs.

- **Quantum efficiency variant:** High quantum efficiency (HQE) cathodes increase the photodetection efficiency (PDE), but show higher DCRs and ionic afterpulse (iAP) rates. They also have a somewhat higher transit time spread (TTS), since the quantum efficiency (QE) in the red is higher than for standard quantum efficiency (SQE) cathodes. This introduces a larger variation of the photoelectron (pe) emission energy. Since almost no photons are expected above 500 nm in LENA [446], this should play no large role. However, if the DCR of HQE photocathodes is too high, SQE cathodes have to be used instead.

- **Size:** Large PMTs are beneficial, since the radioactive background – which is caused to a large extent by PMTs – is likely reduced by using less channels. This is because the PMT weight scales approximately with the active area (using values from [447]) but for each sensor an optical module (OM) and cable(s) are required. For small sensors this increases the radiological load from materials and surfaces.
  Furthermore, larger PMTs are more cost-efficient per sensitive area, and using less channels also reduces costs (OM, electronics, cables).

- **Shape:** A hemispherical window improves timing (due to cathode time spread (CTS)), QE and collection efficiency (CE).

- **Window surface:** Sandblasting or applying a matt coating on the window might improve QE, but it would be necessary to check whether this also applies to HQE cathodes.

- **Glass:** Ultra-low background borosilicate glass with radiopurity comparable to the PMTs installed in Borexino should be used. In order to introduce less background and reduce absorption, the glass should only be as thick as necessary to withstand the expectable pressure in an OM.

- **Dynode type:** A dynode structure providing good timing and CE must be used, i.e. box-and-line or linear focused. Circular cage, fine mesh and metal channel dynodes are not available for large PMTs (the latter two are proximity focused, which is not possible for hemispherical PMTs). Venetian blind and box-and-grid are excluded.

- **Dynode material:** BeCu would cause less backscattering than Cs_3Sb, which is beneficial for the multiplication efficiency (ME) and reduces time shift effects (TSEs; for example E1 and L1, see 2.3.5). Using a K_2CsSb first dynode instead of Cs_3Sb would slightly reduce backscattering as well. Also, BeCu dynodes appear to have a slightly better linearity. On the other hand, Cs_3Sb has a higher secondary emission yield (SEY), and thus requires less stages to achieve the same gain. This decreases the TTS, rise time and fall time, and increases the pulse height, which leads to a higher signal-to-noise ratio (S/N) vs. baseline noise. Applying a higher voltage for BeCu dynodes can compensate these factors,

but is expected to increase light emission and iAP rates and introduces higher running costs.

- **Number of dynodes:** The number of stages should not be too low, since this results in either small gain or – when using high voltages – large DCRs and afterpulse (AP) rates. Neither should it be too high, as a low interstage voltage leads to poor linearity and TTS, and the pulse width increases with the number of stages. Using about 10–14 dynodes is typically best.

- **Gain:** A high multiplier gain shifts many underamplified pulses above the threshold, which improves PDE. However, this comes at the cost of worsening timing, since additional TSEs are detected. A high amplification also makes the pulse shape more compact. A low gain, on the other hand, reduces DCR, iAP rate, EAP rate and light emission. Using a value around $1 \cdot 10^7$ usually produces a good S/N.

- **Gain of first dynode:** $\delta 1$ should be as high as possible (e.g. using K_2CsSb; BeO is excluded), because this improves charge resolution [2] and timing [3]. In addition, a high $\delta 1$ reduces the contribution of TSEs skipping one or more gain steps. This allows to set a lower threshold and improves the TTS, transit time distribution (TTD), peak-to-valley ratio (P/V), and charge resolution and thus enhances photon counting capabilities. However, then the PDE is slightly reduced, since more underamplified pulses are lost.

- **Voltage divider:** The voltage divider (VD) should be front and back tapered to improve linearity, timing [4] and CE (smaller spot size). This also reduces the amount of detectable TSEs (higher $\delta 1$) and the susceptibility to magnetic fields. On the other hand, a tapered divider has a lower gain, which requires to increase the high voltage (HV). The divider current I_d must be large enough to ensure linear behavior for the largest expected counting rate. It should, however, not be much larger, since a higher current increases power consumption, which introduces heat into the detector, raises the DCR, and occasions costs. In the last stages decoupling capacitors and damping resistors should be used. Alternatively, a VD with Zener diodes or an active divider could be employed.

- **Voltage:** A large voltage improves the CE, timing [5] and pulse shape[6] but increases the DCR, TSE probabilities, EAP rate and iAP rate.

- **Polarity:** The HV polarity should be positive, since the cathode is in contact with scintillator. A positive voltage also provides a lower DCR and avoids the risk of cathode decomposition.

- **Threshold:** A large enough hardware threshold should be chosen to exclude most TSEs, EAPs, and interdynode ionic afterpulses but low enough to not reduce the PDE too strongly.

[2] Through the excess noise factor (ENF); $\delta 1$ has the largest influence.
[3] $\delta 1$ affects all contributions to the TTS from dynodes.
[4] Reduced spot size on the first dynode, lower variance of tk1 from CTS and emission energy and angle.
[5] Smaller TTS; reduced earliness, lateness and delay of TSEs.
[6] Smaller rise time, fall time, and pulse width.

- **Transit time spread:** The TTS must be as low as possible and the TTD should be as devoid of TSEs as feasible.

- **Pulse shape:** A small rise time is required. Fall time and pulse width should also be small.

- **Dark count rate:** Obtaining a very low DCR is crucial. Fortunately, this quantity improves greatly after long times in the dark, as is the case during operation of an LSND.

- **Afterpulses:** All types of afterpulse rates should be as low as possible, since they cause (un)correlated background. To obtain low iAP rates, a good initial vacuum, the use of getters, and operation of the dynode mount as ion trap is beneficial. EAP occurrence increases the fall time and pulse width. To reduce the EAP rate, backscattering must be minimized (see above) and light feedback must be prevented (see below).

- **Light emission:** This causes crosstalk in the detector [7], EAPs, higher DCRs, and possibly late afterpulses (LAPs) and thus must be avoided. For this reason, light shields around the dynodes and between dynode stack and cathode should be used (external paths). Furthermore, the dynode layout should be designed to frustrate internal photon reflection and direct transmission from the anode to the second dynode [8] (internal paths). A dynode mount acting as ion trap also lowers the light emission from gas constituents.

- **Flashers:** In order to prevent the occurrence of flashers, one should employ OMs, watertight cables, intransparent casting slip, and use only materials in the PMT base which were tested for electric discharges.

- **Magnetic shield:** Use of a passive magnetic shield is required to avoid the deterioration of the CE, gain, resolution, P/V and timing.

- **Pressure housing:** Furthermore, the PMTs must be housed in pressure encapsulations.

- **Failure rate:** The mean life must be sufficiently high. A good vacuum reduces the gain decrease from ion impact. To lower the corrosion from contact with ultrapure liquids, closed OMs can be used. A burn-in phase during testing helps to increase the early failure rate. In addition, it is important that the cables and OMs are watertight.

- **Temperature:** The ambient temperature should be as low as possible, as this results in a smaller DCR, improved QE, higher gain, and possibly longer cathode lifetimes.

- **Radiopurity:** Like for the PMT glass (see above), low background materials must be used for all components of the PMT (e.g. the ceramic dynode mount), VD and OM.

[7] Random and correlated background; using light concentrators and OMs reduces this.
[8] In the following, dn denotes the n-th dynode.

Data

Based on these considerations, promising PMT series which are commercially available or currently being developed were identified. For these series previously measured properties were compiled using data of manufacturers and independent groups.
The results of this literature research are summarized in table 3.1 and elucidated in the following.

The included series comprise fast large PMTs with diameters from 5” to 20” from Hamamatsu [9] and ETE [10] with enhanced (HQE variants) or standard QE. At present, no other manufacturer offers PMTs in this size category which meet the demanding requirements of LSNDs.
The Hamamatsu R6091 (3”) was originally included as candidate and a sample was begun to be characterized in this work. However, it was found that such small PMT sizes do not conform with radiopurity demands and funding considerations, so the R6091 was excluded.
For the Hamamatsu R7081 (10”) also a multianode variant with 4 anodes exists (see p. 608 and 4.1.2.2) [448, 449], which was developed for NEMO and makes this an image sensor, unlike regular large PMTs.
Since for the Hamamatsu R12860-HQE (20”) only a HQE variant is produced, it is designated simply R12860 from now on.

The CE of the R12860 was obtained by Hamamatsu through simulations – not measurements – for a reduced (46 cm) and the full cathode diameter (50 cm); photoelectrons from the brink of the cathode suffer most from collection losses [178, 279].

Where the TTS was listed as σ_T (see 2.3.3), this was converted to T_{FWHM} with eq. (2.68) to enable the comparison of series. For the R12860 the TTS was measured without timewalk correction (see p. 335) [236], so the actual TTS is better than the listed value – which for such a large sensor already is excellent.

Early pulse (EP) and late pulse (LP) measurements depend on the used definition (see p. 255), the time range for which such effects are registered, and the used threshold. For the measurements of the ET [11] 9351 (8”), EPs and LPs were defined as all pulses earlier or later than $3\,\sigma_T$ from the main peak (earliness up to 20 ns, lateness up to 100 ns) and the threshold was 0.17 pe [12] [336, 450]. In the studies of [309] for the ETE 9354 (8”) and the Hamamatsu R5912 (8”), a threshold of 0.25 pe and a mean number of detected photons of $\lambda = 50$ pe was used. For the R7081 a low threshold of 0.1 pe was used, but here in the EP rate only $\gamma 1$ (see p. 282) was included [301]. For the Hamamatsu R11780 (12”) the EP rate was measured by Hamamatsu for earliness values between 10 and 80 ns at 0.25 pe threshold, which only gives a lower limit, since most EP types are excluded by this window. The LP rate was determined for a lateness between 15 and 100 ns and the same threshold [439]. For the measurement

[9] Hamamatsu Photonics K. K.
[10] Electron Tubes Enterprises Ltd.
[11] Electron Tubes Ltd. (ET)
[12] The unit pe denotes the mean height or charge of a single photon pulse or the mean number of pulses produced in response to a light signal.

Table 3.1: Properties of currently available PMT series with large diameters, bialkali cathodes, borosilicate glass and favorable properties.
Series names starting with R are from Hamamatsu Photonics K. K., all others from Electron Tubes Enterprises Ltd.. Two excluded series are listed for comparison: The R6091 is too small but allows to study the dependence of properties on diameter, and the 9351 is out of production but well studied. All PMTs have hemispherical windows except for the R6091 (plano-concave window). Box-and-line (B+L) or linear focused (LF) dynodes are used in all series.
Values are listed if known, numbers from manufacturers are italicized, values which do not fulfill requirements are in red (see table 1.1; the DCR is colored red if it is higher than 3 times the required value). The number of tested samples is listed for the most significant publications from independent groups; PMTs appearing in several publications of the same collaboration are listed only once. Magnetic field tolerance entries marked with * are detailed in the text. The table is continued on the next page.
Sources: R6091 [189, 213], R6594 [213], 9351 [80, 170, 275, 336, 451], 9354 [201, 263, 309, 342], R5912 [212, 213, 323, 335, 389], R5912 HQE [212, 213, 276, 342, 428, 451], R7081 [160,161,212,213,261,292,301,342,431,452], R7081 HQE [212,213,261,342,452,453], D784 [186,262,306,454], R11780 [190,261,306,439], R11780 HQE [190,261,306,455], R12860 HQE [41, 165, 178, 213, 225, 236, 251, 279, 307, 326, 346, 454, 456–459].

PMT series	**(R6091)**	**R6594**	**(9351)**	**9354**	**R5912**	**R5912 HQE**	**R7081**	**R7081 HQE**	**D784**	**R11780**	**R11780 HQE**	**R12860 HQE**
Number of samples			2200, 80	4, 1	1625, 324, 1	255, 110, 1	474, 1	1, 1	15	7	10	4620, 145, 1, *260*
Diameter window [cm (inch)]	*7.6* (3)	*12.8* (5)	≈ 20.3 (8)	*20* (8)	*20.2* (8)	*20.2* (8)	*25.3* (10)	*25.3* (10)	*28* (11)	*30.5* (12)	*30.5* (12)	*50.8* (20)
Diam. cathode [cm]	*6.5*	*11*		*19*	*> 19*	*> 19*	*> 22*	*> 22*	*27*	*> 28*	*> 28*	*> 46*
Dynodes	12 LF	10 B+L	12 LF BeCu	12 LF Cs_3Sb	10 B+L	10 B+L	10 B+L Cs_3Sb	10 B+L Cs_3Sb	12 LF Cs_3Sb	10 B+L Cs_3Sb	10 B+L Cs_3Sb	10 B+L
PDE [%] at 420 nm							23.0 ± 1.4					28.1
QE [%] at 420 nm	*26*	*25*	27	23, *26.5*	≈ *25*	37 (408 nm), *35 (390 nm)*	27, *25 (390 nm)*	34.6, *35 (390 nm)*	≈ 26.5	*22*	≈ 32	≈ 28.5
CE [%]							*90*					*95/85* (46/50 cm)
TTS T_{FWHM} [ns]	*1.5*	*1.5*	2.76 ± 0.27	1.58, *2.7*	≈ 2.9 ± 0.3, *2.4*	2.60 ± 0.12, *2.4*	2.8 ± 0.2, *3.4*	2.67, *3.4*	4.66 ± 0.40, *3*	3.23 ± 0.35, *2.6*	3.04 ± 0.33	< 2.65 ± 0.15, *2.4*
Transit time T_0 [ns]	*40*	*45*	70	*65*	*54*	*54*	*62*	*62*	*62*	*70*		*95*
EP probability [%]			0.75 ± 0.35, $\gamma 1 > 0.012$	0.2–0.4	0.1–0.5		$\gamma 1 > 0.1$			*> 0.5*		0.05
LP probability [%]			7.90 ± 1.26			2.3 ± 0.05	3		4.51 ±0.74	4.48 ± 0.32, *4*	4.3 ± 0.35	2–3

PMT series	**(R6091)**	**R6594**	**(9351)**	**9354**	**R5912**	**R5912 HQE**	**R7081**	**R7081 HQE**	**D784**	**R11780**	**R11780 HQE**	**R12860 HQE**
Charge res. R_F [%]			106	82–111	71–111		59	67.4	90 ± 25	89 ± 25	103 ± 39	70.0 ± 7.9
P/V [1]			2.10 ± 0.39	2.4–3.2, *2*	3.1 (2.5 –4.3), *2.8*	*2.8*	3.2–5.5, *2.8*	4.1, *2.8*	2.32 ± 0.67, *2*	2.8 ± 0.28, *2.8*	2.24 ± 0.27	3.8, *2.5*
Pulse height [mV]												6.5
Gain g (max.) [10^7]	*1 (> 10)*	*0.2 (> 2)*	2	1 (*3*)	1 (*4*)	1 (*4*)	1 (*4*)	*1 (4)*	1 (*3*)	1	1	1 ($\approx$ *3.5*)
Linear range (dynamic range) [pe]							250–300					370 (> 3000)
Rise time t_r [ns]	*2.7*	*3.5*		3.1–3.4, *3.5*	3.1–3.2, *3.6*	*3.6*	*3.8*	*3.8*	*5*	*4*		6.9, *5.2*
Fall time t_f [ns]				4.1–4.4	3.5–4.6							10.2, *8*
Pulse width t_w [ns]				*4.5*					*6*			11.6, *11*
DCR [kHz]			2.48 ± 3.49	≈ 1, *4*	1.3, *4*	5.80 ± 0.78, *6*	2.2 ± 0.5, *7*	5, *8*	18.4, *8*	3.7 ± 5.1, *10*	4.4 ± 1.9, *15.5*	6.4 ± 1.9, 15.5, *18.6*
EAP rate [%]					3.4 ± 0.6	2.7 ± 0.01						
iAP rate [%]			4.94 ± 2.64	2.1–3.2, *5*	< 2, 2.2–2.5	7.1 ± 1.8	2.68 ± 1.46	16		*10*		5
Pressure capability [MPa]				*0.202*	*0.7*	*0.7*	*0.7*	*0.7*	*0.808*	*0.8*		1.25, *0.9*
Magnetic field tolerance [µT]				*, *40*	*		50		> 45*		32	
Radiopurity K_{nat} [ppm], ^{232}Th [ppb], ^{238}U [ppb] (Bq/kg)				*60* (4.0), *30* (0.3), *30* (1.0)	(0.6), (0.2), (0.9)				*60*, *30*, *30*			

of the R11780 and the R11780-HQE the LBNE-WCD [13] group counted all pulses with a lateness between $5\,\sigma_T$ and 80 ns as LPs, and for the ETE D784 (11") for up to 95 ns [190, 306].

In most sources, the charge resolution R_σ (see p. 119) was determined from the charge distribution (single electron response, SER; see 2.2.2) by fitting a Gaussian to the single photon (spe) peak or using the relative charge variance. This was converted to the R_F definition using eq. (2.68) to allow the comparison of all values.

The dark count rate depends strongly on the time in the dark, threshold and temperature. Unfortunately, most publications do not list either of these values. For the 9351 the DCR was measured after 3–70 h in dark and at 0.2 pe threshold [336, 450]. For the 9354 [309] measured a rate of 1 kHz after 15 h settling time. [160] examined the R7081 after 20 h at 20 °C and found that the DCR increased with rising temperature by 100 Hz/°C. Hamamatsu DCR numbers were obtained after 15 h and with 0.25 pe threshold, except for the R11780 (24 h) [439] and the R12860 (2 d, 25 °C) [178]. For the latter, JUNO (Jiangmen Underground Neutrino Observatory) measured a DCR of 15. 5kHz for 4680 PMTs (unknown parameters) [307], whereas Hyper-Kamiokande found that the DCR dropped to 6.35 ± 1.93 kHz for 60 PMTs at room temperature after many days in the dark [346].

The EAP rate also depends on the acquisition range and threshold. For the R5912 delays between 8 and 60 ns and a threshold of about 0.1 pe were used by [335], for its HQE variant 0–100 ns and also about 0.1 pe threshold [276].

For iAP rates the same considerations as for EAPs apply. The 9351 was characterized with spe light, 0.17 pe threshold and a delay range from 0.4 to 12.8 µs [336]. In [309] the 9354 and R5912 were measured with single photons, 0.25 pe threshold and delays up to 24 µs, while ETE used 0.4–12.8 µs for the 9354. Daya Bay [389] found an iAP rate of $< 2\%$ for 1625 R5912 PMTs and spe intensity, while in the PMT series selection for Daya Bay [309] still rates between 2.2 and 2.5% were observed for 7 samples. The R5912-HQE was measured between 0.1 and 10 µs [428] with an occupancy (see p. 89) of $o = 5$–10%, the R7081 within 0.1–16 µs and with $o = 10\%$ [292] and the R7081-HQE from 0.5 to 13 µs [453]. Hamamatsu determined a typical iAP rate of 10% (max. 20%) between 0.1 and 16 µs at 0.25 pe threshold and spe illumination. [279] also used single photons for the R12860 measurement according to [41]. All groups which specified the occupancy used spe illumination for the iAP characterization, which shows a strong preference over multi-pe acquisition with its additional errors.

The radiopurity of the R5912 is listed for the ultra-low background glass variant in the table; low background glass would have 5.1 Bq/kg ^{40}K, 1.3 Bq/kg ^{232}Th and 1.9 Bq/kg ^{238}U [309]. For the numbers in Bq/kg (all from [309]) the compliance with requirements could not be calculated, since they were measured using glass samples and the fraction of glass in the total PMT mass was unknown.

The magnetic field tolerance was defined as a drop of CE, gain or output to 50% of the value without fields. For the 9354, ETE states an output decrease of

[13] Long-Baseline Neutrino Experiment Water Cherenkov Detector

50% at 40 µT applied in the most sensitive direction [263]. On the other hand, [309] found a charge variation of up to 27% for 50 µT for the 9354 and up to 11% for the R5912 at the same field. At 50 µT in the most sensitive direction the R7081 showed a decrease of the P/V from 4.2 to 1.2, the gain decreased by 27% or increased by 7%, the CE decreased to 50% and the output dropped to 40% [431]. A reduction of the output to 50% occurred at about 40 µT. The D784 appears to be quite field-resistant, as for transverse fields of up to 45 µT only a PDE reduction of less than 10% was observed [306]. The R11780-HQE, on the other hand, shows a PDE decrease of 50% at 32 µT [190]. The properties measured by [261] for the R7081-HQE and the R11780 were not included in the table, since no magnetic compensation was used, but can allow to assess the influence of the magnetic field.

Despite a thorough literature search no values could be found for many important properties.

Comparison

Table 3.1 can now be used to draw a comparison between series.

High and standard quantum efficiency

For HQE variants the dynode structure is unaltered compared to the SQE type to the author's knowledge, so both are expected to show similar properties – with the exception of QE, DCR, EAP rate, iAP rate and possibly TTS.

For the R5912 a strong increase of DCR and iAP rate are observed with a HQE cathode. Interestingly, the EAP rate *decreases* for HQE although the acquisition window is longer, identical thresholds were used, and larger γnk rates are expected (see p. 379). However, the difference is almost within error bars. Also unexpectedly, the TTS is slightly better for HQE, but the discrepancy is within uncertainties. The P/V is identical within fluctuations and the rise time is slightly higher for HQE, but this also depends on the used VD.
In the HQE variant of the R7081 the DCR shows a moderate to large increase and a tremendous rise of the iAP rate is found. Again, the TTS is slightly better, but the increase is within error bars and only one sample was measured. The charge resolution is higher, although only within expectable fluctuations (confer other series); the P/V values match.
For the R11780 HQE version, the DCR increase is moderate to strong, the TTS is a bit smaller (within errors), the LP rate is identical considering uncertainties, the charge resolution is higher (but within the large variations), and the P/V is lower (within errors).

All in all, as expected, a marked rise of DCR and iAP rate is found. No notable increase or even a decrease of the EAP rate is observed. The TTS did not grow slightly as anticipated, but actually became somewhat smaller. However, the listed measurements were performed with wavelengths in the blue/UV, for which a TTS increase is not expected, and the use of higher voltages might explain the observed differences. Interestingly, a minor deterioration of charge properties (resolution, P/V) is observed for HQE cathodes, but this might originate from low statistics in some of the listed sources. All other properties were identical within errors.

Suitability for neutrino detectors

These candidate series shall now be compared to the LENA requirements to preselect models.
For the EP, LP, EAP and iAP rates the values of the requirements listed in table 1.1 are for a classic pulse-shape-agnostic analysis, as usually employed in data taking.

For the majority of PMTs no values are listed for PDE and CE, in which case a pulse efficiency (see 2.1.3) of 80% was assumed to estimate the PDE. The modest PDE requirements are fulfilled by all models except the R11780-SQE; for this series the HQE variant would be required.

The P/V is not known for the Hamamatsu R6594 (5”). All other PMTs fulfill the requirement, although for the D784 and R11780-HQE some samples were below the limit, which could raise the fraction of rejected devices.

A sufficient maximum gain was required for inclusion in the list and thus is fulfilled by all sensors.

The dynamic range is unknown for most series, but is expected to fulfill the requirements based on available values: The minimum dynamic range is 152 pe for the R7081 and 608 for the R12860, which is easily fulfilled by these sensors.

Regarding the TTS, for the independent measurement of the R5912 some samples exceed the limit, but the listed value and its error bars have some uncertainty since they were extracted graphically. The values from Hamamatsu for the R7081 and R7081-HQE are above the limit, but the independently measured values fulfill the demands. The value from ETE for the D784 is just within the acceptance region, but independent measurements show far higher numbers which regrettably excludes this recently developed sensor. The R11780 should comply with demands according to Hamamatsu, but in non-manufacturer measurements many sensors lie above the limit, which is similar for the R11780-HQE. Finally, the R12860 – the newest sensor on the list – fulfills the TTS requirements despite having the largest cathode.

The *total* EP rate (not only $\gamma 1$) is unknown for most sensors. The 9354 and R5912 seem to fulfill requirements, but [309] did not provide much detail about the measurement, so this remains uncertain.

Almost all series satisfy the LP rate limit. The value is unknown for the R6594, 9354, R5912 and R7081-HQE, but the latter two likely have similar numbers as the R5912-HQE and R7081, respectively.

The DCR criterion imposes severe limits. The value from table 1.1 translates to the following upper DCR limits: R6594 278 Hz, 9354 679 Hz, R5912 692 Hz, R7081 1086 Hz, D784 1330 Hz, R11780 1578 Hz, and R12860 4378 Hz.
Most DCR values in the table were determined after 15–48h in the dark. During operation in a detector the sensors will spend many weeks and months in the dark, which greatly lowers the DCR. For the 9351, Borexino found a decrease from 2.48 kHz in the PMT characterization (3–70 h in dark) [336] to about 800 Hz in the first days of detector operation (starting 2007-05-16, day 1) after commissioning [460]. The rate continued to decrease to a plateau around 400 Hz starting on about day 620, which extended to the end of the period studied in [460] (about day 1425)

with fluctuations to lower and higher values. As of July 2020, the DCR had decreased even further to about 200 Hz for the remaining PMTs [461]. This additional decrease by a factor of about 3–10 during commissioning and in the early operation phase can be considered when analyzing the listed values. As a conservative estimate, an expectable reduction by a factor of 3 was used for the comparison with requirements. The DCR is too high by a factor of 1.47 for the 9354, 1.88 for the R5912 (HQE 8.38), 2.03 for the R7081 (HQE 4.60), 13.83 for the D784, 2.34 for the R11780 (HQE 2.79), and 1.46 for the R12860. No value was available for the R6594. This excludes the D784 and possibly also the R5912-HQE and R7081-HQE; for the latter only one sample was measured, which does not allow a final statement at this point. The value of 6.4 kHz for the R12860 was obtained after many days in dark, so it is questionable whether a sufficient further decrease is possible. All other series should be able to fulfill the limit after longer times in dark.

The EAP rate requirement is fulfilled by the R5912 and its HQE variant, but unknown for the rest.

The iAP rate criterion is met by the 9354, R5912, R7081 and R12860, while the R5912-HQE, R7081-HQE (one sample only) and R11780 have too high rates.

Regarding radiopurity, it is assumed that a glass with comparable amounts of radioactive isotopes as the Borexino PMTs can be procured and used in the manufacturing. Such a glass would also meet the LENA requirements.

For the R12860 actually 50 samples withstood peak pressures of 1.25 MPa during tests without damage [165, 279]. This would make it the only PMT series capable of operation in LENA without pressure housings. However, failure due to pressure depends on the load rate and the PMTs were only subjected to such pressure for far shorter times than the foreseen 30 a of operation. Therefore, a factor of safety of at least 1.5 (Super-Kamiokande), better 2 should be employed. Hamamatsu states a far more conservative pressure capability of 0.9 MPa [178]. This means that in the end none of the listed series can withstand 1.2 MPa and all will have to be housed in pressure encapsulations.

To ensure sufficient operational properties in the presence of the geomagnetic field (GMF), a passive magnetic shielding is required for all series, since even a change of the PDE or gain by 10% is too large.

For a fraction of functional PMTs of 90% after 30 a of operation a failure rate of not higher than $4 \cdot 10^{-7}$ failures/h is allowed. This is within the range reached by Hamamatsu PMTs and very close to the number for screened Hamamatsu PMTs (see 2.7.3) and thus is considered possible for both manufacturers. It is of crucial importance to eliminate all risk factors which could artificially raise the failure rate and therefore a PMT housing should be used (see p. 442).

Preselection

The most severe requirements are imposed by DCR, TTS and iAP rate. It would have to be checked whether these requirements can be loosened by additional Monte Carlo detector simulations.

Based on the current limits and the data available in literature, the R11780-SQE, D784 and R5912-HQE are excluded as candidates. The R11780 does not meet the

PDE and iAP demands and has a TTS near the exclusion limit. The D784 shows too large TTS and DCR. And the R5912-HQE has a too high iAP rate and potentially a too large DCR.
However, also other series show properties which might not meet demands: The R11780-HQE will likely also be excluded by its iAP rate, which should be even higher than for the SQE variant. Apart from that, its TTS and DCR are border cases. The R7081-HQE might have too high DCR and iAP rate. And even the R12860 shows a DCR which could be problematic for some analyses.

If all models with potential problems are eliminated from the list, this leaves the R6594, 9354, R5912 and R7081.
Nevertheless, several requirements are not known for the last three PMTs, and there are only a few which *are* known for the R6594, of which none are from independent measurements. Furthermore, the cost optimum is expected for the largest sizes (i.e. 20") [165], which also is a non-negligible factor in the selection. Regarding the other types, the R11780-HQE is a good PMT except for its expected iAP rate. The R12860 even has *excellent* properties, where only the DCR might be problematic. Finally, the R7081-HQE should not be rashly eliminated from the short list either, since the value of its iAP rate is based on one sample only and for iAPs large fluctuations between devices are routinely observed.

The short list thus comprises the following PMT models (in order of rising diameter): R6594, 9354, R5912, R7081, R7081-HQE, R11780-HQE and R12860.

Since nearly all series are excluded by strict application of the current limits, this suggests to revise and loosen the criteria in coordination with simulations which assess their influence on the physics case.

3.2 Measurement of photomultiplier performance

Even these extensive literature studies could not produce values for all needed properties for each model.
For the short-listed sensors further measurements are required to obtain the following properties:

- **P/V**: R6594
- **TTS**: R6594, 9354[14] and R7081-HQE[14]
- **EP probability**: R6594, 9354, R5912, R7081, R7081-HQE, R11780-HQE, R12860
- **LP probability**: R6594, 9354, R5912, R7081-HQE
- **DCR**: R6594, R7081-HQE[14]
- **EAP rate**: R6594, 9354, R7081, R7081-HQE, R11780-HQE, R12860
- **iAP rate**: R6594, R7081-HQE[14], R11780-HQE

[14] Since the independently measured value in the table is based on one sample only.

For this reason, samples were procured and measured for all of the above series with the exception of the R12860 (not yet available at the time of measurements), R7081-HQE and 9354. Since this was done at an early stage of this work, most of the sources listed in table 3.1 were not published yet and from then available data the latter two were not identified as short list candidates. Instead, the R6091 (3”) was still included in the studies.

3.2.1 Setups

The bulk of the characterization was carried out with an existing setup of the Borexino collaboration at the Laboratori Nazionali del Gran Sasso (LNGS) in Italy.
In parallel, a new, improved setup for PMT analysis was developed at the TUM in Garching (SPAX). Based on the experience and results gained first at the LNGS and later in Garching, the TUM setup was iteratively upgraded and optimized for the analysis of timing, EAPs, iAPs, LAPs and light emission for exceptionally low thresholds. SPAX was used for the later PMT characterizations and for the dedicated analyses aimed at the theoretical understanding of PMT properties. The evolution of time shift theory (TST) and SPAX were closely intertwined and mutually dependent, since the theory was required to explain the observations, while the setup was needed to study the predictions of the theory.

Laboratori Nazionali del Gran Sasso

The Borexino collaboration developed a dedicated PMT testing setup at the LNGS in order to characterize the 2350 ET 9351 PMTs purchased for the inner and outer detector [80, 275, 336, 450]. This allowed to characterize up to 60 PMTs in one cycle in the dark room, with simultaneous data acquisition for 32 channels.
In the present work, this setup was reactivated to measure several candidate PMT series for large LSNDs in collaboration with Gyorgy Korga and with support from Oleg Smirnov, Laszlo Papp, Paolo Lombardi, and Gioacchino Ranucci. The measurement site had not been used for several years, which required to readjust the configuration by using ET 9351 PMTs with known properties. Several repairs had to be carried out and some improvements could be made. The latter comprised placing a low-noise preamplifier (fixed gain of 55) near the PMT inside the dark room, which improved the P/V by a factor of two. Furthermore, a second channel was used to measure the first-AP delay distribution (DD) of EAPs down to delays of 25–30 ns.

Layout

The final setup is shown in fig. 3.1. The original setup was based on a majority logic unit (LeCroy 4532) to improve acquisition speed [450]. In the configuration used here, this effectively is equivalent to a coincidence between the laser trigger output and a PMT signal and is thus shown in simplified form in the schematic (fig. 3.1a). Since the analysis of EAPs was only tried near the end of the measurements, some problems with crosstalk between channels could not be resolved and no EAP DD without background could be produced. Thus, the second channel is not depicted.

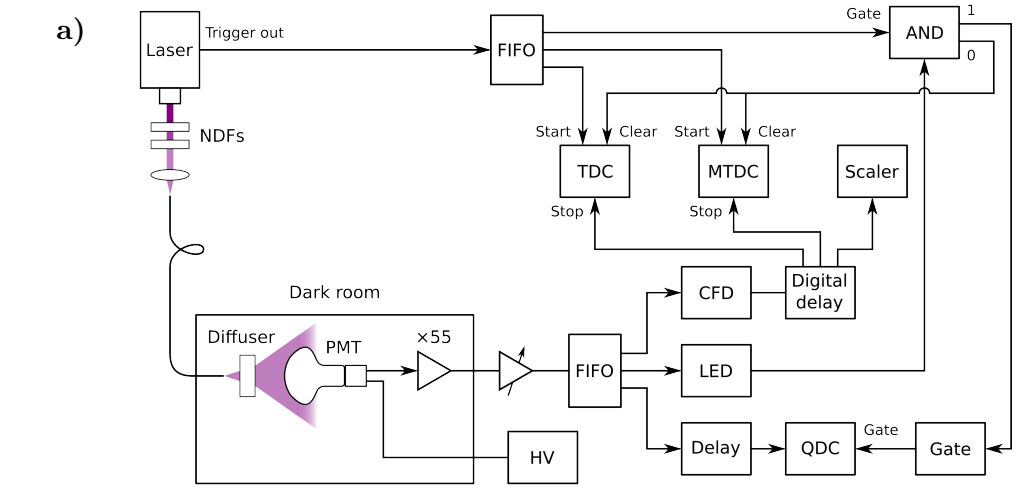


Figure 3.1: PMT characterization setup at the Laboratori Nazionali del Gran Sasso (LNGS), based on the Borexino PMT testing facility.
a) Schematic layout of the improved setup, description see text.
b) Photo of the electronics room. The entrance to the dark room can be seen on the right.
c) Picture of the table holding the PMTs. The sensors are placed facing upward in holes with centering rings. Up to 64 PMTs can be accommodated on the table, which is separated into four sections by black shrouds to suppress reflections. The whole table is surrounded by field-compensating coils for the two main components of the local geomagnetic field (GMF) (partially visible at the lower left). The diffuser is placed at the ceiling and illuminates 16 measuring sites. The depicted PMTs are (from left to right) a Hamamatsu R7081, R5912 and R6091.

The measurement is based on the start–stop method described in 2.3.8 (fig. 2.48a). The setup can record the TTD with a time-to-digital converter (TDC, ca. 120 ns window), the iAP DD (32 µs) with a multi-hit time-to-digital converter (MTDC), the SER using a charge-to-digital converter (QDC) and the DCR through a scaler. The occupancy can also be measured using the fraction of next laser pulses which is visible in the iAP DD, since the laser period was set to 30 µs so that the next peak is still contained in the DD. The occupancy is then simply $o = N_2/N_1$ with N_1 the number of detected prompt pulses from the initial light pulse and N_2 the amount of subsequent light pulses (compare eq. (2.10)).

Premises

The setup is located in two adjacent rooms, a dark room which can house 64 PMTs on four tables surrounded by field-compensating coils, and a separate room for the light source and electronics, which allows adjustment of measurement properties without access to the dark room.

Light source and optics

The ps-laser and optics are mounted on an optical bench in a small dark box. As light source a Hamamatsu PLP-02 laser with 410 nm wavelength and 30 ps pulse width is used (laser diode head SLDH-041, pulser C4725). The trigger output jitter is < 10 ps with respect to the light pulse [462] and the repetition rate was set to 33 kHz (30 µs period). The emitted light is attenuated by exchangeable neutral-density filters (NDFs), coupled into a 12 m long fiber which enters the dark room, and is fanned out by a diffusing beam probe with a transmissive diffuser.
The PMTs are fixated and centered by their own weight in circular holes (with aperture rings for smaller diameters) in a wooden table with 16 holes. The cathode is facing upwards and has a distance of > 1.5 m to the diffuser. The four tables are separated by black shrouds to reduce crosstalk from reflections, but here only one table with one PMT in the center was used at the same time in order to eliminate this effect and optimize the homogeneity of illumination.

Field compensation

The GMF is compensated with rectangular coils for its two dominant components (Merritt coil system with four square coils for the main component, two square Helmholtz coils for secondary direction). This results in measured residual fields of < 8.5 µT (uncompensated component) and < 1.5 µT (Merritt coils) perpendicular to the PMT rotational axis and < 5 µT along the axis (two-coil system) with a non-uniformity in the table plane of $\leq 10\%$. From the small remaining fields no strong influence on PMT properties is expected.

Electronics

Most components of the electronics (discriminators, logic, digital delay, logic, all data acquisition components, voltage supply via CAEN C117B controller) are CAMAC-controlled by a computer (CAEN C111 interface). With this system acquisition rates of up to 1000 events/s (depending on occupancy) are possible.
The PMT is supplied with HV by a CAEN mainframe SY403 with a modified A505 board (16 channels) or a Bertan 205A-03R, which both display low ripple and allow to check the current on each channel.

The PMT output is preamplified by a low-noise amplifier (gain 55, self-built by G. Korga) in the dark room and with variable gain (Phillips Scientific 777) outside the dark box. The signal is fed through a fan-in fan-out (FIFO) (Phillips Scientific 740) to a leading-edge discriminator (LED) with 0.1 pe threshold (LeCroy 4413). The discriminator is connected to a coincidence (majority logic unit LeCroy 4532, labeled AND in the diagram) with the laser trigger, which inhibits acquisition if no PMT pulse is observed.
One output of the signal FIFO is used for charge acquisition with a QDC (LeCroy 2249A) and is delayed to wait until the processing time of the coincidence branch has passed, since the coincidence opens a gate for charge integration.
The laser trigger is also fanned out by a FIFO and starts both a TDC (CAEN C414) to record the TTD and an MTDC (LeCroy 2277) for the iAP delay distribution. The third output of the signal FIFO is connected to a constant fraction discriminator (CFD; ORTEC CCF8200, 0.2 pe threshold). The logical output of the CFD is delayed [15] digitally (LeCroy 4518/300), stops the TDC and MTDC, and as is counted by a scaler (LeCroy 4434) to obtain the DCR.
The electronics jitter was $\approx 45-90$ ps and the upper limit for the time of flight difference from the diffuser to the cathode brink and center ranged between 3.4 ps (R6594) and 13.4 ps (R7081). Combined with the width of the light pulse (< 30 ps) and the trigger jitter (< 10ps) this results in a maximum total setup jitter between 48.6–92.7 ps (statistic addition through convolution of Gaussians) and 73.3–128.0 ps (systematic, linear addition). As a conservative estimate, the linear addition using the maximum value of the electronics jitter was assumed as the time resolution of the setup. This amounts to 118.3–128.0 ps depending on PMT diameter.
The CAMAC control and data acquisition was carried out by an MS-DOS based software run on an emulator under Linux. This automatically corrected drifts of the pedestal position and produced histograms of the TDC, iAP DD (including primary pulses, since an MTDC was used) and SER.

Evaluation

The acquired data was saved in form of histograms which are split over multiple data files. These were read in, analyzed and visualized using ROOT [463, 464] and PEST [16] (see p. 481).

Measurements

Measurements were performed at about 10^7 gain, spe intensity (occupancy 1.8–5.5%), 23 °C and with active magnetic compensation. The CFD threshold was set to 0.2 pe and for the LED to 0.1 pe to measure the SER down to the pedestal.
The manufacturer's voltage divider of the R6594 was designed for negative HV and thus was retrofitted by G. Korga for positive polarity.

Shielded Photosensor Analysis eXperiment

Following the measurements at the LNGS, in Garching at the TUM an existing wooden dark box and NIM [17] rack were used to construct an improved photosensor

[15] Due to the long time of flight of the photons to the PMT.
[16] Photosensor Evaluation Software Toolkit
[17] Nuclear Instrumentation Module

characterization setup, which was subsequently upgraded multiple times using the complete know-how contained in ch. 2. This setup was dubbed Shielded Photosensor Analysis eXperiment (SPAX), is shown in figs. 3.2–3.7, and was used for the measurements studying TST, EAPs, LAPs and light emission.

In its final variant, SPAX consists of a wooden dark box (hence the name) into which a Faraday cage was integrated and an electronics rack.
The Faraday cage was realized using aluminum plates, L-profiles covering the slits, and interlocking U- and T- profiles for the lid.
The light source, optics, and photosensor are positioned on an optical breadboard for improved stability and reproducibility. All surfaces were covered with removable antireflective high-grammage stage molleton to suppress scattered light. As light source, a fast violet LED (light emitting diode) driven by a self-built Kapustinsky pulser [356, 357] (made by Laszlo Papp) with a measured combined jitter (see p. 478) of 472 ps RMS [18] (FWHM [19] 1.10 ns) was employed. The light source emission time distribution (ETD) of LED and Kapustinsky pulser and the jitter distribution of the electronics can be deconvolved from the TTD with the evaluation software (PEST), since both were determined for all measurements. The light source was characterized with a low-TTS silicon photomultiplier (SiPM), whose TTD was also unfolded from the obtained ETD. Originally, a picosecond diode laser with 48–50 ps emission jitter was used; however, this device ceased to function and could not be repaired, which required to switch to a fallback solution. The photosensors are illuminated homogeneously by reflection off a Spectralon plate and by maximizing the distance between light source and sensor, and the light level is reduced to single photons through NDFs.
The electronics is in large part NIM-based and located in a rack next to the dark box. An analog coincidence on the trigger signal starting the Kapustinsky pulser and the photosensor output is used for acquisition as it preserves the trigger time. The sensor opens a gate with adjustable window and the trigger is delayed to lie within this window. This method allows to obtain both low 2 pe contamination (see p. 90) and short acquisition times, since the flash analog-to-digital converter (FADC) readout and storage capacity is the statistical bottleneck. As FADC an Acqiris DC282 with 8 gigasamples/s (GS/s), 10 bit resolution and acquisition rates of up to 5 kHz was used, which allowed offline analysis (see pp. 127 and 336).

To speed up the measurement preparations, a LabVIEW script [425; v. 2011] was written (programmed by Mario Schwarz) which allowed to adjust the occupancy, gain, and CFD threshold.
The data analysis was carried out with a flexible dedicated evaluation software (PEST) based on ROOT [463, 464], which was programmed for this purpose and is detailed on p. 481.
SPAX can measure the TTD, pulse shape, gain, SER, DCR and APs with delays from 0.8 ns to 100 µs (EAPs, iAPs, LAPs).

In the course of this work, the setup was relocated no less than four times (TUM E15 neutrino laboratory → Max-Planck-Institut für Physik (MPP) → neutrino laboratory → Maier-Leibnitz laboratory → neutrino laboratory; the dark box moved twice).

[18] Root mean square
[19] Full width at half maximum

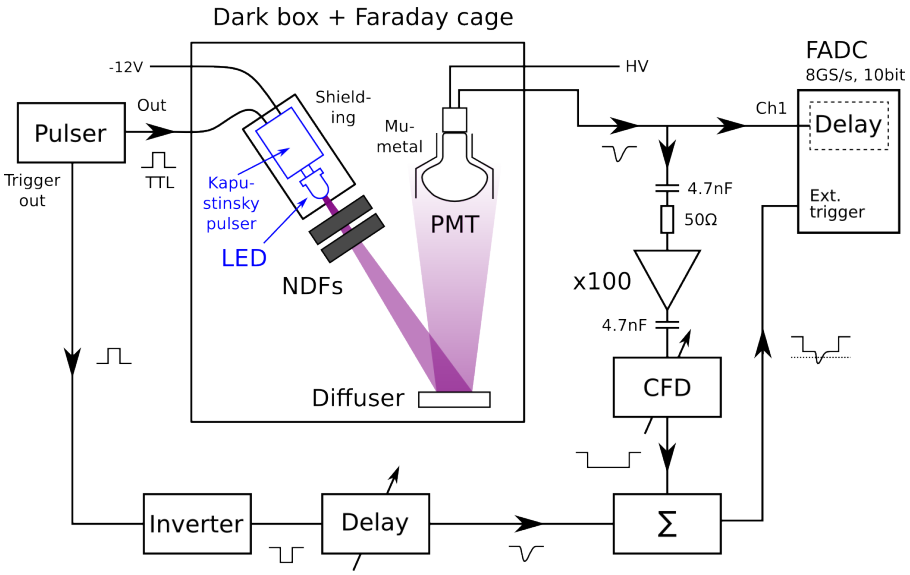


Figure 3.2: Shielded Photosensor Analysis eXperiment (SPAX): Schematic of the standard layout of the setup. This configuration is used for measurement of the transit time distribution (TTD), early afterpulse (EAP) delay distribution (DD), pulse shape, gain, and charge distribution. In addition, the dark count rate (DCR) is measured with a scaler on another constant fraction discriminator (CFD) output (not shown). Inside a Faraday cage dark box, a fast LED is driven by a Kapustinsky pulser and illuminates the photosensor homogeneously by reflection off a diffuser plate. A coincidence on light pulse and photosensor output which preserves the time information of the pulser trigger output allows fast acquisition at low occupancies. The unaltered sensor signal is detected by a flash analog-to-digital converter (FADC) which provides the benefits of offline analysis. A detailed explanation is given in the text.

During its stay at the MPP, an early variant of the electronics and a different dark box, laser, and Faraday cage were used. However, since the setup was still in the development phase at this time and no magnetic shielding was implemented yet, these measurements are not reported here.

Layout

Standard configuration

In fig. 3.2 the schematic layout of the standard electronics configuration is shown, which was used for all purposes except iAP and LAP measurements. The layout is based on the FADC acquisition method which was explained in 2.3.8 (see fig. 2.48b). After several iterations, the setup was reduced to the essential components, which are depicted in figs. 3.2 and 2.48b, to reduce noise and jitter.

A start pulser triggers light generation by a fast LED which is driven by a Kapustinsky pulser. The light is attenuated by NDFs and hits a reflective diffuser, which illuminates the magnetically shielded PMT (or other photosensor). The sensor output is connected directly to an FADC if possible, or amplified beforehand (not shown) if the pulses are too small.
The main target is to provide a coincidence between light pulse and sensor output to eliminate the acquisition of empty waveforms. This requires to wait for the

signal from the sensor, mostly due to its transit time, and thus normally the FADC acquisition would start when the output pulse arrives. However, in order to determine the TTD, the acquisition must be initiated by the trigger output. This is achieved by inverting and delaying the positive trigger pulse, while the PMT pulse is split [20], amplified and fed to a CFD which opens a gate. The delayed trigger output is added to the logical NIM signal from the CFD with an analog sum (three 17 Ω resistors in star connection). The output of the sum is then fed to the external trigger input of the FADC, whose threshold is set to trigger acquisition only if both sum signals overlap. Since the photosensor opens the gate and the trigger signal sits on top, the trigger time is determined by the latter.
The acceptance window is set by the CFD gate length and the delay line adjusts the position of the main peak in the TTD.
In order to minimize the electronics jitter, an analog sum is used instead of a FIFO (which also would introduce a cutoff frequency). For the same reason, the threshold is chosen to lie on the point of the steepest initial flank (see p. 335) of the trigger output, which is broadened a bit by the delay line which acts like a low pass.
The FADC can acquire the waveform in expectation of a trigger in advance, which is equivalent to an internal delay and is used to record the sensor output before the processing time of the coincidence has passed.
Since the start pulser only produces one pulse per period and the CFD output is either on or off, data taking is only started by primary pulses (PPs) and not APs (see fig. 3.3b).
Because the FADC input was found to float at about −10 mV to the common ground, the amplifier input was isolated by a high pass. Since the amplifier output also introduced a constant offset, the CFD input was isolated as well. To protect the pulser and CFD output from each other, 18 dB attenuators were included on both inputs of the analog sum.

Afterpulse configuration

The iAP and LAP measurement layout (fig. 3.3) is based on the standard arrangement and extends it with additional components to realize a triple coincidence of light pulse, PP and AP which preserves the time stamp of the trigger output. This coincidence allow to greatly reduce both the acquisition time and the amount of recorded waveforms without APs. Since the information about the AP rate or probability is lost in the recorded data when using a coincidence, the AP probability is determined separately with a scaler using the number of occurring PPs and the number of waveforms containing APs (see 2.5.2.1).

To realize the triple coincidence, the analog sum is connected to a CFD (CFD2) instead of the FADC trigger in. This serves the same function, namely to trigger on the coincidence of light pulse and PP within a TTD window determined by the CFD1 gate and Delay1. The PP number is counted with a scaler and another output of CFD2 opens a gate (Gate1) which sets the length of the DD. A second output of the amplified sensor signal is fed to another CFD (CFD3). The outputs of CFD3 and Gate1 are connected to a logic AND (AND1), which registers whether an AP is present. To avoid triggering on PPs by mistake, the output of Gate1 is delayed (Delay2) to lie behind the PP gate from CFD1. In this way only APs produce an

[20] At the T-crossing in the figure a 50:50 splitter is placed which consists of three 17 Ω resistors in star connection.

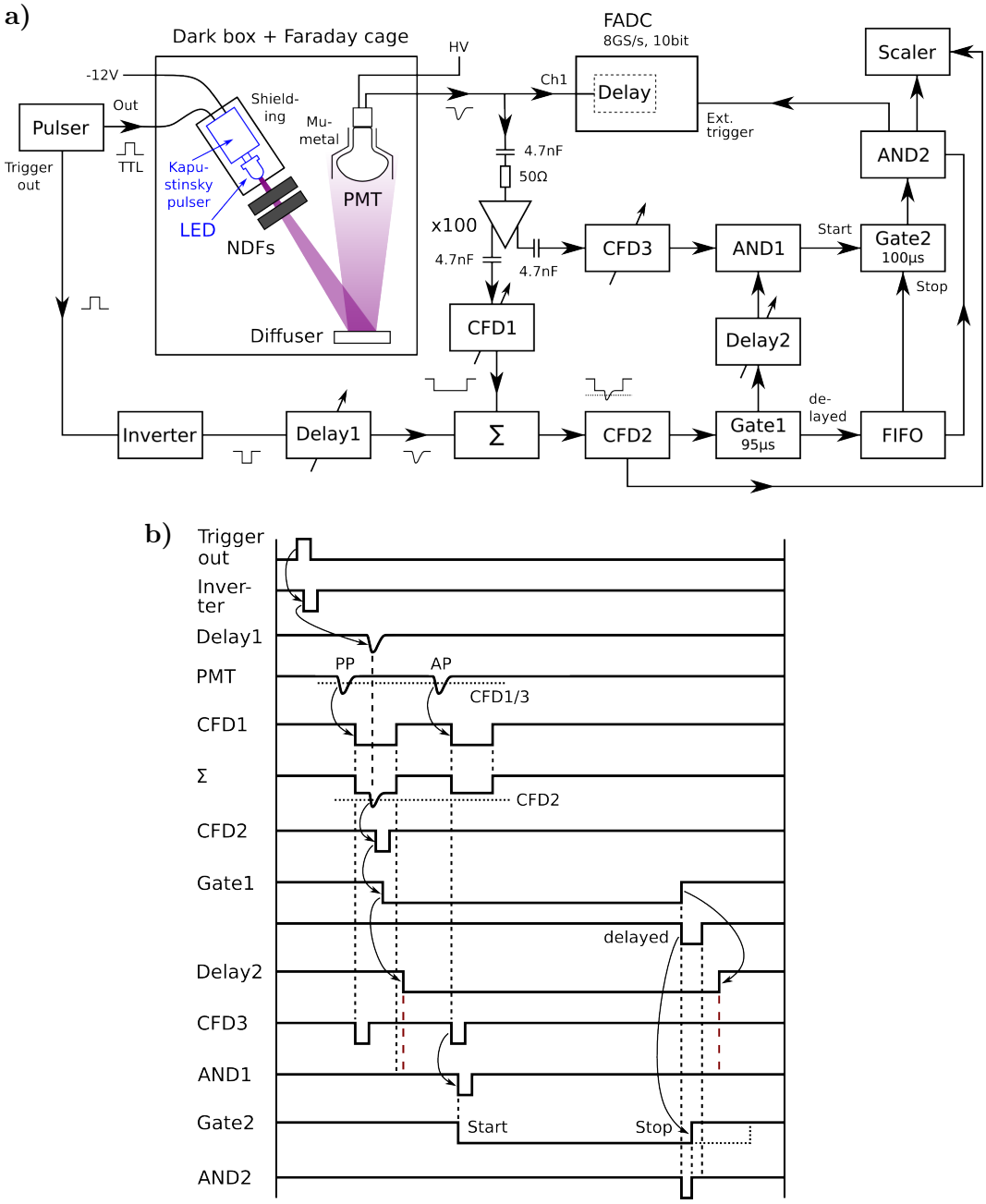


Figure 3.3: SPAX: Layout for ionic afterpulse (iAP) and late afterpulse (LAP) measurements using a triple coincidence between light pulse, primary pulse (PP) and afterpulse (AP). This requires to determine the AP probability with an additional scaler.
a) Schematic layout. The values of Gate1 and Gate2 are for an LAP measurement. Discussion see text.
b) Timing overview of the outputs of the components in a). Small transit and processing times are neglected for better readability. Red long dashes denote the AP acquisition window, which is delayed to produce no overlap with PPs. Gate2 would be longer (dotted line) but is stopped by the delayed output of Gate1. For detailed explanations see text.

output on AND1. This has the effect that Delay2 introduces a minimum delay before which no APs trigger the coincidence. Since a large part of the EAP DD overlaps with the TTD, this excludes most EAPs[21] and is the reason why for EAP measurements the standard layout was employed. In the iAP/LAP configuration, CFD3 is used instead of a second output of CFD1 to allow to set an independent threshold for the APs.
At this point a triple coincidence of light, PP and AP is realized and could be used to trigger the FADC. However, then the AP would start acquisition and the time of light output and the TTD information would be lost. Therefore, three further components are added (on the right in the schematic) to preserve the time information, similar to the standard layout.
The triple coincidence (AND1) starts another gate (Gate2), whose length should exceed that of Gate1. A delayed output of Gate1 which produces a logical signal after the end of the gate is used to stop Gate2. The signal of Gate2 and the delayed signal of Gate1 are connected to a logic AND (AND2). Due to the processing time of Gate2 and slightly different transit times in the electronics there is an overlap between both signals in AND2. This produces a logic signal which is tied to the pulser trigger out with a fixed delay and thus preserves the start time information. Because the trigger signal passes through several electronic components, a modest amount of jitter is introduced, but since for iAP and LAP measurements µs time scales are used, an uncertainty in the order of a few nanoseconds is tolerable. The output of AND2 produces a signal only if one or more APs are present which is used to trigger the FADC and count the AP *probability* with a scaler – the AP *rate* could be determined using AND1. The FADC acquires the waveform in advance and stores it if a trigger signal is detected. The FADC window should be larger than the sum of the CFD1 gate, Gate1 and Delay2 in order to fully record the PP TTD and the AP window opened by Gate1.
The inputs of CFD1 and CFD3 were decoupled with high-pass filters to eliminate floating grounds. Optionally, for small signals an additional amplifier can be used before the T-crossing.

Dark box

In figs. 3.4e and 3.4f the final layout of SPAX is shown. In figs. 3.4a–3.4d the different construction stages of the setup are shown, starting with the original wooden box after removal of the worn felt lining of walls and lid. After addition of all installations, the box provides an interior space of about 200 cm length, 110 cm width, and 110 cm height for experiments.

The dark box was made light tight through several measures.
Slits and gaps in the box were filled with black silicone.
Unused cable feed-throughs (BNC, SHV) were covered with caps and in addition both unused *and* used feedthroughs were coated with several layers of black fabric tape.
To seal the critical gaps between lid and box, a rectangular ring of aluminum U-profiles was welded together and bolted on the top of the box, while a matching ring of T-profiles was welded on the lid. When closing the box, both rings mesh and create a light trap. In order to avoid gaps between the T-profiles and the opposing surfaces,

[21] Uncorrelated EAPs can, however, still be contained in the waveforms.

Figure 3.4: Shielded Photosensor Analysis eXperiment (SPAX): A photosensor characterization setup specialized on timing and small-charge effects which was developed and continuously improved during this work.
a) Original wooden dark box with removed lid. The predecessor box was twice as long and split to create two dark boxes; the second box is visible in the top right. The open side was closed with a flange.
b) Upgrade to an aluminum (Al99.5) Faraday cage with 1 mm sheets as walls, L profiles covering the edges, and a hinged lid with light trap through interlocking T- and U-profiles. On the floor two rectangular hollow sections are placed to distribute the weight of the optical table. Foam rubber pedestals were used for vibrational decoupling; one is visible in the picture.
c) Antireflective (AR) cover with tailored, removable (hook and loop fasteners), high-grammage (345 g/m^2) stage molleton.
d) Two optical breadboards (M6 thread, 25mm spacing, 10 mm thick Al, 27 kg each) are mounted on the table for a stable and reproducible alignment of the optics.
e) With light source, optics, PMT and mu-metal installed. A self-built Kapustinsky pulser with a fast LED is used. The intensity is attenuated with neutral-density filters (NDFs) placed in a tube which is tilted vs. the emission direction to reduce the amount of reflected light. A reflective Spectralon diffuser plate is used to maximize the distance between light source and photosensor and optimize the homogeneity of illumination. All surfaces are covered with molleton to minimize reflections and scattered light. A second experiment (PoLiDe) shares the same dark box, and is covered at the lower edge.
f) Final site of the setup. Note the different surroundings in the images; the setup moved four times, the dark box twice.

a foam rubber inlay was placed in the U-profiles, an additional layer of softer foam rubber was applied in the region of the corners, and the lid was pressed onto the foam rubber with clamps on all sides. This design proofed to be light tight when tested with a PMT and a strong flashlight. As an added degree of redundancy, two layers of the heavy molleton used for the antireflective (AR) lining of the interior, which also is nearly light-tight, were fixed on top of the lid. Through gravity these layers automatically cover the gaps when closing the box.

The light tightness was tested (see p. 161) after each step of the construction (wooden box with old wooden lid, Faraday cage with Al lid, application of molleton), and where necessary leaks were sealed as described above. Testing was performed by first placing an experimentalist with dark-adapted eyes in the box, who looked for leaks while a co-experimentalist passed a flashlight over critical spots[22]. The current test position was indicated acoustically. In a second step, a small PMT was placed in the dark box (fig. 3.6f). Its voltage was carefully, slowly ramped up until the muon hits and then the thermionic dark counts became visible. At this voltage, first an integral test (room illumination turned on and off / curtains drawn or open) was performed. Then the procedure with the flashlight was repeated.

Faraday cage

A PMT acts as an antenna for electromagnetic waves[23] due to its large conductive components (cathode, aluminized glass, dynode chain, wires). Unfortunately, it is quite efficient in this regard: When connecting a 10" inactive PMT with a short shielded cable to a digital oscilloscope and running a fast Fourier transform (FFT) analysis of the recorded waveforms, the present author could detect all locally registered radio stations and sort them by signal strength. This was judged to be problematic for the investigation of underamplified pulses and thus the dark box was retrofitted with a Faraday cage on the inside.

The skin depth was calculated for relevant frequencies and materials available for selection and an aluminum shield with 1 mm thickness was found to provide sufficient shielding.

To avoid unpleasant surprises, the expected resonance frequencies $f(i,j,k)$ of the Faraday cage were calculated [465]:

$$f(i,j,k) = \frac{c}{2}\sqrt{\left(\frac{i}{x}\right)^2 + \left(\frac{j}{y}\right)^2 + \left(\frac{k}{z}\right)^2} \tag{3.1}$$

with $i,j,k \in \mathbb{Z}$, while x,y,z are the dimensions of the Faraday cage, and c denotes the light speed. The lowest resulting frequencies were 74.9, 136.3, and 149.9 MHz, meaning that the public radio band is skipped. No source from the official frequency plan was found to match resonance frequencies up to $i = 3$ with x chosen as the longest dimension.

To construct the Faraday cage, 1 mm thick plates of Al99,5 were used for walls, bottom and lid, and the gaps between plates were covered with 3 mm thick Al angle sections to prevent the formation of slot antennas.

[22] Edges of the box, sealed gaps, gaps between lid and box, feed-throughs and screws.
[23] Radio, WLAN, cell phones, non-public broadcasts, emissions from laboratory electronics, ...

The Faraday cage was earthed independently from the laboratory ground, and the optical breadboards were connected with copper braids to the shielding. The mu-metal shield was linked to the breadboard in the same manner and the ground of all feed-throughs was conductively connected to the Faraday cage as well. The earth of all devices in the electronics rack was connected with broad copper braids and the rack was connected in the same way to the Faraday cage. The ground of the Acqiris FADC board in the measurement computer was found to float and thus was connected to the rack. Care was taken to avoid the formation of ground loops by using a starred connection pattern.
As a result of these endeavors, it was possible to achieve a baseline RMS of down to 0.24 mV in the measurements, which allowed to set pulse thresholds of as low as 1.1 mV.

Light source and optics

For a fast, application oriented characterization of large photosensors, it is important that the light source has low emission jitter and that the sensor is illuminated homogeneously (see also p. 332 and [170; pp. 572–573] for a discussion of light sources).
For SPAX LEDs and lasers were used, which each have individual advantages and drawbacks:

- **Emission angle:** LEDs have a conical emission profile ("viewing angle"), which in principle allows integral illumination without a diffuser for sufficiently large distances and viewing angles. Nevertheless, using a diffuser helps to improve the homogeneity.
  Lasers, on the other hand, emit collimated light, which is good for surface scans, but requires a diffuser for integral illumination.
  Both light source types can be coupled into a fiber. Due to the emission angle of the fiber this allows homogeneous lighting if the fiber end is positioned in a distance from the sensor, and enables surface scans when placed in an adjustable holder close to the sensor surface.

- **Emission time distribution:** An LED has an emission time spread (ETS) between of 0.6 and 6.3 ns FWHM according to [466] (over 5000 tested LEDs). The value varies strongly between series and even specimens. While some samples show strong trailing flanks characterized by long exponential decays with decay times in the order of 10 ns or more, in fast LEDs this behavior is missing or present with far shorter lifetimes.
  Picosecond diode laser have far smaller pulse widths of 30–90 ps[24] at wavelengths close to 400 nm (to match scintillator emission). However, they typically also show a tail extending to 1 ns or more for a few percent of pulses, which should be kept in mind.

- **Emission spectrum:** LEDs emit over a small to broad wavelength range. This is closer to the real light source and allows to study the wavelength-dependence of effects using optical band-pass filters.
  Lasers, in turn, produce (almost) monochromatic light.

[24] Judging from a product survey of the manufacturers ALS, BEC, Edinburgh Instruments, Hamamatsu, onefive, and PicoQuant.

- **Coherent emission:** Laser light is coherent. This means that it produce speckles, and one cannot use sandblasted glass as a diffuser, since it provides only one scattering surface and can amplify the speckle size. Furthermore, for lasers it is also necessary to calculate the propagation of the beam profile with Gaussian beam optics (see for example [467; ch. 6]).
- **Stability:** For both lasers and LEDs the emission intensity and ETD is very sensitive to variations of the supply voltage and temperature. For LEDs the spectrum also depends on this to a lesser extent. Therefore it is recommendable to use stabilized power supplies and, if possible, control the temperature with e.g. a Peltier cooler, since otherwise drifting will occur.

Laser

Originally, a ps diode laser (Edinburgh Instruments EPL-405mod) with 48-50 ps pulse width, a wavelength of 403 nm and a modified minimum repetition rate of 5 kHz (to allow LAP measurements) was used. For the MPP measurements a PicoQuant PDL 800-B with 70 ps, 405 nm was employed which was coupled into a fiber. Unfortunately, the EPL-405mod ceased to function before the final measurements were made, could not be repaired, and a replacement laser could not be procured.

Light-emitting diode

For this reason, the laser was substituted by a self-built Kapustinsky pulser (made by Laszlo Papp) driving a fast LED. This provided a sufficiently low ETS for the characterization of large PMTs.

Two versions of the pulser were constructed (see figs. 3.5b–3.5d and 3.5e–3.5i), where the second design was improved based on the experience gained during construction of the first. The dimensions of the pulsers were minimized by using surface-mount device (SMD) components to lower the jitter (figs. 3.5b and 3.5e show the pulsers soldered to the LEDs with a 1 euro cent coin for size comparison). The LEDs with attached pulsers were fixated in LED holders and placed in shielding housings to suppress the emission of electromagnetic waves from the fast voltage spikes. These housings were made light-tight except for a small hole in front of the LED. The completed pulsers were fixed on optical posts for reliable positioning on the optical table and covered with several layers of molleton for redundancy. The pulsers were supplied with the stabilized −12 V power supply of the NIM crate, which showed the lowest ripple of all available power adaptors.

SiPMs with small jitter were used to measure the ETD of both pulsers and various LEDs and select the combination with the shortest ETS. This included a Hamamatsu S10931-100P (3×3 mm, 500–600 ps TTS FWHM) with an amplifier from Patrick Krause (see figs 3.6a and 3.6b). Furthermore, the same amplifier was used with a faster Hamamatsu S10362-11 (1×1 mm, 200–300 ps). Finally, a Ketek 3325-WB (3×3 mm, 200–300 ps) was employed in combination with a Ketek SiPM evaluation kit including a preamplifier (PEVAL-KIT) from Felix Henningsen (see fig. 3.6c and 3.6d). The SiPMs were amplified in the dark box, and the Hamamatsu SiPMs and their amplifier were also shielded in a grounded high-frequency box, since the sensors picked up correlated noise from the pulser.

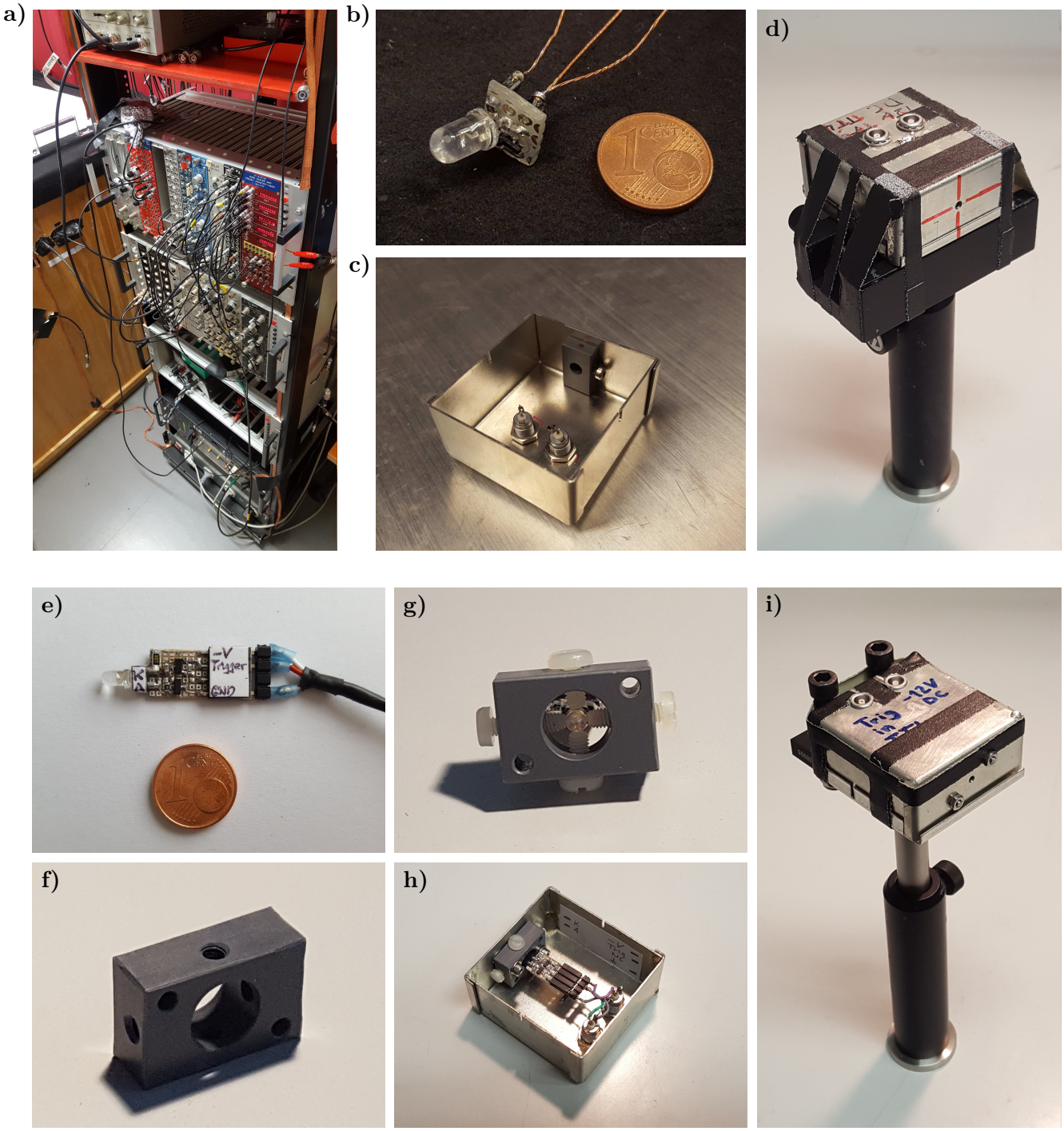


Figure 3.5: SPAX electronics and light sources
a) Electronics for timing and iAP measurements. The pulser is visible on the top left, the FADC is situated on the bottom. All cables were kept as short as possible to reduce electronic transit times, jitter from electronics and pulse shape distortion; the electronics was arranged in a manner to facilitate this. The processing electronics is NIM-based, and all active components are placed in the upper rack, while the lower crate (not powered) only houses passive components (delay lines, inverter) with no connection to its ground in order to reduce noise. Noise reduction is also the reason for the separation between the FADC, which is a board in a measuring computer, and the NIM electronics. The Kapustinsky pulser was powered with the stabilized −12 V supplied by the crate (red connectors on right. The silicon photomultiplier (SiPM) power supply is visible above the FADC.
b)–d) LED pulser mark 1: **b)** Fast LED on a Kapustinsky pulser (built by Laszlo Papp). **c)** Shielding box for the pulser with LED holder at top right. **d)** Assembled pulser on optical mount
e)–i) LED pulser mark 2: **e)** LED and Kapustinsky pulser (also by L. Papp). **f)** Adjustable LED holder. **g)** With LED mounted. **h)** Pulser installed in shielding enclosure with attached LED holder. **i)** Finished pulser mounted on post

Figure 3.6: SPAX setup characterization and PMT equipment
a)–**d)** SiPMs used to determine the light source emission time distribution (ETD), which allowed to select the fastest LEDs, optimize the pulser, and deconvolve the ETD from the TTD.
a)–**b)** Hamamatsu S10931-100P (3×3 mm) and amplifier (Patrick Krause) in shielding housing. The SiPM is on top of the C-shaped extension and located directly below the window in the lid visible in b).
c)–**d)** Ketek 3325-WB (3×3 mm) attached to Ketek SiPM evaluation kit with preamplifier (PEVAL-KIT) mounted on post.
e) Measurement of the illumination homogeneity in the cathode plane. The monitor PMT (on left) was fixed and used to eliminate drift of the light source. The distribution was measured by moving the roaming PMT (right) in the left-right and up-down directions.
f) Light leak search: A small PMT was placed in the dark box and powered on. The illumination was varied either integrally (turning the light on and off) or by pointing a flash light on critical regions (lid, feed-throughs, edges) while monitoring the dark count rate (DCR) for an increase. Found leaks were sealed and tested again. This procedure was carried out at each stage (bare wood, Faraday cage, AR fabric). Here the last iteration is shown.
g)–**h)** Self-built mu-metal for the 12" Hamamatsu R11780-HQE, for which no commercial magnetic shielding was available. Two segments were cut from a coil of mu-metal, glued to a cylindric shape, connected with overlap and painted with antireflective coating. The complete length of the PMT and a region of over half the cathode diameter in front of it was shielded, which reduced the measured residual field from 44 µT to $\leq$ 3.2 µT.
i) PMT holder used for PMTs with diameters $\leq$ 10".

Despite the design improvements of the second pulser, the lowest overall ETS of 1.10 ns FWHM after deconvolution of the SiPM TTD and statistic subtraction of the electronic jitter was found for the first model (see fig. 3.8 for the ETD). Therefore, the earlier version was used for all measurements. According to [466], this is a quite good value and typically the ETS can not be lowered much further without extensive LED screening.

Neutral density filters

An occupancy of $< 10\%$ was set and adjusted by use of NDFs, which were contained in a tube. In order to minimize stray light and the avoid introducing an optical jitter from photons which are reflected on the surfaces of the NDFs and the LED housing, the following measures were adopted:

- The NDF barrel was placed directly in front of the LED aperture.
- Within the tube the NDFs were moved as close as possible to the LED.
- The NDF with the largest optical density (OD) was positioned closest to the LED and was AR-coated.
- Calculations were carried out to find the best combination of tube length and the tilt angle between tube and pulser housing to reduce reflections. An angle of 21° and a short tube length of 1" eliminated all rays which were reflected on the first NDF surface and the pulser housing for LED viewing angles up to 30° while transmitting all direct rays. Since for all measurements an AR-coated NDF with OD 2 was placed closest to the LED, reflections on later surfaces are greatly suppressed and can be considered negligible (suppression by a factor of at least 10^4 due to having to pass the first filter *twice*). The tube was chosen and positioned accordingly.
- The tube and LED housing were wrapped with several layers of molleton to eliminate stray light from non-direct rays.

Diffuser

The light emitted by the LED was directed onto a reflective diffuser in form of a Spectralon plate.
The dark box coordinate system (x, y, z) was chosen to have its origin in the center of the LED spot on the Spectralon plate (compare fig. 3.4e). Here, the x axis lies in the planes of the diffuser plate and the table, y shows in direction of the table surface normal and the z axis corresponds to the direction of the diffuser surface normal. The coordinate system is left handed (the x-axis points to the right when looking from the diffuser plate towards the sensor).

Factors contributing to inhomogeneity

In this arrangement, six factors contribute to non-homogeneity:

- The emission angle distribution of the diffuser in dependence of incidence position and angle
- The intensity profile of the diffuser illumination

- The distance between diffuser and sensor
- Reflections and scattering on setup surfaces
- The tilt of a sensor surface vs. incident light
- The intensity reduction with $1/r^2$ to sensor surfaces not on the optical axis

The curvature of the PMT is *not* considered. The setup tests how a PMT performs for front illumination, and the curvature of the cathode is part of device performance and not of the setup.

The emission of an ideal Lambertian reflector follows a cosine-law, as does the intensity drop from the tilt of a plane surface. If no reflections from other surfaces occur and the diffuser emits from a point-like region, the intensity distribution therefore scales with $\cos(\theta)^2$ where θ is the emission angle from the diffuser.
To calculate the expected intensity drop at a position on the photosensor, θ can be related to a distance z to the sensor along the diffuser surface normal and a lateral distance x from the z-axis by

$$\theta = \arctan\left(\frac{x}{z}\right) \tag{3.2}$$

Spectralon is a *near*-ideal Lambertian reflector. In [358] for the illumination of Spectralon under 40° vs. the surface normal small non-Lambertian components were observed and the emission intensity distribution $I(\theta)$ was slightly asymmetric. The amount of reflected light increased towards larger scattering angles (away from the light source incidence angle) with a 2% increase beyond the expected value per 30°. [468] observed for small illumination angles that more light is reflected close to 0° than expected from the cosine-law (6–9% more at 0°, depending on sample). This was attributed to a small specular reflection component. The measured curve crosses the cosine function at about 51° and at large angles lies below that of an ideal diffuser. From this follows that for small incident angles a steeper intensity gradient occurs, which is detrimental for homogeneous illumination. On the other hand, the emission distribution is more symmetric for small than for large incident angles.

If an area of the diffuser is illuminated, the effective area of the light source increases which further improves the homogeneity. Based on this consideration, the LED should be placed as far as possible from the Spectralon plate. Since different times of flight to the near and far end of the illuminated region also increase the optical jitter, however, a moderate distance was chosen as a compromise.

Backscattered light from other surfaces in the setup can in principle improve the homogeneity by compensating the $\cos(\theta)^2$ drop. However, as this introduces optical jitter and is hard to control, all areas were covered with AR fabric (heavy black Molleton or felt).

Layout of the optics

In order to optimize the illumination homogeneity of the photosensors, diffuser plate and sensor were placed at the maximum possible distance of 138.3 ± 0.3 cm from each other (see fig. 3.4e). This distance z was determined by the dimensions of the R11780-HQE, its VD and the minimum cable curvature radius and was also used

for other PMTs for reasons of comparability. Since the part of the cathode which is closest to the diffuser lies at this z value, it is labeled cathode plane in the following.
A large Spectralon plate with 181.5 mm width, 181 mm height and 12.5 mm thickness was used to allow to reach a better homogeneity through wide-area illumination.
The center of the Spectralon plate was illuminated at an angle of 30°, which is a compromise between non-ideal effects (see above), and from a distance of 37.1 cm. Compared to an angle of 40°, this allows to set a larger distance between LED and Spectralon, the intensity difference on the LED-near and -far sides of the diffuser plate is smaller, and the PMTs are not hit by a potential specular reflection peak. Furthermore, for 30° the LED housing casts no shadow on sensors with radii of up to ca. 40 cm.
The rotation symmetry axis of the LED beam was set at a height of 21.0 cm in order to accommodate photosensors with diameters of up to ≈ 42 cm between central beam and table.
The diffuser surface normal was aligned with the optical axis (along breadboard holes) with $\pm 1°$ accuracy.

Combined with the LED viewing angle, the large dimensions of the diffuser plate and the nigh-perfect Lambertian emission profile of Spectralon [358, 468], these measures provide the most homogeneous illumination possible without using a larger dark box.

Homogeneity calculation

The expected homogeneity was calculated based on an ideal Lambertian diffuser. This was corrected by a shift of the emission intensity[25] of 1.5%/30°, the tilt of the cathode plane vs. incident light from the diffuser center for an $x > 0$, and the accompanying intensity reduction due to the increased traversed distance for off-axis rays.
The strongest influence was found for the distance dependence followed by the (identical) reductions from ideal diffuser and surface tilt. The emission distribution correction only resulted in small changes. For the cathode radius of the R11780-HQE (14.0 cm, corresponding to 5.8° from the diffuser center) an intensity reduction of up to 2.31% is expected for the LED-near side and 1.74% for the LED-far side (for the R5912 up to 1.14 and 0.75%, resp.).
The results of these calculations are compared to measured data in fig. 3.7b.

Homogeneity measurement

The homogeneity in the cathode plane was measured using a reference PMT and a roaming PMT (see fig. 3.6e).
For both PMTs an ETE 9111 with a diameter of only 1" was used for better accuracy. The reference PMT remained at the optical axis ($x = y = 0$ cm around a height of 21 cm), while the roaming PMT moved around to map the intensity distribution. The relative intensity $I_r = I(x, y)/I(0, 0)$ was determined by the ratio of the counted pulses detected by both PMTs within a fixed window. Before and after each measurement the DCR was determined with the same discriminator and scaler channels for both PMTs, and the DCR was interpolated linearly and subtracted. The dark box was opened in the dark to move the (inactive) roaming PMT and

[25] Linear interpolation between the measurements reported in literature for 0° and 40° incidence angle on Spectralon.

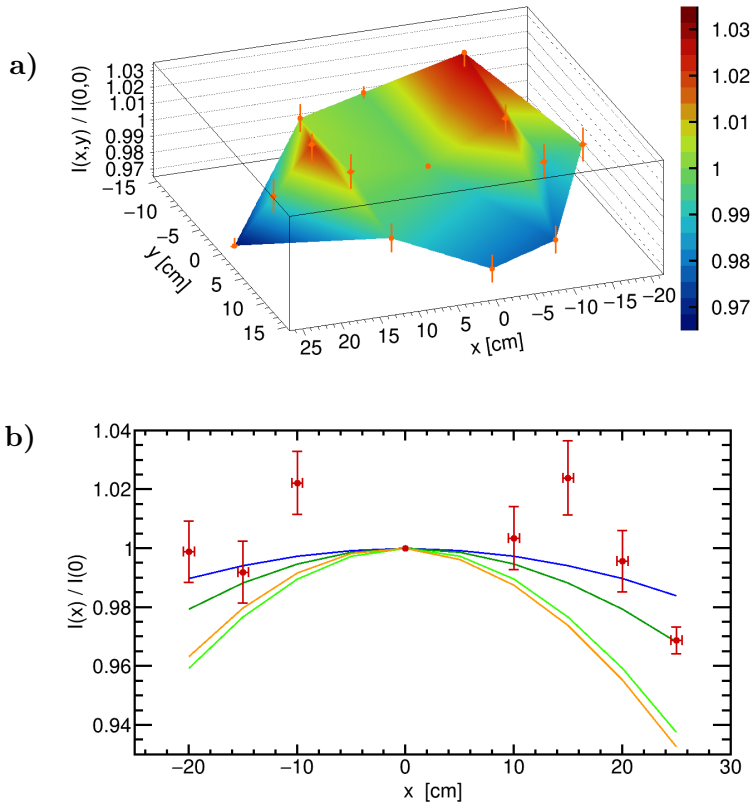


Figure 3.7: Measured homogeneity of illumination in the cathode plane
Relative intensity $I_r = I(x, y)/I(0, 0)$ compared to the center in a region of 15−25 cm radius around the beam axis. The x-axis is parallel to the table surface, the y-axis vertical to the table surface. **a)** Homogeneity map. I_r is plotted on the z-axis and also indicated by colors. The error bars for all three axes are drawn in orange. I_r lies between 96.9 and 103.3% in the complete area of interest. **b)** Intensity profile along the x-axis ($y = 0$). Measured values are drawn in red and can be compared to calculated curves for point-like emission from the diffuser (in 5 cm steps). Blue: ideal Lambertian reflector ($\cos(\theta)$ law). Dark green: in addition tilt of a plane surface vs. incident light (also a $\cos(\theta)$ law). Bright green: plus $1/r^2$ intensity reduction with distance. Orange: all of the above including an asymmetric emission profile as observed in [358]. The measured intensity roughly follows the theoretical curve but lies above it, shows fluctuations, and has a better homogeneity towards large x values. This behavior can be attributed to the wide-area illumination of the diffuser and to light which is scattered off the (AR-coated) surfaces in the setup. The excess around 15 cm might be caused by specular reflection on first Spectralon and then the AR-coated walls.

after closing the lid the PMT was allowed to cool down for several minutes with applied HV before the next measurement. Furthermore, the intensity values were corrected by the PDE factor between both PMTs, which was obtained from measurements where both PMTs were switched.

The measured intensity distribution is shown in fig. 3.7. The values on the x-axis (fig. 3.7b) lie above those which were calculated for point-like emission from Spectralon for large x-values.
This is not surprising considering that a large part of the diffuser plate was illuminated by the LED to improve the homogeneity.
In addition, scattering and reflections off the setup walls contribute to increased intensities near the brink: The dark box ceiling was located at $y = 29.8$ cm, the table top at $y = -20.9$ cm, the left wall (seen from the diffuser) at $x = -33.3$ cm and the covered PoLiDe experiment at about $x = 46$ cm. Therefore, more scattered light is

expected for large negative x and y values, which agrees with the observations in fig. 3.7a.
The specular reflection off the diffuser would lie at $x = -80\,\text{cm}$ and thus falls on the molleton on the left wall. An additional specular reflection off the fabric would hit the PMT at $x = 13.2$, $y = 0\,\text{cm}$, which could explain the increase observed at 15 cm in fig. 3.7b.
The other fluctuations in both plots are likely either a result of scattering on other surfaces or due to statistic fluctuations.

All in all, the wide-area illumination of the Spectralon and the stray light from the setup surfaces lead to a maximum deviation from the center intensity of $3.03 \pm 0.87\%$ within 15 cm radius ($2.22 \pm 1.07\%$ for 10 cm). The spot-to-spot deviations amount to up to $5.16 \pm 1.72\%$ within 15 cm ($2.22 \pm 1.79\%$ for 10 cm).
This was found to be acceptable for a high-quality characterization of large PMTs.

Transmissive diffuser

During this work also a study on Spectralon as a *transmissive* diffuser was carried out in collaboration with Hanna Kellermann.
First tests were performed with a slab of 3−4 mm thickness with Kellermann's setup, which showed an intensity drop to around 90% at $\theta = 10°$. This decrease is stronger than expected from an ideal Lambertian diffuser, sensor tilt [26] and the reduction due to higher distances (in sum 94.1% expected), but not by much. A large part of light is lost through backscattering in the Spectralon and above about 8 mm thickness no light is transmitted through the plate. For single photon measurements this is no problem, however, since usually NDFs have to be used anyway. This might make this a promising field of study in the future.
For SPAX in the end a classic reflective diffuser was used, since this allowed to maximize the diffuser–sensor distance.

Lens as diffuser

In an early stage of SPAX, as part of the diploma thesis of Michael Nöbauer [469], it was shortly tried to diffuse the laser beam with ball lenses and gradient-index lenses to achieve sufficient homogeneity of illumination.
On the horizontal axis the intensity dropped to about 89% at 8.5° (ca. 80% at 12.4°), which is not much worse than the measurements with Spectralon mentioned above. However, strong local fluctuations of up to 15% within 1.1° angular distance were observed, likely due to dust particles on the lens or its surface roughness. Furthermore, outside of the x-axis the intensity showed an *increase* up to 113.5% towards the ceiling and the same strong fluctuations. The rise might in part be due to stray light from the lid, which was still covered with felt (higher reflectance than molleton) at this time.
Due to the large variations, this approach was not pursued further, but might merit further analysis regarding the performance after thorough cleaning or possibly etching.

Setup fluorescence

As already discussed on pp. 297 and 299, after evaluation of the measurements a slight exponential decay following the main peak became evident, which could stem

[26] A pin diode was used, see p. 621.

from fluorescence of some material in the setup or from slow secondary recombination processes in the LED. This decay was not present in measurements of the same PMT at identical HV and similar conditions at the LNGS.
In a reanalysis of preliminary measurements performed at the MPP, the same decay as observed in SPAX was found with similar decay times (roughly 22 ns). Since at the MPP a ps diode laser and no magnetic shielding were used, this excludes the LED and effects caused by the Lorentz force as origins. However, at the MPP measurements a Spectralon plate was used as reflective diffuser as well, which supports the assumption that Spectralon fluorescence [344] might be the cause for this behavior.

Antireflective measures

The complete setup was coated respectively covered with AR material to reduce the amount of stray light as far as possible. For the walls, ceiling and table, the same high grammage (345 g/m^2) black stage molleton from Jostra as for the lid cover was used.
In order to select the fabric with the best properties regarding light tightness and the suppression of reflections at glancing angles, several dozen samples of heavy black fabric of different types (felt, velvet, molleton, untreated cotton cloth, ...) were ordered and screened for suitability.

The self-built mu-metal shielding (figs. 3.6g and 3.6h) was coated with two layers of over-pigmented black AR silicone resin paint from Berger. According to the vendor this paint has comparable properties to the AR coating used by large optics manufacturers.

Photomultiplier holder

PMTs with diameters of 10" or less were fixated in a self-built PMT holder (fig. 3.6i), which is based on two rectangular plastic blocks with triangular recesses. The holder is thus capable of fastening any PMT whose stem has a smaller diameter than the recesses. The PMT can also be fixated with a surrounding mu-metal shield.
The blocks are held with nuts on threaded rods which are fixed on a metal plate. The upper block is carefully secured to provide enough stability while not exposing the PMT envelope to mechanical stresses. For this purpose the inside is covered with foam rubber. To provide stable and reproducible positioning, the complete holder can be screwed to the optical breadboards.

Magnetic shielding

For SPAX passive magnetic shielding with mu-metal was employed (see p. 436).

For the R11780-HQE (12") no commercial magnetic shielding was obtainable, so a self-built one was fabricated from available mu-metal using eq. (2.225) (see figs. 3.6g and 3.6h). Due to the large length of the PMT, two single-turn segments were glued to cylinders and connected with an overlap (diameter 32.0 cm, total length 53.7 cm). The resulting shield covers the complete dynode stack and extends to more than half the cathode diameter in front of the cathode, as ideally recommended. The shielding was coated with two layers of AR paint.
A Samsung Galaxy S6 magnetometer was used to determine the residual fields and a reduction from a local field of 44 µT to $\leq$ 3.2 µT total field (the detection limit of

the device) was found. This results in a shielding factor of ≥ 13.75 and allows to reconstruct a permeability of 20400 for the mu-metal (whose handling history was unknown), which is still rather good. No significant influence on PMT properties is expected for such low fields.

Electronics

A schematic layout of the electronics of the setup was given in figs. 3.2 and 3.3 and was already discussed before. The electronics rack is shown in fig. 3.5a.

As pulser, a Hewlett-Packard 8012B was used. Five delay lines with delays adjustable from 2.5 to 66 ns (0.5 ns digit) and one with 2.5–34 ns delay (0.25 ns digit) from various manufacturers (FE290, MPI, berer co, CERN N9053) were used. For Delay2 one of the 66 ns delay lines was employed, all others were used for Delay1.
The PMT HV supply was a CAEN N471, which has two channels, can provide both polarities, allows to monitor the current, ramps up a set voltage to avoid a too fast increase, and inhibits voltage after power failures. For the SiPMs an EA-4003 power supply was used.
The pulse height fraction which is transmitted to the FADC by the combination of splitter and high pass was measured to be 80.425%.
As amplifiers one or multiple channels of a CAEN N979 16-channel fast amplifier were used. If no preamplifier is used for the FADC signal, a high-pass was required to decouple the amplifier input from the FADC, since the Acqiris Ch1 ground floats at about −10 mV.
The discriminator CFD1–3 were different channels of an EG&G ESN CF8000 octal CFD. This amplifier has a minimum threshold of 10 mV and thus requires one or more preamplifiers to measure underamplified PMT pulses.
The logic ANDs and the FIFO were realized with three channels of a Phillips Scientific 754 quad four-fold logic unit.
Gate1 and Gate2 were created by a LeCroy 222 dual gate Generator.
Pulses were counted with a CAEN N145 quad scaler. This device allows to set a fixed time for the acquisition gate, which allows the accurate determination of the DCR and AP rate/probability.
The pulses were recorded with an Acqiris DC282 FADC with 10 bit vertical resolution, up to 8 GS/s (interlacing), 4 channels, acquisition rates of up to 5 kHz, < 1 ps sampling jitter, and 15 ps trigger time precision.
The complete setup was subjected to extensive, repeated interference reductions to achieve the lowest possible noise on the signal.

Setup jitter

The electronics jitter was measured as described on p. 336 for all measurements, with resulting values between 150.66 ± 0.10 ps and 187.69 ± 0.13 ps. The electronic jitter probably originates in large part from the HP pulser, as this showed some signs of instability, and by replacing the start pulser the values can likely be improved significantly.
The time of flight differences (optical jitter) were calculated to lie between 40.25–61.18 ps assuming full illumination of the Spectralon plate, which is larger than for the LNGS measurements (9.76–20.92 ps for the same PMTs). The increased optical jitter is a disadvantage of the integral illumination of the diffuser, which improves

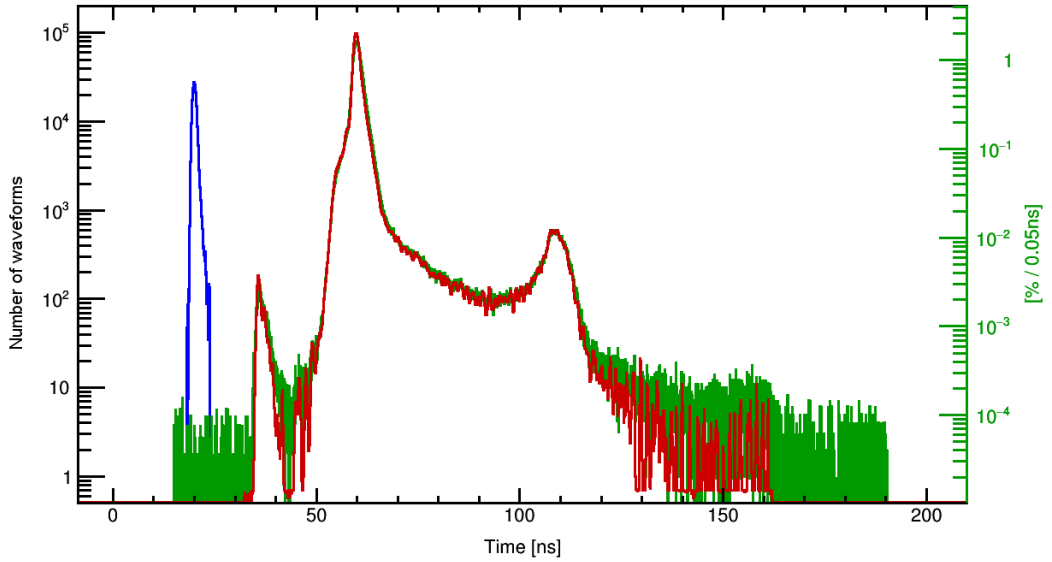


Figure 3.8: Transit time distribution (TTD) for a Hamamatsu R5912 PMT before (green) and after (red) deconvolution of the light source emission time distribution (ETD, blue). The measurement was acquired for +1425 V ($g = 1.3 \cdot 10^7$), 2.9% occupancy, $h_s = 4.5\%$ pe setup threshold, and $h_f = 1.44\%$ pe fit threshold. The artificial oscillations introduced by the deconvolution algorithm are clearly visible in flat regions. The SiPM TTD (approximated by a Gaussian) was unfolded from the shown ETD.

homogeneity, however. For spot illumination of the diffuser center (with a laser), the flight time differences amount to only 10.56–22.61 ps, which is comparable to the LNGS values. To reduce the optical jitter, one could use a lens or reduce the distance between LED and diffuser, but since the total setup jitter is dominated by the LED ETD, no large overall improvement would be expected.
The ETD of the LED (fig. 3.8) was measured with SiPMs as described above. The ETS was found to be 1.17 ns FWHM (497 ps σ using (2.68)). After unfolding of the SiPM TTD (approximated with a Gaussian) and statistic subtraction of the electronics jitter a value of 1.10 ns (472 ps σ) results. The σ of the total setup jitter (LED, optics, electronics) then lies between 497–511 ps (statistic sum) and 663–721 ps (systematic sum), where the ranges denote differences between measurements due to PMT diameter and electronics. With the EPL-405mod laser the total setup jitter (σ) was only between 152–190 ps (statistic sum) and 182–231 ps (systematic sum), which is far lower and is why a ps laser was used originally. For the laser the total jitter is dominated by electronics. Unfortunately, the laser ceased to function and it was not possible to repair or replace it. In a next step, the setup should be upgraded with a new ps-laser laser, as this improves the emission jitter and optical jitter, and a better start pulser, which would reduce the electronics jitter.

Despite these extensive measurements, it is difficult to quote a timing error for the TTS or the TTD. While the setup jitter stated above applies for individual pulses, for features in the TTD the fluctuations of statistically independent jitter components and TSEs can partially cancel.
The only way to obtain the integral timing error is to deconvolve (and thus eliminate) the total setup jitter from the TTD and compare the resulting distribution with the original TTD. This unfolding was implemented in PEST, but the employed

deconvolution algorithm was found to introduce oscillations in the TTD and could not be considered reliable yet. Therefore, the algorithm was only used to deconvolve the SiPM response function from the narrow ETD, where no oscillations occurred.
Alternatively, one can estimate the TTS without setup contributions by approximating the setup jitter components (ETD, optics, electronics) and the TTD as Gaussian functions (statistic subtraction). From (2.74) follows that the resulting combined distribution then also is a Gaussian whose σ equals the statistic sum of all individual standard deviations. This equation can then be used to calculate the original σ without the influence from the setup.
The reliability of the deconvolution algorithm and the statistic subtraction was assessed by comparing two measurements of the R5912 at +1425 V at the LNGS and in SPAX (occupancy 2.92% in SPAX, 1.83% at the LNGS). The measured total T_{FWHM} was 2.27 ns in SPAX and 2.04 ns at the LNGS. When statistically subtracting LED/laser jitter, optical jitter and electronics jitter, a value of 1.94 ns results for SPAX and 2.03 ns for the LNGS. The difference between both numbers is only 84 ps and likely originates from non-Gaussian contributions or errors introduced by the unfolding of the SiPM. It could also be that the SPAX errors are slightly overestimated or the LNGS errors were underestimated. When performing the deconvolution for the SPAX measurements, a TTS of 1.74 ns results (fig. 3.8), which is far lower than expected from LNGS data. The unfolding algorithm thus requires further work and was not used in the following.

Measurements

In order to measure a PMT, its VD and mu-metal screen were attached, the PMT was placed on the optical table, and the mu-metal was grounded. The PMT front was positioned in the cathode plane, the device was oriented along the optical axis, the cables were connected, and all surfaces in the dark box were covered with molleton or felt. Then the lid was closed and the sensor was tested for functionality and light leaks by carefully ramping up the voltage while monitoring the output on a digital oscilloscope.
Next, the electronics were adjusted. First, a discriminator gate length (CFD1) which is suitable for acquisition of the TTD (and EAPs) was selected using the oscilloscope. Delay1 was adjusted until the pulser trigger was at the position on the gate which produced the desired EP and LP windows, including a window of 10–20 ns before the EPs to determine the baseline. The pulser period was set and measured. Following this, the HV was adjusted to the target gain (using (2.102) in a spreadsheet) with the LabVIEW script. Amplifier channels were added as needed between PMT and FADC (for $g \geq 10^7$ no amplifiers were needed). The electronics noise was checked, interference reduction measures were carried out as needed, and the discriminator threshold was set as low as possible above the pedestal. For AP-coincidence measurements also the thresholds of CFD2 and CFD3, Gate1 and Delay2 were adjusted. The data acquisition with Acqiris Live (software from the FADC manufacturer) was prepared by adjusting the program settings.
Now, the occupancy o was adjusted to a suitable level of a few percent (see p. 92) by replacing the NDFs in the dark box and using the pulse counting method (see p. 94) in the LabVIEW script.
After a sufficient cool-down time with applied HV for the PMT, the DCR was measured with the scaler and the occupancy was determined again. The FADC was recalibrated after having reached a stable temperature, the measurement properties

were noted down (HV, threshold, occupancy, temperature, DCR, time in the dark), and the acquisition was started. Due to the analog coincidence, the pedestal in the SER is not recorded, which has the advantage of a far higher statistics per acquisition time; otherwise, about 95% of events are only usable to obtain the occupancy. On the other hand, without the pedestal, o has to be measured separately, which here was done with the LabVIEW script. The occupancy can change slightly during a measurement due to drift of the light source and PMT, but the mean o can still be obtained in the offline analysis by fitting the SER (see p. 99). This was not necessary here, since o was sufficiently stable.
For an AP-coincidence, in addition, the AP probability was measured with the scaler just before the start of data taking.
After data acquisition was completed, an electronics measurement was carried out as described on p. 336 to determine the transit time offset and electronics jitter.

Evaluation

Concomitantly with the development of SPAX, a dedicated evaluation software for the waveforms acquired with the FADC was written based on C++, Bash and the ROOT library [463, 464]: the Photosensor Evaluation Software Toolkit, or short PEST. In 2017 a revised object-oriented class layout was implemented and the original version (v1.0) was discontinued. In its current version (v3.6), PEST has become a generic all-round program for the analysis of arbitrary waveforms or photosensor measurements. Including the associated ROOT++ library, which contains stand-alone and ROOT-based functions for statistic data analysis, PEST has over 56000 lines of code.
The software is based on the contents of ch. 2 and all results shown there and in the current chapter were produced with PEST. The LNGS measurements were analyzed using ROOT and a subset of the PEST routines, since in these analyses data was available only in form of histograms.
The most important features of PEST comprise:

- **Signal processing and waveform analysis**
  - Waveform scaling to compensate amplification/attenuation (p. 128)
  - Subtraction of the baseline offset (p. 124) and of an average empty waveform (correlated noise, p. 341)
  - Noise cuts (baseline RMS and mean)
  - Detection of clipped waveforms
  - Smoothing (Savitzky-Golay or moving average, pp. 129 and 340)
- **Pulse shape analysis**
  - Threshold-based pulse search using a Schmitt trigger (p. 127), a constant fraction (CF) threshold to determine the time of a pulse (p. 335), and noise recognition (from the charge–height ratio and statistic analyses)
  - Average pulse shape (APS) determination (p. 339) using cuts for the selection of included pulses

  - Pulse-fitting with the APS [153; pp. 658–659] or an arbitrary pulse shape (see also 2.3.6 and 2.4):
    - ▷ Intelligent start parameter selection
    - ▷ Recognition of overlapping pulses (pileup) down to peak distances of 0.17 pulse widths (0.8 ns for the R5912) and charges differing by a factor of 40
    - ▷ Automatic determination of the number of overlapping pulses and improvement of the fit parameters using the reduced χ^2 of previous fits
    - ▷ Pulse subtraction from the waveform (fitted pulses or simple subtraction of a height/charge-scaled APS)

- **Timing corrections**
  - Transit time offset (see p. 336)
  - DCR subtraction (p. 163)
  - Unshadowing (p. 164)
  - Deconvolution of light source ETD (experimental, see pp. 167 and 479)

- **Output and measured parameters**
  - 1D and 2D histograms for:
    - ▷ PPs: transit time, pulse shape parameters (rise time, fall time, pulse width, pulse length), charge, pulse height
    - ▷ APs: same as PPs plus charge fraction; histograms for the first AP and all APs
    - ▷ Waveforms: mean value and RMS of the baseline window

    The charge can be scaled to pe. All plots are saved as png images.
  - TST results are read from a csv file and plotted into copies of the corresponding histograms (see 2.3.6 and 2.5.2.2).
  - Persistent plot of all waveforms (see p. 339 and fig. 2.51)
  - Can plot individual waveforms to better analyze effects (see figures in 2.4).
  - Determined photosensor properties:
    - ▷ Gain, mean charge and spe charge (from clean PPs using cuts, see pp. 261 and 488) with occupancy correction (pp. 97, 137, 128)
    - ▷ P/V (p. 122)
    - ▷ DCR (pp. 162 and 350)
    - ▷ EP and LP rates (p. 255)
    - ▷ AP rate properties (see 2.5.2.1), including the different photonic afterpulse rates (p. 380) and the remaining fractions of EAPs and PPs vs. threshold (p. 377)

    All results are written to a log file.

- **Customization**
  - Cuts can be applied on all waveform and pulse properties for the filling of the histograms and the waveform plots.
  - At the moment 551 options are implemented.
  - The program flow is directed via a control file.
- **Performance**
  - Multithreading
  - Batch-processing of several measurements; remote control option to suppress graphics display
  - Preprocessing: The results from the time-consuming waveform analysis and pulse search are saved in a ROOT-file and during analysis only the evaluation step (see below) has to be repeated.
  - Can concatenate data files for faster readout.

  These features significantly speed up the analysis. During processing (see below) PEST is only limited by the hard disk readout speed if the pulses are not fitted (200 GB in 29 min for an external hard disk → 115 MB/s). Due to the preprocessing, a complete evaluation is finished within typically 1–2 min.
- **Applicability:** PEST was written for the FADC type Acqiris DC282, but is usable for every setup (and completely different experiment types) with slight modifications or additions, since it is a generic pulse shape analysis software.

A side benefit of pulse-fitting is that it should allow the separation of 1-pe and 2-pe pulses more reliably than a simple charge cut. This was not tested so far, but would be interesting to study, as it could be employed in experiments such as JUNO or THEIA to improve the charge resolution.

Analysis procedure

The basic procedure of an analysis with PEST is described in the following.
First, the waveforms are processed ("processing"). Quality cuts are applied (baseline RMS and mean, limit the TTD to the acquisition window), the thresholds for PPs and APs are set, and the gain is determined. In a next step, the average empty waveform is obtained from waveforms where the PP is very late or absent (fake triggers) and then subtracted from all waveforms. Subsequently, the APS is determined iteratively (described on p. 339) and the CF point of the pulse time is set. Once a reliable APS is obtained which can not be improved further, the pulses are fitted with the APS. For this, first the cuts used to identify pileup and the χ^2 thresholds for the escalation of the fits (n → n+1 pulses) are selected using a sufficiently large fraction of all waveforms. In the next iteration, the fitted pulses are subtracted, which reveals further pulses which were previously hidden by oscillations. Now, the fit parameters are finetuned (minimum resolvable height and time difference, noise cuts). When this is finished, all waveforms are fitted and subtracted in this manner, which is the

part of the processing which is numerically most costly. The results are saved in a ROOT-file so that this step does not have to be repeated.
The second part of the analysis uses the processed waveforms to evaluate the measurement and produce most plots and measured parameters ("evaluation"). Here, first the time offset is corrected using the previously analyzed corresponding electronics measurement. In addition, the pulse time is shifted to the peak (see p. 154), the TTD and DD are unshadowed, the DCR is subtracted from both, and optionally the ETD is unfolded. Based on the processed histograms the EP and LP rates, AP rate parameters, charge resolution, P/V, etc. are obtained through an iterative adaptation of cuts and fits.
Following the completed evaluation, a "classic" analysis is carried out based on the last settings in the control file. In this analysis the pulse fits and pulse subtraction are not used and it thus can only recognize a part of the effects which were predicted by TST and observed with SPAX + PEST. The classic analysis is carried out to be able to compare the results of a full analysis to other setups which are pileup-agnostic, such as the one used at the LNGS.

3.2.2 Results

In the following, the results of the PMT characterization campaigns with SPAX and at the LNGS shall be presented.
The findings are arranged in sections which correspond to property groups, following the structure of ch. 2. In every section the measured distributions which best characterize the behavior are shown and the main properties are summarized in a table. In the text, for each characteristic the obtained values are compared for different PMT types, voltages (where applicable), setups, and evaluation methods (classic and pulse-fitting) as well as to literature. To describe occurring effects, the TST nomenclature and calculations are used. TST is introduced in 2.3.5, while it is summarized on p. 197 where the respective resources about time shift effects (abbreviations, schematic drawings, waveforms, detailed descriptions and calculated properties) can be found.
The most reliable measurements whose gain was closest to $1 \cdot 10^7$ were used for the final results, which are summarized in the next section, 3.3. For better comparability with literature, here values from *classic* analyses were taken. While pulse-fitting offers much more information and allows, for instance, a better discrimination of 1- and 2-pe events, it is not yet implemented in most current event reconstruction routines and is not used for photosensor characterizations in literature so far.
For the PMTs which were measured with SPAX, the gain was also varied to study the effect of a different HV and gain on sensor properties. The knowledge of the behavior for a gain range permits to adapt the target gain if this should prove to be beneficial for the physics agenda.

In SPAX timing and EAP properties were determined in the same measurement, since the required acquisition parameters are similar. The same applies to iAP and LAP characteristics.

Overview of measurement and evaluation parameters

Gain

For the baseline timing and iAP measurements which were performed for all PMT types (for PMTs studied with SPAX also EAPs and LAPs), the gain was set to $1 \cdot 10^7$ to the extent possible. This was done by applying the following voltages: R6594 +1670 V, R5912 +1425 V, R7081 +1725 V, R11780-HQE +1648 V.
As mentioned in 2.2.4, the gain measurement – and thus also the gain adjustment – suffer from a considerable error for all methods. In addition, the gain drift of the PMT, uncertainties and fluctuations of the amplification or attenuation in the setup, the discriminator threshold, and the baseline offset contribute to this error.
As was discovered later during evaluation, in the LNGS measurements for the R5912 and the R7081 a gain of about $1.3 \cdot 10^7$ was set due to an incorrect calibration of the QDC in the original state of the setup. Since the value is close to the target gain of $1 \cdot 10^7$, the corresponding voltage was reproduced in the SPAX measurements to allow to compare results with the other setup.

For the R5912 in addition timing and EAP measurements with gains of about $5 \cdot 10^6$ (+1275 V) and $5 \cdot 10^7$ (+1725 V) and iAP and LAP acquisitions with $5 \cdot 10^7$ were performed. For the R11780-HQE additional measurements were carried out with approximately $5 \cdot 10^7$ (+2092 V) for timing, EAPs, iAPs and LAPs.

Voltage divider

As voltage divider the Hamamatsu E7694-01 (positive HV, front-tapered) was used for both the R5912 and R7081. The R6594 was fitted with a circuit board soldered to the PMT leads (Hamamatsu assembly ref. no. 6594-091009, front-tapered; originally negative polarity, modified by G. Korga for positive voltage). The R11780-HQE was measured with a Hamamatsu E7693-01 MOD [27] (positive HV, front-tapered, enlarged $R_{10,a}$).

Amplification

To reduce errors, in SPAX measurements no preamplifiers were used in the signal branch. Since a part of the signal was split off for the CFD branch (see fig. 3.2), a total amplification of 0.804 was measured.
At the LNGS, amplifiers and attenuators with a total amplification of 35.5 (R7081), 36.2 (R5912) and 48.0 (R6594) were employed.

Noise

In SPAX the measured baseline RMS ranged from 0.24 to 1.45 mV for timing and EAP measurements (R5912: +1275 V 0.24 mV, +1425 V 0.31 mV, +1725 V 1.29 mV; R11780-HQE: +1648 V 0.34 mV, +2092V 1.45 mV).
The R11780-HQE showed no broader noise bands than the R5912 despite having a cathode area (i.e. antenna) which is 1.91 times as large. This confirms that the shielding of both the Faraday cage and the light source inside it is effective and no significant amount of noise is picked up.

[27] Repaired by L. Papp since the supplied VD did not produce output signals due to an incorrect internal connection.

The correlation of the noise RMS with the gain points to a contribution from frequent strongly underamplified late-stage dark counts. The PMTs did not pick up noise from the light source, the signal branch was not modified between measurements with $1 \cdot 10^7$ and $5 \cdot 10^7$ gain, and the noise correlates with gain. Therefore, the amplification process of the PMT, which is based on the emission of electrons inside the dynode chain, must be involved in the production of this noise source. Since the noise is not correlated with the signal and has very small pulse heights, this indicates thermionic emission from the middle to late dynodes.

For iAP and LAP measurements in SPAX, the noise RMS was higher due to the additional electronics used for the triple coincidence. Several of its components showed signs of instability due to old age, but unfortunately no replacements were available at that time. This also affected measurements where the coincidence was not used ("simple" AP measurements), since the modules were installed in the same NIM rack and polluted the common ground. The resulting baseline RMS ranged from 1.40 to 5.52 mV (R5912: +1425 V coinc. 1.41 mV, simple 1.40 mV, +1725 V coinc. 5.52 mV; R11780-HQE: +1648 V coinc. 1.40 mV, simple 1.44 mV, +2092 V 5.50 mV). Again, the R11780-HQE showed no stronger noise pickup despite its larger size.

Threshold

For the LNGS measurements a threshold of about 0.2 pe (mean height) was used [450].

In SPAX measurements the charge and height corresponding to 1 pe were set to the peak of the clean PP distribution (peak of the normal PP distribution for classic evaluations), since this is the best approximation of the γk charge. As detailed in 2.3.6.1, γk is the foreseen regular amplification process and is used as a reference for TST analyses. Since this definition of 1 pe results in lower pe-thresholds than when using the mean value, the height and charge are listed in the tables for both definitions to allow comparison. For iAP and LAP measurements the pe-height of the corresponding timing measurement was used, because the higher sampling rate provides a more precise value.
Due to the low noise RMS, which resulted from the installation of the Faraday cage, for the timing and EAP measurements low pulse search thresholds (identical for PPs and APs) between 5.4% pe and 13.4% pe could be achieved (see table 3.2). Through pulse-fitting the threshold could be further lowered to 1.2–2.8% pe.
In the iAP/LAP measurements the higher noise RMS required to to use larger thresholds – 0.28 to 0.56 pe (PP) and 0.34 to 0.69 pe (AP) – than for the associated timing measurements, which excludes many underamplified pulses. The threshold could be lowered in future measurements by exchanging the defective components. This is expected to reveal an even higher contribution from interdynode APs than already visible when comparing the measurements for the R5912 at +1425 V at the LNGS (0.2 pe) and with SPAX (0.34 pe) in fig. 3.23 for delays below 1 µs.
The R11780-HQE has very shallow, broad pulses – for $g = 10^7$ the pulse height amounts to only 8.7 mV (mean) and 10.8 mV (peak for clean PPs). This results in higher effective pe-thresholds than for the R5912 at comparable gains. For all PMTs an increase of the gain is found to reduce the pe-threshold, since the height scales with the gain, while the contributions from noise components are unaffected.

Temperature

In the LNGS measurements the temperature was maintained at about 23 °C through air conditioning.
In SPAX also air conditioning was used and in addition the temperature in the dark box was measured. The temperature lay between 21.8 and 23.1 °C for all shown SPAX measurements during the course of seven weeks (see table 3.6), which confirms that the measures which were taken for temperature stabilization were sufficient.

Occupancy

In the measurements shown here, the occupancy was adjusted to values between 1.83% and 5.53% (LNGS) and 2.03% to 3.47% (SPAX); the rationale for this choice of values is discussed on pp. 90, 91 and 92.

Time in the dark

Due to the limited available measurement time at the LNGS, the PMTs spent about 3 to 11.5 h with applied HV in the dark before the start of the measurement.
In SPAX for the R5912 4.7 to 17.3 h were allowed to pass before data acquisition. The R11780-HQE showed an extremely high initial DCR of about 3 MHz for which reason an extended cool down period of 5.1 to 21.9 d was required before measurements with acceptable dark noise contaminations (around 10 kHz) could be carried out.

Magnetic compensation and orientation

In the LNGS an active compensation with a coil system was used (see p. 459) and the PMTs faced upward.
In SPAX a passive mu-metal shielding was applied. For the R5912 an ETE MS200A shield for 8"-PMTs was used. For the R11780-HQE a custom-made shield was fabricated in-house (see p. 477), since no commercial shielding of sufficient size was available. The PMTs were facing west due to the orientation of SPAX, which was determined by the layout of the laboratory.

Light source and optics

For the LNGS measurements the PMTs were illuminated by a ps diode laser coupled to a fiber ending in a transmissive diffuser ball (see p. 457).
In SPAX a fast LED driven by a Kapustinsky pulser was reflected off a Spectralon plate (see p.460).

Setup jitter

For the LNGS a total setup jitter FWHM of 215 ps (R6594) to 217 ps (R5912 and R7081) resulted from the Gaussian sum of ETD, light source trigger jitter, time of flight differences and electronics jitter.
For SPAX a setup jitter FWHM of 1.170 ns (R5912), 1.204 ns (R11780-HQE +1648 V) and 1.199 ns (R11780-HQE +2092 V) was determined.
These values were quadratically subtracted from the measured TTS variants T_{FWHM} and σ_T (scaled using (2.68)) to estimate the true TTS for zero setup jitter. The values after quadratic subtraction are denoted by the subscript s in the variable name.

Sampling rate

An FADC sampling rate of 8 GS/s was used for all SPAX timing/EAP measurements. For the SPAX iAP/LAP runs the FADC memory and the far longer acquisition window limited the rate to 1 GS/s.

Pulse time

In the PEST evaluation of SPAX data, the pulse times were determined using a software CF search with the CF-point set at the steepest position in the rising flank of the APS. The obtained pulse time was corrected to correspond to the peak by adding the difference between CF-point and peak in the APS. Interestingly, the resulting CF-point was quite consistent for all SPAX measurements and ranged between 54% (R11780-HQE +1648 V timing and iAPs, +2092 V iAPs) and 55% (rest) pulse height.

Pulse-fitting

In the PEST pulse search via fitting (pulse-fitting), which is the base for all TST results, the minimum pulse distances were between 0.7 ns (R5912 +1725 V) and 0.8 ns (all others). The minimum pulse heights ranged from 0.26 to 1.35 mV (R5912: +1275 V 0.26 mV, +1425 V 0.35 mV, +1725 V 1.35 mV; R11780-HQE: +1648 V 0.3 mV, +2092 V 1.2 mV).
The resulting mean reduced χ^2 of the pulse fits lay between 1.22 and 1.34 (R5912: +1275 V 1.22, +1425 V 1.24, +1725 V 1.34; R11780-HQE: +1648 V 1.32, +2092 V 1.24). The fits were performed with a single APS model, which is equivalent to assuming that the γk APS can describe all pulses independently of their charge and origin effect. The fraction of pulses which were discarded due to a too high reduced χ^2 was very low (0.06% to 0.76%). Based on these quality factors the fit accuracy was considered acceptable to obtain reliable results.

Clean primary pulses and pulse clusters

As was found in 2.3.6 and 2.5.2.2, the signal which a PMT produces in response to a photon typically does not consist of one but multiple pulses from various high-probability TSEs which produce EAPs (called sum-pulse here); the probability for pileup ranged from 63 to 68% in the measurements reported here (see table 3.6).

In the SPAX analyses, the knowledge of the waveform and the contained pulses which is gained through pulse-fitting enables to select the primary pulses which do *not* suffer from pileup or subsequent afterpulses. These are the PPs for which the quality parameters of pulse subtraction [28] do not indicate undetected pulses in the residual waveform and for which no further pulse is found in the waveform (no APs). These selected PPs are referred to as "clean" primary pulses in the following, since due to the exclusion of all EAP-producing effects they allow to obtain the γk properties (charge, height, transit time) far more accurately than from the "regular" PPs. It is important to note that a selection of clean PPs does not only leave clean RPs (i.e. γk), as e.g. a γ1 will also pass all pileup and EAP cuts and enter the clean PP distributions. Undersized pulses without afterpulses will be present in the selection, while oversized pulses from pileup will be eliminated. This is visible in the SER and TTD plots shown in the following.

[28] Cuts for pileup detection which are based on a comparison of the absolute height and the relative height and charge of the residual values with the subtracted pulse and the noise band.

For classic analyses a selection of clean PPs is not meaningful. This is because no pileup can be detected and an AP exclusion is not reliably possible if APS oscillations are present, since these cause a chain of fake APs for pulses for which the oscillation peaks exceed the pulse search thresholds. This would introduce a strong bias towards small pulses.

PEST also allows to group the pulses which are found with fitting to clusters, which contain all fitted pulses which overlap between 10% height on their rising and falling flanks. This enables to reconstruct a classic, pulse shape-agnostic analysis, where overlapping pulses can not be distinguished and are counted as one pulse, from evaluations which use pulse-fitting. The obtained cluster values closely resemble those of pulses in a classic evaluation but are not completely identical due to several factors:

- The charge is integrated between the first and last datapoints above the noise band. For the sum-pulse only the first peak of the APS, which extends until the first zero-crossing, can be considered, while for the constituent pulses the complete APS is taken into account through an extrapolation factor. On the other hand, the constituent pulses are smaller, which reduces the usable integration range due to the larger influence of the noise band.
- The 10%-height times which are used to obtain pulse shape parameters vary between sum-pulse and constituent pulses.
- The (one) APS used for fitting might not perfectly describe all pulse types from different TSEs. This affects the charge, height, time and pulse shape parameters.
- The cluster time is set to the charge-weighted average peak time and not the time of the sum-peak, which strongly depends on pileup.

The first cluster in a waveform is referred to as primary cluster (PC), subsequent clusters as secondary clusters.

3.2.2.1 Photon counting

The distributions which summarize the results regarding gain, charge and pulse height are shown in fig. 3.9 (SER for LNGS measurements), fig. 3.10 (charge distributions for SPAX), and fig. 3.11 (height histograms from SPAX). The main findings are summarized in table 3.2.

Gain

Reference gain

The gain is defined as the mean primary pulse charge $\overline{q}_{\mathrm{PP}}$ divided by the elementary charge (eq. (2.90)). This charge definition takes into account all time shift effects and gives the most comprehensive picture of PMT charge behavior.
Unsurprisingly, the gain values resulting from $\overline{q}_{\mathrm{PP}}$ are very different for fitting and classic analyses, since pulse-fitting detects many small constituent pulses, which results in a far lower gain.

Series	**R6594**	**R5912**				**R7081**	**R11780-HQE**	
Voltage [V]	+1670	+1275	+1425		+1725	+1520	+1648	+2092
Setup	LNGS	SPAX	SPAX	LNGS	SPAX	LNGS	SPAX	SPAX
Baseline RMS [mV]		0.24	0.31		1.29		0.34	1.45
Threshold		*11.6*	*6.2*		*5.4*		*13.4*	*10.9*
[% pe]	20	11.4	6.4	20	6.2	20	12.3	9.7
Fit threshold [% pe]		*2.7*	*1.4*		*1.2*		*2.8*	*2.3*
Voltage divider	6594-091009	E7694-01	E7694-01	E7694-01	E7694-01	E7694-01	E7693-01-MOD	E7693-01-MOD
Occupancy [%]	5.53	3.14	2.92	1.83	3.29	2.91	2.03	2.09
Time in the dark	9 h	30.8 h	4.7 h	3 h	17.3 h	11.5 h	17.0 d	18.0 d
DCR [kHz]	4.95	0.51	0.77	1.51	1.12	2.46	6.32	9.06
Temp. [°C]	23	21.8	22.4	23	23.1	23	22.0	21.8
N	$1.73 \cdot 10^7$	$5 \cdot 10^6$	$5 \cdot 10^6$	$1.52 \cdot 10^7$	$5 \cdot 10^6$	$3.19 \cdot 10^6$	$5.02 \cdot 10^6$	$5.02 \cdot 10^6$
Gain [10^7]: from								
$\overline{q}_{\mathrm{PP}}$		*$3.79 \cdot 10^6$*	*$8.87 \cdot 10^6$*		*$3.58 \cdot 10^7$*		*$7.47 \cdot 10^6$*	*$3.39 \cdot 10^7$*
	$9.64 \cdot 10^6$	**$6.11 \cdot 10^6$**	**$1.35 \cdot 10^7$**	**$1.29 \cdot 10^7$**	**$5.42 \cdot 10^7$**	**$1.34 \cdot 10^7$**	**$1.01 \cdot 10^7$**	**$4.60 \cdot 10^7$**
$\overline{q}_{\mathrm{PPc}}$		*$4.89 \cdot 10^6$*	*$1.08 \cdot 10^7$*		*$4.63 \cdot 10^7$*		*$7.76 \cdot 10^6$*	*$3.39 \cdot 10^7$*
$\overline{q}_{\mathrm{PC}}$		*$5.62 \cdot 10^6$*	*$1.30 \cdot 10^7$*		*$5.08 \cdot 10^7$*		*$1.02 \cdot 10^7$*	*$4.69 \cdot 10^7$*
$q_{p,\mathrm{PP}}$		*$5.52 \cdot 10^6$*	*$1.31 \cdot 10^7$*		*$5.28 \cdot 10^7$*		*$9.55 \cdot 10^6$*	*$4.44 \cdot 10^7$*
	($1.05 \cdot 10^7$)	$6.82 \cdot 10^6$	$1.57 \cdot 10^7$	($1.41 \cdot 10^7$)	$6.34 \cdot 10^7$	($1.43 \cdot 10^7$)	$1.05 \cdot 10^7$	$4.90 \cdot 10^7$
$q_{p,\mathrm{PPc}}$		*$5.87 \cdot 10^6$*	*$1.50 \cdot 10^7$*		*$6.09 \cdot 10^7$*		*$9.28 \cdot 10^6$*	*$4.42 \cdot 10^7$*
$q_{p,\mathrm{PC}}$		*$6.22 \cdot 10^6$*	*$1.49 \cdot 10^7$*		*$5.85 \cdot 10^7$*		*$1.06 \cdot 10^7$*	*$4.95 \cdot 10^7$*
Charge [pC]:								
$\overline{q}_{\mathrm{PP}}$		*0.61*	*1.42*		*5.74*		*1.20*	*5.42*
	(1.54)	**0.98**	**2.17**	**(2.06)**	**8.69**	**(2.15)**	**1.61**	**7.37**
$\overline{q}_{\mathrm{PPc}}$		*0.78*	*1.74*		*7.42*		*1.24*	*5.43*
$\overline{q}_{\mathrm{PC}}$		*0.90*	*2.08*		*8.14*		*1.64*	*7.52*
$q_{p,\mathrm{PP}}$		*0.88*	*2.10*		*8.45*		*1.53*	*7.12*
	(1.68)	1.09	2.51	(2.25)	10.16	(2.30)	1.68	7.85
$q_{p,\mathrm{PPc}}$		***0.94***	***2.40***		***9.77***		***1.49***	***7.08***
$q_{p,\mathrm{PC}}$		*1.00*	*2.39*		*9.38*		*1.69*	*7.93*
Resolution [%]:								
$R_{F,m,\mathrm{PP}}$		-	-		-		-	-
	(60.3)	**87.3**	**86.6**	**(68.2)**	**85.6**	**(66.9)**	**87.9**	**88.6**
$R_{F,m,\mathrm{PPc}}$		*95.8*	*95.1*		*81.0*		-	-
$R_{F,m,\mathrm{PC}}$		*85.2*	*85.1*		*81.8*		*85.9*	*85.6*
$R_{F,p,\mathrm{PP}}$		-	-		-		-	-
	(55.3)	**78.3**	**74.8**	**(62.5)**	**73.2**	**(62.5)**	**84.2**	**83.1**
$R_{F,p,\mathrm{PPc}}$		*79.8*	*68.8*		*61.5*		-	-
$R_{F,p,\mathrm{PC}}$		*77.0*	*74.0*		*71.0*		*83.0*	*81.2*
$R_{\sigma,\mathrm{PP}}$		*67.9*	*70.6*		*69.2*		*63.2*	*64.5*
	(35.1)	**46.6**	**49.9**	**(35.5)**	**50.1**	**(37.4)**	**51.0**	**55.4**
$R_{\sigma,\mathrm{PPc}}$		*49.0*	*61.4*		*57.3*		*60.3*	*66.9*
$R_{\sigma,\mathrm{PC}}$		*44.7*	*48.7*		*48.7*		*48.8*	*52.0*
F_{PP}		*1.46*	*1.50*		*1.48*		*1.40*	*1.42*
	(1.12)	**1.22**	**1.25**	**(1.13)**	**1.25**	**(1.14)**	**1.26**	**1.31**
F_{PPc}		*1.24*	*1.38*		*1.33*		*1.36*	*1.45*
F_{PC}		*1.20*	*1.24*		*1.24*		*1.24*	*1.27*
$\mathrm{P/V}_{\mathrm{PP}}$		*1.33*	*1.16*		*1.34*		*1.83*	*1.80*
	3.76	**2.84**	**2.83**	**2.85**	**2.88**	**2.79**	**2.89**	**2.96**
$\mathrm{P/V}_{\mathrm{PC}}$		*2.72*	*2.85*		*2.96*		*2.79*	*2.86*
Height [mV]:								
$\overline{h}_{\mathrm{PP}}$		*6.3*	*14.5*		*65.9*		*8.7*	*40.5*
		8.7	**19.8**		**86.7**		**11.5**	**54.7**
$\overline{h}_{\mathrm{PPc}}$		*8.1*	*17.8*		*85.1*		*9.0*	*40.5*
$h_{p,\mathrm{PP}}$		*9.0*	*21.2*		*94.6*		*11.1*	*52.1*
		9.7	23.5		97.3		11.7	57.7
$h_{p,\mathrm{PPc}}$		***9.5***	***24.3***		***110.9***		***10.8***	***51.5***

Table 3.2: PMT characterization results: Gain, charge and pulse height.
Measurement parameters above line, results below. Main results bold. Discussion see text. Italic values from pulse-fitting, normal font from LNGS or SPAX with classic analysis; if both are available, fitted values in first row, classic in second, identical values in one row with normal font. Uncertain values in parentheses (see p. 496).
PMTs and voltage dividers from Hamamatsu. Mean baseline root mean square (RMS) listed to indicate extent of pedestal. Threshold as percent of primary pulse peak height (for SPAX clean PPs). DCR: For SPAX from TTD using classic evaluation, for LNGS from iAP DD. N is the number of acquisitions.
Results for primary pulses (PPs), clean PPs (no pileup or APs; subscript PPc) and PCs (primary clusters; clusters are the sum of all fitted pulses which overlap between 10% height on rising and falling flank). $\overline{q}$ mean charge, q_p peak charge, $\overline{h}$ mean height, h_p peak height. Resolution: $R_{F,m}$ FWHM divided by mean charge, $R_{F,p}$ FWHM by peak charge, R_σ standard deviation by mean charge; no value for $R_{F,m}$ and $R_{F,p}$ if $\mathrm{P/V} < 2$ (FWHM not defined). F denotes the excess noise factor (ENF), P/V the peak-to-valley ratio.

The gain from the classic mean PP charge was selected as reference gain since, this is the definition used throughout literature. In view of the findings of TST this is not quite accurate though, since it is the gain of the primary sum-pulse and not of the actual primary pulse. However, the constituents of the sum-pulse can not be resolved without pulse-fitting and so for most practical purposes the effective gain of the sum-pulse is relevant. The TSEs in this case constitute additional gain paths for the true PP, which actually *reduces* the charge variation (compare classic and fitting R_σ PP values in table 3.2). This is because branching effects are included in the classic sum-PP when they lead to pileup, whereas such effects reduce the charge of the true PP, which is found via fitting.

Classic mean primary pulse gain

The classic gain from $\overline{q}_{\mathrm{PP}}$ is close to the target gain (see p. 485) with deviations of 1.0% to 22.2% (except for the R5912 at +1425 V and the R7081 where the different gain had another reason) and a mean deviation of 8.6%.
At the LNGS the gain was set using the SER peak and empirical values of the factor between mean and peak charge which were obtained in preceding measurements. In SPAX the gain was adjusted with the mean charge using the LabVIEW script. While the gain setting accuracy is within typical values, it would be desirable to try to improve the precision in SPAX by implementing another of the methods listed starting on p. 137.
The gain rises with HV as expected. Using (2.102) the gain–voltage power law growth coefficient β can be calculated from the outermost gain values. The resulting value is 7.22 for the R5912 and 6.37 for the R11780-HQE, which has a flatter curve and requires higher voltages to reach the same gain.
The classic gain values for the R5912 at +1425 V (same PMT, same HV) at the LNGS ($1.35 \cdot 10^7$) and in SPAX ($1.29 \cdot 10^7$) are in good agreement. The small difference can be caused by baseline offsets, gain drift, and the different measurement method (QDC vs. FADC and numeric integration). The agreement of both values, which were obtained with completely different methods, verifies the correctness and accuracy of the charge measurement with both setups.
From the manufacturer datasheets [315, 439], the R6594 is expected to reach $g = 10^7$ at 1775 V and from the gain–voltage curve in the same source, a gain coefficient β of 7.01 can be extracted. For both R5912 and R7081, $g = 10^7$ lies at 1505 V ($g = 1.3 \cdot 10^7$ at 1550 V) with $\beta = 7.69$ according to Hamamatsu. For the R11780-HQE $g = 10^7$ is predicted at 2005 V with $\beta = 7.25$. It is well-known that the dynode gains and the total gain at a fixed voltage vary considerably between series and even specimens of the same series. This is apparent when comparing the measured values to the datasheets: All sampled PMTs had larger gains than expected. The R6594 reaches $g = 9.64 \cdot 10^6$ already at 1670 V and the R5912 $g = 1.35 \cdot 10^7$ at 1425 V, while the R7081 (which uses the same VD as the R5912) attains $g = 1.34 \cdot 10^7$ only at 1520 V. The R11780-HQE showed $g = 1.01 \cdot 10^7$ at a voltage of as low as 1648 V. In the datasheets, β is smaller for the R11780-HQE than for the R5912, like in the measurements, but both measured values are lower.

Mean primary pulse gain from pulse-fitting

The gain of $\overline{q}_{\mathrm{PP}}$ from pulse-fitting is smaller than the classic one (factor 0.62–0.66 for R5912, 0.74 for R11780-HQE). The power law is slightly steeper for the

R5912 ($\beta = 7.43$), which might be an effect of different pe-thresholds, and nearly identical for the R11780-HQE (6.34). The factor between fitted and classic values is closer to 1 for the R11780-HQE, which results from a lower contribution of undersized pulses than for the R5912. This is also visible in fig. 3.15, the EP fraction vs. threshold, which drops far faster for the R5912.

Mean clean primary pulse gain

The values of $\overline{q}_{\mathrm{PPc}}$ lie between the numbers from classic and fitting $\overline{q}_{\mathrm{PP}}$, since pulse branching effects and their resulting pileup are excluded. The β values are 7.43 (R5912) resp. 6.18 (R11780-HQE); the decrease of the latter might also be due to varying thresholds. The ratio of $\overline{q}_{\mathrm{PPc}}$ to the classic $\overline{q}_{\mathrm{PP}}$ is 0.80–0.85 (R5912) and 0.77–0.74 (R11780-HQE), which shows that for the R11780-HQE the contribution of EAPs to the sum-PP is stronger than for the R5912. This agrees with the higher pileup probability found for the R11780-HQE (see table 3.6).

Mean primary cluster gain

The mean PC gain closely resembles that of classic PPs. For the R5912 the values are slightly smaller than for classic PPs, for the R11780-HQE a bit larger. This can be understood from the combination of the integration method and the overlap of EAPs: The APS extrapolation factor is smaller for the R5912 (0.873–0.915) than for the R11780-HQE (0.960–0.966), and in both cases is < 1 which reduces the PC pulse charges compared to classic PPs. This is compensated in part due to the recognition of EAPs which would lie in the first negative APS peak but are found through fitting. This raises the factor between mean PC gain and reference gain to 0.92–0.96 (R5912) resp. 1.01–1.02 (R11780-HQE).
β is very similar to the classic PPs (R5912 7.28, R11780-HQE 6.39).

Classic peak primary pulse gain

The classic peak PP gain $q_{p,\mathrm{PP}}$ is higher than the classic mean PP gain, since the influence of undersized pulses is mostly excluded when studying the peak. This applies to all peak-gain–mean-gain pairs.
The peak charge for the R5912 at 1425 V is lower in LNGS data (2.25 pC) than in SPAX (2.51 pC). The peak position is very sensitive to even small influences and the difference might result from a broader pedestal in the LNGS measurement, since the spe response is folded with the setup noise.
The gain coefficient β for the R5912 is 7.38, for the R11780-HQE 6.46. This is slightly higher than for the mean: The peak apparently grows faster than the average charge, from which can be concluded that the undersized fraction due to TSEs grows with voltage.
The ratio of $q_{p,\mathrm{PP}}$ to the classic $\overline{q}_{\mathrm{PP}}$ is higher for the R5912 (1.12–1.17) than for the R11780-HQE (1.04–1.07), since more undersized pulses are present in the R5912.

Peak primary pulse gain from pulse-fitting

The fitted peak PP gain $q_{p,\mathrm{PP}}$ actually lies *below* the classic mean PP gain due to frequent undersized EPs contributing to the former and EAP pileup on the latter. It is far closer to the classic peak than the fitted mean is to the classic mean, because undersized pulses influence the peak position less than the average.

For the R5912 $\beta = 7.47$, for the R11780-HQE 6.44. These values are higher than for the corresponding mean values for the same reason as discussed for the classic peak PP gain.
The ratio of the fitted peak PP gain to the classic $\overline{q}_{\mathrm{PP}}$ is 0.81–0.84 (R5912) and 0.91 (R11780-HQE), which is closer to 1 than the fitted mean and shows the same behavior.

Peak clean primary pulse gain

The peak clean PP gain is a very relevant quantity, as it is the best indicator for the γk charge (without a 2D fit in the charge–time distribution (CTD)) and thus is used as the value for 1 pe. It is close to the classic mean gain; the value is slightly higher for the R5912 (except for 1275 V) due to the exclusion of undersized pulses, and slightly lower for the R11780-HQE due to the elimination of pileup.
The ratio to the classic peak is below 1 (R5912 0.86–0.96, R11780-HQE 0.88–0.90) due to pileup exclusion.
The coefficient β is far higher (R5912 7.74, R11780-HQE 6.55) than for the classic mean gain due to the double elimination of undersized pulses by the peak and the clean PP selection: The gain would rise far faster if not for the growing fraction of TSEs in the SER.

Peak primary cluster gain

Finally, the peak PC gain is close to the classic peak (ratio for R5912 0.91–0.95, for R11780-HQE 1.01), with some exceptions which may be due to the factors listed on p. 489. The power law coefficients of 7.42 (R5912) and 6.47 (R11780-HQE) are very close to the classic ones.

Charge

Charge distributions

In fig. 3.9 the PP SERs measured at the LNGS for the R6594, R5912 and R7081 at gains of about 10^7 are shown. In fig. 3.10 the SERs for PPs, clean PPs, PCs and APs measured with SPAX for the R5912 and R11780-HQE for varying voltages are depicted using both pulse-fitting and classic analyses.

The charge distributions for classic PPs (LNGS and SPAX) show the familiar shape from fig. 2.21: The spe peak resembles a (compound) Poissonian distribution, the 2-pe population scales with the occupancy (eq. (2.18)), and a large fraction of underamplified pulses is present which creates a valley and a subsequent rise down to the effective charge threshold. Since the actual threshold is height-based, the corresponding charge boundary is smeared out.
For LNGS measurements the valley is barely contained; while the recorded SER extends to lower charges, below the shown region notable contributions from the pedestal distort the shape[29] which renders this range unusable.
The R11780-HQE shows a very broad spe distribution and no 2-pe peak is visible anymore, but a fluent transition is observed instead.

The PPs which are detected through pulse-fitting contain far more undersized pulses than classic PPs, since more small EPs are recognized. Therefore, the peak

[29] This was concluded from the extent of the left pedestal flank as well as a comparison with SPAX data for the R5912 at +1425 V.

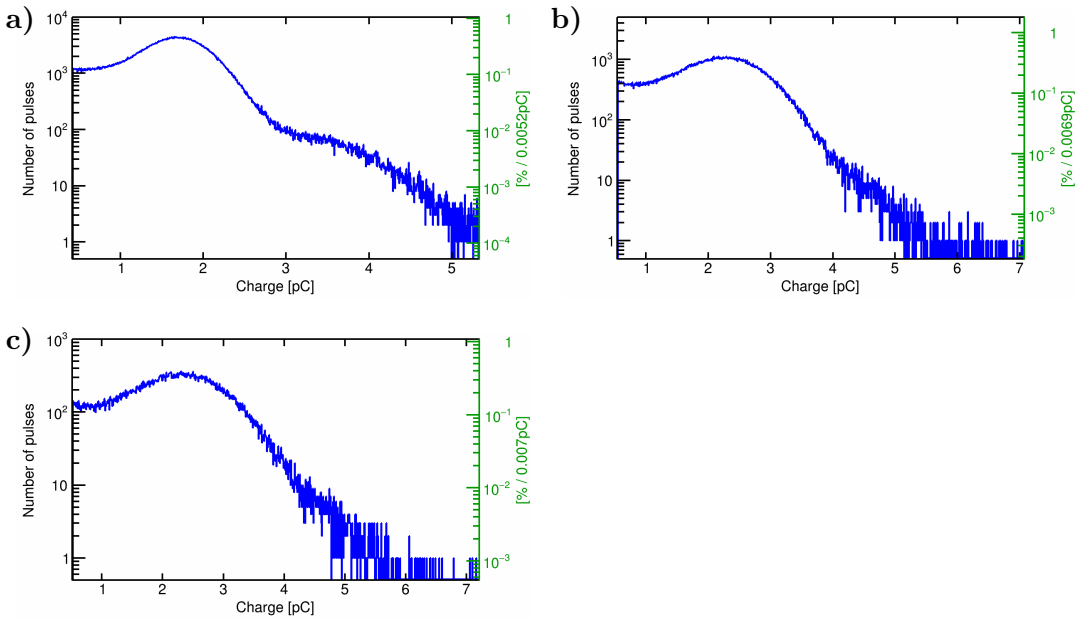


Figure 3.9: Charge results from the LNGS measurements: Single electron response (SER) for PPs. Fraction of all PPs per bin on right axis. Discussion see text.
a) R6594 +1670 V, **b)** R5912 +1425 V, **c)** R7081 +1520 V.

is very flat and its position is lowered; the exact position is not very accurately determinable as it is easily influenced by small effects. The valley is too flat to define an FWHM for all measurements, so no FWHM-based resolution values could be determined.
The 2-pe fraction is far higher in the SERs of classic PPs and PCs than those of fitted PPs and clean PPs. This shows the 2-pe discrimination power of pulse-fitting, which can separate two (or more) close PPs. This *also* means that one cannot use a fit of the SER of fitted/clean PPs to determine the occupancy accurately – instead classic PPs or PCs have to be used. Furthermore, the regular occupancy correction would reduce the mean charge of these distributions too much and a fit determining the remaining 2-pe population in the respective SER is required to calculate the correct correction factor.
Interestingly, for the R11780-HQE the peak showed a substructure with a sub-peak at lower charges. This is best observable for +1648 V in the clean PP clean and fitted AP distributions.

In the transition from fitted PPs to clean PPs for the R5912 mostly underamplified pulses are eliminated. For the R11780-HQE also more underamplified than overamplified PPs are removed, but the strongest relative difference occurs for regular charges around the peak. Thus, for the R11780-HQE a larger fraction of pulses at regular charges seems to stem from pileup than for the R5912.
Plotting the classic clean PPs is not reasonable, since large pulses have APS oscillations which exceed the threshold and are counted as fake APs, while small pulses do not. This would introduce a strong bias towards small pulses without gaining additional knowledge.

For primary clusters the SER shape is very similar to that of classic PPs. At very low charges more pulses are recognized for PCs by the fit routine, because

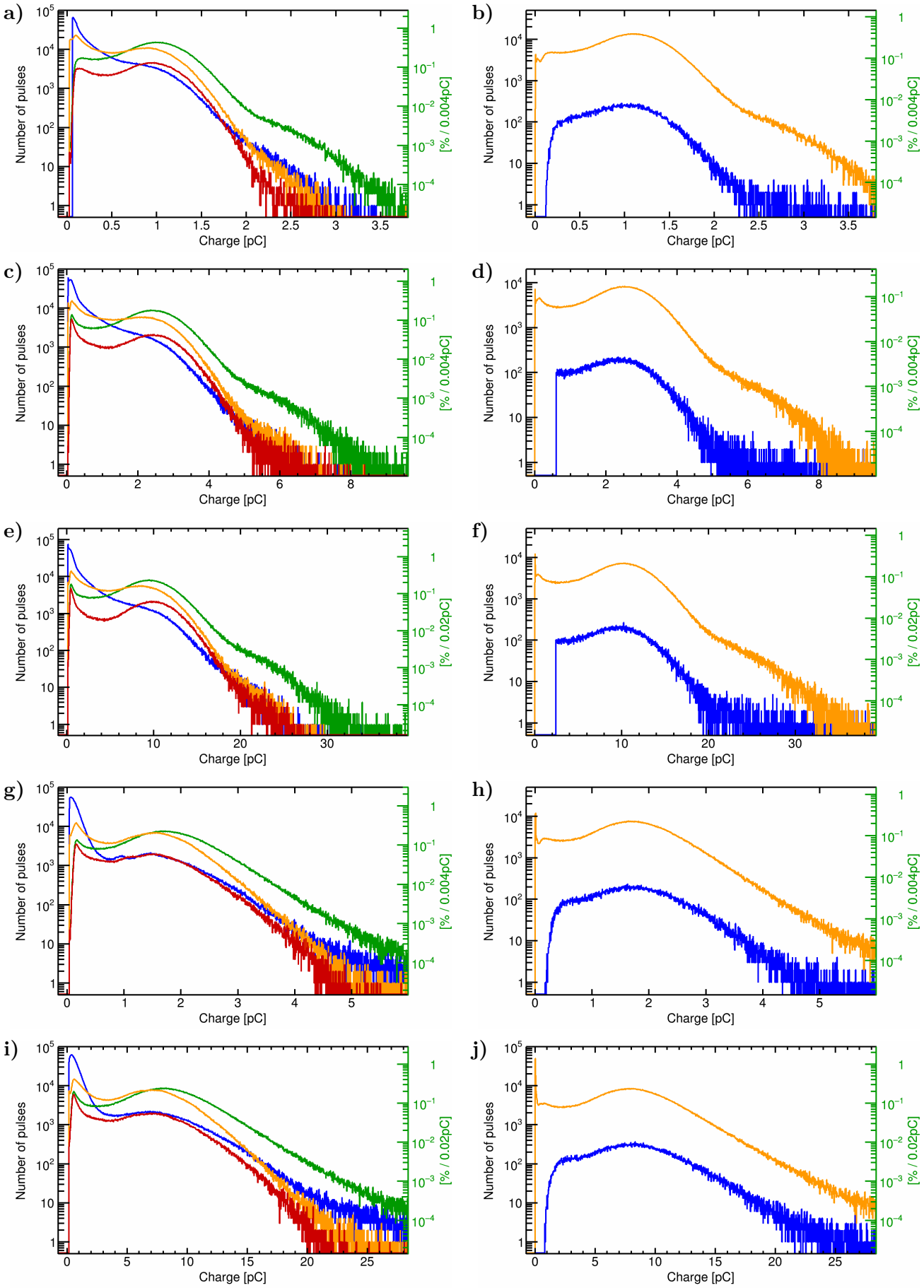


Figure 3.10: Charge results from SPAX: SER for clean PPs (red), PPs (orange), PCs (green, only for pulse-fitting), and APs (blue). Fraction of all PPs per bin on right axis. Figures on left from pulse-fitting, on right from a classic pulse search for the same measurement. Discussion see text.
a)–f) R5912: a) and b) +1275 V, c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

the fit threshold h_f is smaller than the hardware (setup) threshold h_s. However, since not all pulses which are larger than h_f but smaller than h_s can be found (see p. 267), this only constitutes a lower limit for the true number.

The AP charge distributions which result from pulse-fitting are strongly biased towards small charges.
For the R5912 no spe peak occurs anymore. Several subcomponents are clearly discernible, including a peak at the lowest charges which results from a missed amplification on the first dynode (relative charge $1/\delta 1$) and more skipped gain steps[30], as well as quasi-exponential decays at intermediate charges.
For the R11780-HQE the peak at the smallest charges is particularly large, but less events are encountered at intermediate charges, which is why the peak at regular charges is still discernible. At 1648 V an extra peak appears below the PP peak, for which the most likely candidate effect is BE1 (see p. 284). The localized peak in this case might be due to a combination of two effects: First, a stronger suppression of the BE1 PPs with large earliness than in the R5912, which would eliminate small APs. And second, the invisibility of APs with small delays and large charges due to the resolution of the fit routine. For larger voltages the delays shrink and the effect might slip below the temporal resolution and disappear, which could explain the absence of the peak for 2092 V.
Starting at the regular pulse (RP) peak the curves closely follow the PC distributions. Both high and low charges in the classic PP SER thus have a heightened probability to originate from (multiple) EAPs.

The charge curve of classic APs follows that of classic PPs: These are mostly sum pulses from γnk photonic APs (see p. 324) and their accompanying satellite pulses.

Generally, at smaller voltages the pe-threshold is higher due to constant noise components, for which reason less undersized pulses are visible. This affects the calculated mean charge and the charge standard deviation which are used to calculate R_σ and F.

Finally, the classic PP SER shall be compared for LNGS and SPAX data of the R5912 at 1425 V.
For SPAX the mean charge is 5.3% larger (possible reasons stated above), the peak charge is 11.5% higher, the FWHM is 33.3% larger, and the standard deviation σ is 47.9% higher (partly due to a lower SPAX threshold). In a word, the LNGS SER is notably more compact, which raises the question as to the origin of this discrepancy. While the R_F resulting from the SPAX SER is in agreement with literature values (compare tables 3.1 and 3.2), the LNGS values are in tension. A higher 2-pe contamination in SPAX ($o = 2.92\%$, LNGS 1.83%) could only partially explain the differences. Altered values of the bin-to-charge scaling and the charge offset in the QDC used for the LNGS measurement, however, would explain both the observed compression of the SER and that the reference gain matches the one measured with SPAX. Likely, this was caused by an insufficient correction of the pedestal position during the charge calibration.
Since this calibration was used for all LNGS measurements, the majority of charge properties obtained from the LNGS SERs is unreliable and is therefore enclosed in parentheses in table 3.2. Since the gain was calibrated to 10^7 with a pulser, gains set to this value are *correct*, however, because the errors from offset and scaling cancel.

[30] For two or more missed gain steps, no discernible peaks are expected anymore. Rather, the slope should show an increase for each additional skipped dynode.

The deviation between true and measured gain increases the more the gain differs from 10^7; higher gains will be reconstructed with too low values. The classic mean PP gain of the R5912 at 1425 V at the LNGS was $1.29 \cdot 10^7$, which – as expected – is slightly lower than the SPAX value of $1.35 \cdot 10^7$. This also shows that the predicted gain error for the LNGS measurements (the classic mean PP gain ranged from $9.64 \cdot 10^6$ to $1.34 \cdot 10^7$) is sufficiently low to allow the use of these measurements. For the PP peak charge, resolution and ENF values, however, only relative comparisons for similar gains can be carried out. The P/V is *not* affected by an offset or charge scaling and can be used.

Properties

The mean and peak charges for the various pulse types were already discussed for their gain equivalents.

Resolution

The charge resolution is extracted from the SER and can be defined in several ways, see eqs. (2.64) to (2.67). For comparison, values for all definitions are presented here. For R_σ the standard deviation of the SER is used and not that of a Gaussian fitted to the peak, as is frequently done, since this would result in too low values.

The resolution $R_{F,m,\mathrm{PP}}$ is based on the FWHM of the spe peak and the mean charge of PPs. For classic PPs the value at $g \approx 10^7$ is comparable for the R5912, R7081 and R11780-HQE. With increasing voltage a slight decrease is found for the R5912 and a small increase for the R11780-HQE. No values from pulse-fitting could be obtained for lack of a deep enough valley.
The values for clean PPs (only available for the R5912) are larger than the corresponding classic numbers for low voltages due to the increased detection of small pulses. At high voltages the value is smaller, however. This likely is due to a growing $\delta 1$, which better separates γk from underamplified PPs.
For PCs, the resolution $R_{F,m,\mathrm{PC}}$ is similar, if a little lower, than that of classic PPs. A slight decrease with HV is observed.

The peak-based FWHM resolution $R_{F,p,\mathrm{PP}}$ of classic PPs, as expected, produces smaller values than the mean-based $R_{F,m,\mathrm{PP}}$, since the mean charge lies below the peak due to the undersized pulses. The values are comparable for R5912 and R7081, where the mean charge notably differs from the peak, and higher for the R11780-HQE, where both values are closer. The numbers decrease with voltage. Again, no values for pulse-fitting are possible.
For clean PPs (only for R5912) a strong decrease with HV is observed and the value is far lower than for classic PPs except for the lowest voltage, which can be attributed to the growing $\delta 1$.
For PCs the values are once again close to the classic PPs (slightly smaller) and drop with voltage.

For the R5912 and R11780-HQE at $g \approx 10^7$ both $R_{F,m,\mathrm{PP}}$ and $R_{F,p,\mathrm{PP}}$ are in agreement with literature values (table 3.1; R5912 71–111%, R11780-HQE 103±39%) and the more realistic $R_{F,m,\mathrm{PP}}$ is used as the final result.

The resolution $R_{\sigma,\mathrm{PP}}$ is based on the SER standard deviation and the mean charge. The benefit of this number is that it contains the influence of the underamplified pulses.

While these increase the standard deviation, the resolution is still far smaller than any FWHM-based definition. However, here an *increase* with voltage is found due to the growing contribution from pileup. For classic PPs the resolution is smaller than for fitted PPs, which contain far more small pulses, and the growth with HV is more pronounced. Values at $g \approx 10^7$ are similar for the R5912, R7081 and R11780-HQE. For clean PPs the resolution is higher than for classic PPs but lower than for fitted PPs (with the exception of the R11780-HQE at 2092 V) and rises with HV.
Using PCs instead gives slightly lower values than for classic PPs and shows the same behavior.

Excess noise factor

The excess noise factor F_{PP} for classic PPs (defined by eq. (2.52)) at the standard gain is 1.25 for the R5912 and 1.26 for the R11780-HQE. As expected, this is higher than calculated with (2.55) (1.075 for the R5912). The ENF of the R7081 should be similar. A slight increase with voltage is found. The values for fitted PPs are notably higher due to the higher fraction of underamplified pulses and also appear to increase with HV. For clean PPs the ENF is also higher but typically not as much as for fitted PPs. The numbers for PCs again resemble those of classic PPs.

Peak-to-valley ratio

The P/V of classic PPs at standard gain is highest for the R6594 (3.76), which indicates a very good n-pe separability and photon-counting capability. The R5912 (2.85), R7081 (2.79) and R11780-HQE (2.89) show similar, good values. The results of SPAX and LNGS agree. A small increase with HV is observed. The literature values for the R5912 (2.5–4.3) are consistent with the result, the value of the R7081 agrees with that of the manufacturer (2.8) but lies lower than in literature (3.2–5.5). The measured P/V of the R11780-HQE is above reported values (2.24 ± 0.27), but agrees with those of the R11780 (2.8 ± 0.28, manufacturer 2.8); the SER should be similar for SQE and HQE cathodes.
The PC values are similar and show a stronger increase with voltage. For fitted PPs the P/V is far smaller (below 2 for all measurements) and more or less independent of the voltage.

Pulse height

Pulse height distributions

The pulse height distributions (PHDs) obtained with SPAX for the R5912 and R11780-HQE are shown in fig. 3.11.
The PHDs which result from pulse-fitting are completely equivalent to the corresponding charge distributions, since in an APS-fit the charge and height are coupled by a fixed factor. Therefore all findings noted there also apply here.

Interestingly, slight oscillations are observed in all classic PP PHDs to some extent (stronger for the R5912).
One immediate conjecture would be that these sub-peaks might result from different numbers of secondary electrons (SEs) produced on d1, but this can not be the case for a number of reasons. First, the number of sub-peaks lying below the spe peak does not match the measured $\delta 1$ for the R5912. Second, such sharp distributions would only be possible for a very high $\delta 2$, which is not the case here. And third,

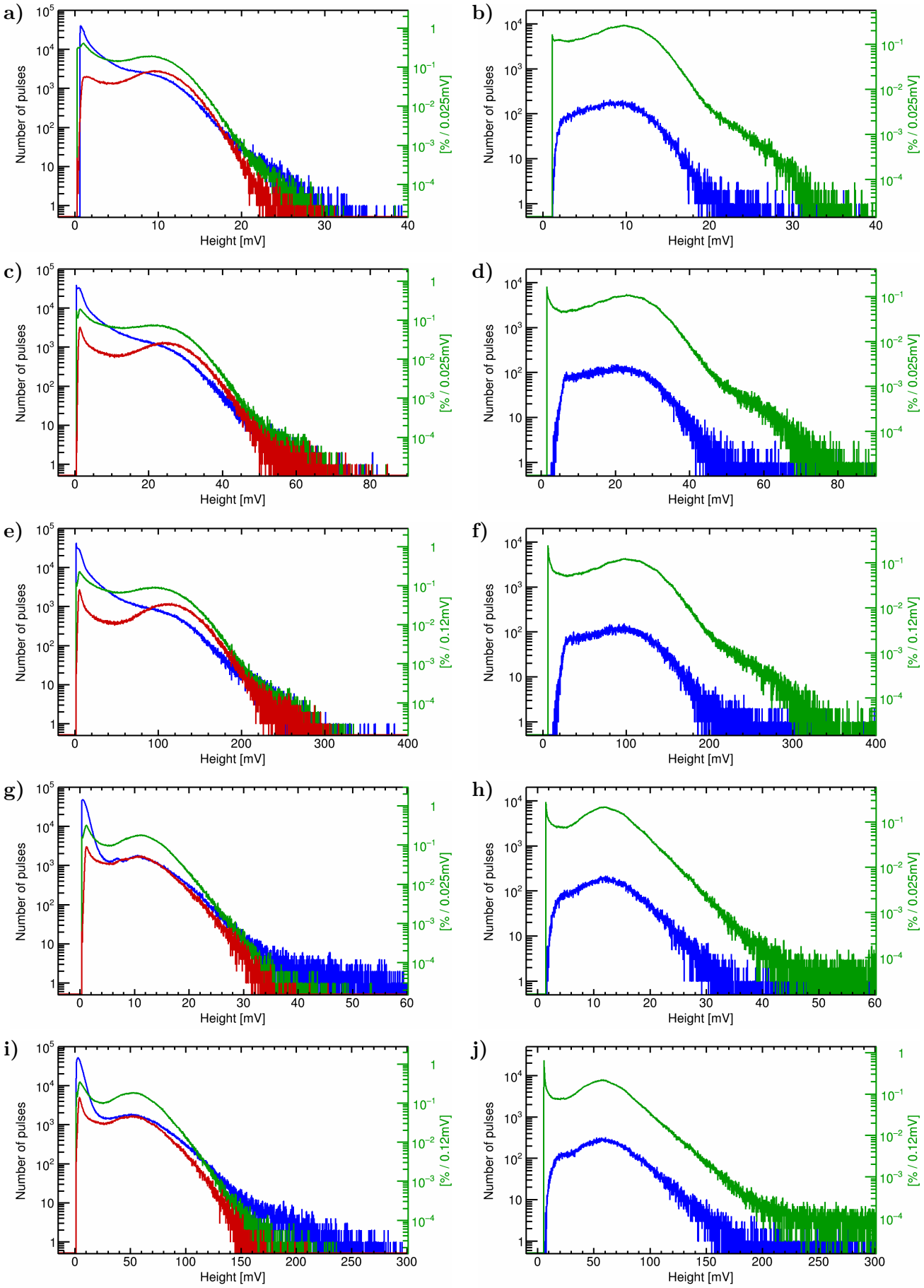


Figure 3.11: Height results from SPAX: Pulse height for clean PPs (red, only for pulse-fitting), PPs (green) and APs (blue). Fraction of all PPs per bin on right axis. Figures on left from pulse-fitting, on right from classic pulse search for same measurement. Discussion see text.
a)–f) R5912: a) and b) +1275 V, c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

none of the corresponding classic SERs and pulse-fitting PHDs show any oscillations. Therefore, these structures must have other origins.
Time-correlated noise oscillations from the light source or the trigger electronics could cause a height modulation of RPs which are shifted by the CTS and have a suboptimal $\delta 1$. However, the average empty waveform showed no oscillations for the respective measurements and was subtracted from all waveforms specifically to eliminate correlated noise.
The observed number of oscillations was independent of gain, which leaves two possible effects: Either the vertical resolution of the FADC shows oscillations or the cause is from a PMT-intrinsic effect. In case of the latter, the oscillations could stem from pileup of small EAPs. These would not be visible in the PHD of fitted PPs, since the EAPs are resolved and treated as independent (after)pulses. Such EAPs would have to have very short delays so that the peak widths still overlap; in the order of 2 ns for the R5912. In the DD (see fig. 3.19) the largest peak lies at these delays, which provides a multitude of possible effects. To confirm whether this indeed could cause height oscillations would require further, extensive studies, however.
The FADC vertical resolution seems to be a far more likely cause. Such small oscillations would not distort the presented results significantly, especially since they average out in the fits and integrations through which most properties were obtained.

Properties

The classic mean PP height h_{PP} at the standard gain is significantly higher for the R5912 than for the R11780-HQE. For both PMTs the height rises faster with HV than the classic mean PP charge, which points to pileup as the cause.
The fitted mean PP height is smaller than the classic height due to the undersized pulses and when compared to the fitted mean charge shows a similar voltage behavior.
Clean PPs have heights close to those of classic PPs for the R5912 (elimination of undersized pulses dominates), but similar to fitted PPs for the R11780-HQE, where the exclusion of pileup is more relevant.
The classic PP peak height lies notably above the mean height for the R5912 and only slightly for the R11780-HQE. The peak still rises faster with voltage than the corresponding charge but not as strongly as the mean. For fitted PPs the values are smaller.
Finally, the peak clean PP height gives the most accurate estimate of the RP height: For the R5912 this value lies above the classic mean height (less undersized EPs), for the R11780-HQE below (less pileup).

3.2.2.2 Time response

In the following discussion of the PMT timing measurements, first the relevant distributions (TTD, CTD) will be discussed using TST. Equipped with the detailed TST analysis of an R5912 at +1425 V (see 2.3.6 and 2.5.2.2, especially figs. 2.44–2.47), the results at other voltages and for other PMTs can be examined qualitatively without having to repeat a full – and complex – TST evaluation.
The transit time and TTS will be addressed together with the TTD.
Afterwards, the EPs and LPs are analyzed, before the fraction of PPs and EPs vs. threshold and the selection of the optimum threshold for EP elimination are covered.
The resulting timing properties are summarized in table 3.3.

Series	**R6594**	**R5912**				**R7081**	**R11780-HQE**	
Voltage [V]	+1670	+1275	+1425		+1725	+1520	+1648	+2092
Setup	LNGS	SPAX	SPAX	LNGS	SPAX	LNGS	SPAX	SPAX
Gain [10^7]		*3.79 · 10^6*	*8.87 · 10^6*		*3.58 · 10^7*		*7.47 · 10^6*	*3.39 · 10^7*
	$9.64 \cdot 10^6$	$6.11 \cdot 10^6$	$1.35 \cdot 10^7$	$1.29 \cdot 10^7$	$5.42 \cdot 10^7$	$1.34 \cdot 10^7$	$1.01 \cdot 10^7$	$4.60 \cdot 10^7$
Threshold		*11.6*	*6.2*		*5.4*		*13.4*	*10.9*
[% pe]	20	11.4	6.4	20	6.2	20	12.3	9.7
Fit threshold [% pe]		*2.7*	*1.4*		*1.2*		*2.8*	*2.3*
Voltage divider	6594-091009	E7694-01	E7694-01	E7694-01	E7694-01	E7694-01	E7693-01-MOD	E7693-01-MOD
Setup jitter FWHM [ns]	0.215	1.170	1.170	0.217	1.170	0.217	1.204	1.199
CF-point [%]		*55*	*55*		*55*		*54*	*55*
Occupancy [%]	5.53	3.14	2.92	1.83	3.29	2.91	2.03	2.09
Time in dark	9 h	30.8 h	4.7 h	3 h	17.3 h	11.5 h	17.0 d	18.0 d
DCR [kHz]	4.95	0.51	0.77	1.51	1.12	2.46	6.32	9.06
Temp. [°C]	23	21.8	22.4	23	23.1	23	22.0	21.8
N	$1.73 \cdot 10^7$	$5 \cdot 10^6$	$5 \cdot 10^6$	$1.52 \cdot 10^7$	$5 \cdot 10^6$	$3.19 \cdot 10^6$	$5.02 \cdot 10^6$	$5.02 \cdot 10^6$
T_{FWHM} [ns]		*2.55*	*2.27*		*2.07*		*2.53*	*2.19*
	1.93	2.52	2.39	2.04	2.14	3.06	2.45	2.12
$T_{\mathrm{FWHM},\,s}$ [ns]		*2.26*	*1.94*		*1.70*		*2.23*	*1.83*
	1.92	**2.23**	**2.08**	**2.03**	**1.80**	**3.05**	**2.13**	**1.75**
$T_{\mathrm{FWHM},\,c}$ [ns]		*2.27*	*2.11*		*1.84*		*2.47*	*2.11*
$T_{\mathrm{FWHM},\,c,\,s}$ [ns]		*1.95*	*1.75*		*1.42*		*2.16*	*1.74*
σ_T [ns]		*0.90*	*0.81*		*0.74*		*1.00*	*0.89*
	0.54	0.78	0.74	0.60	0.66	0.87	0.91	0.80
$\sigma_{T,s}$ [ns]		*0.75*	*0.64*		*0.54*		*0.86*	*0.73*
	0.53	**0.61**	**0.54**	**0.59**	**0.43**	**0.86**	**0.76**	**0.62**
σ_{TTD} [ns]		*8.37*	*8.00*		*7.43*		*11.60*	*10.13*
	4.65	8.25	7.74	7.68	6.76	9.98	11.38	2.40
T_0 [ns]		*62.6*	*59.9*		*55.4*		*80.6*	*72.7*
	[45]	**62.7**	**59.9**	[59.9]	**55.5**	[63]	**80.6**	**72.7**
$\overline{T}$ [ns]		*64.3*	*61.6*		*57.1*		*82.4*	*74.3*
	[46.1]	65.0	62.1	61.8	57.5	[65.5]	83.0	73.9
$p_{\mathrm{EP},\,3\sigma}$ [%]		*6.24*	*5.69*		*4.59*		*6.17*	*5.60*
	2.67	**0.77**	**1.12**	**1.49**	**1.06**	**0.58**	**1.63**	**1.90**
$p_{\mathrm{EP},\,\mathrm{G}}$ [%]		*12.00*	*8.37*		*6.59*		*11.93*	*11.01*
	3.00	**0.88**	**1.27**	**(1.86)**	**1.24**	**0.69**	**3.49**	**3.02**
E1: p [%]		*4.38*	*5.01*		*3.53*		*7.62*	*4.66*
	1.16	0.50	0.69	0.69	0.78	0.49	3.19	2.12
$\lvert T_{\mathrm{PP}}\rvert$ [ns]		*4.11*	*3.82*		*3.77*		*3.32*	*3.11*
	5.54	4.44	4.42	4.27	3.60	3.94	2.12	2.41
σ [ns]		*1.47*	*1.47*		*1.27*		*1.27*	*0.87*
	1.18	1.30	1.34	0.97	1.37	1.13	1.47	0.90
γ1: p [%]		*0.034*	*0.142*		*0.142*		*0.144*	*0.181*
	0.0052	0.035	0.143	0.011	0.142	0.014	0.147	0.188
$\lvert T_{\mathrm{PP}}\rvert$ [ns]		*23.8*	*23.8*		*21.5*		*35.8*	*31.7*
	13.8	24.4	23.3	22.6	21.4	31.3	33.9	31.2
h_{EP} [pe]		*0.191*	*0.166*		*0.143*		*0.296*	*0.282*
		0.531	0.415		0.374		0.653	0.509
$p_{\mathrm{LP},\,3\sigma}$ [%]		*11.7*	*13.8*		*14.1*		*6.31*	*6.00*
	14.3	**18.5**	**19.9**	**18.4**	**20.5**	**18.0**	**8.62**	**7.61**
$p_{\mathrm{LP},\,\mathrm{G}}$ [%]		*23.7*	*26.8*		*26.9*		*12.1*	*12.4*
	34.9	**37.6**	**38.1**	**(37.0)**	**40.1**	**40.0**	**21.3**	**20.7**
$p_{\mathrm{LP},\,\mathrm{p}}$ [%]		*4.10*	*4.76*		*4.84*		*3.49*	*3.44*
	3.41	**3.93**	**4.52**	**2.94**	**4.50**	**3.16**	**3.41**	**2.55**
L1: p [%]		*1.28*	*1.26*		*1.49*		*1.49*	*1.44*
	2.65	1.28	1.22	1.51	1.39	1.37	1.47	1.28
T_{PP} [ns]		*51.2*	*48.8*		*45.0*		*73.5*	*65.8*
	25.5	51.6	49.1	49.1	45.0	63.8	73.7	65.8
σ [ns]		*2.40*	*2.14*		*2.22*		*3.16*	*2.52*
	2.01	2.25	2.04	2.09	2.04	2.66	3.06	2.26

Table 3.3: PMT characterization results: Timing parameters.
See table 3.2 for measurement and formatting descriptions. Values in parentheses for unreliable fit. Discussion in text. All results after unshadowing and DCR subtraction of the TTD. TTS: T_{FWHM} and σ_T, subscript s after Gaussian subtraction of setup jitter, subscript c for clean PPs. σ_{TTD} and $\overline{T}$ standard deviation and mean of complete TTD. T_0 main peak center (γk transit time): mean of peak Gaussian of a 3-Gaussian fit of rising flank and peak; if no sufficiently low red. χ^2 was achievable, the maximum position of a cubic polynomial fit of the peak was used. For transit times in brackets (LNGS only) the time offset could not be determined and for T_0 values from the SPAX measurement or datasheets [315] for $g = 10^7$ were used. $p_{\mathrm{EP},\,3\sigma}/p_{\mathrm{LP},\,3\sigma}$: EP/LP probability up to $3\,\sigma_T$ from T_0. $p_{\mathrm{EP},\,G}/p_{\mathrm{LP},\,G}$: prob. of all EPs/LPs remaining after subtraction of the peak Gaussian of the 3-Gaussian fit. $p_{\mathrm{LP},\,\mathrm{p}}$: prob. after main peak end. p probability, T_{PP} relative transit time to T_0 (lateness if positive, $\lvert T_{\mathrm{PP}}\rvert$ earliness), σ standard deviation of fitted Gaussian. h_{EP} optimum threshold for EP discrimination and PP preservation.

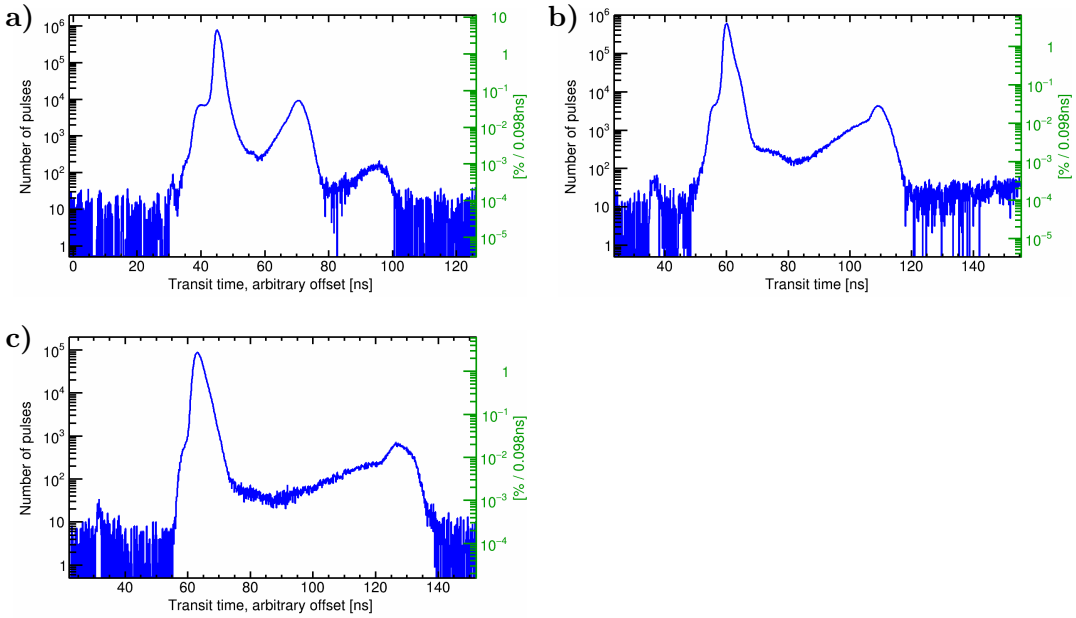


Figure 3.12: Timing results from LNGS measurements: TTD after unshadowing and DCR subtraction. Fraction of all PPs per bin on right axis. Absolute transit time for R5912 (main peak center from classic SPAX evaluation used for offset). For the R6594 and R7081 the main peak transit time T_0 for $g = 10^7$ from the datasheet was used to determine the offset. Discussion see text. **a)** R6594 +1670 V, **b)** R5912 +1425 V, **c)** R7081 +1520 V.

Transit time distributions

In figs. 3.12 and 3.13 the TTDs which were acquired at the LNGS and with SPAX for all PMTs and voltages are shown for several TTD subtypes.
First, the TTD for classic PPs shall be addressed – the baseline histogram, which is also used in literature. Comparison with the fitted and clean PPs will reveal the additional information provided by pulse-fitting and the shortcomings of such a classic TTD.

Classic primary pulses

For the LNGS measurements, the absolute offset of the TTD could not be determined. In order to provide absolute transit times, the main peak center of the R5912 measurement was adjusted to coincide with the corresponding SPAX measurement. For the R6594 and R7081 no such measurements were made, so instead the main peak center was set to the datasheet value for $g = 10^7$. Since this value is averaged over the series and was measured by the manufacturer, the time axis was marked as having an arbitrary offset.

In the classic PP TTD in fig. 3.12a, the R6594 shows a pronounced X_sE1 (E1, ΛE1, BE1) cluster. E1 is well discernible since it has a high probability, lies very early, and the TTS is very small. The γm onset is clearly visible. γ1 has a quite short earliness due to the small k–d1 distance. The main peak is rather Gaussian; the right flank shows CTS and is not as strongly affected by EAP pileup in the sum-PP as other PMTs (see below). This indicates a lower EAP rate. The bump from Rm/ΛL11 is clearly discernible, the transition region which lies before this might stem from L11 and ΛL11 or EAP pileup. The ΛL1 slope before L1 is visible,

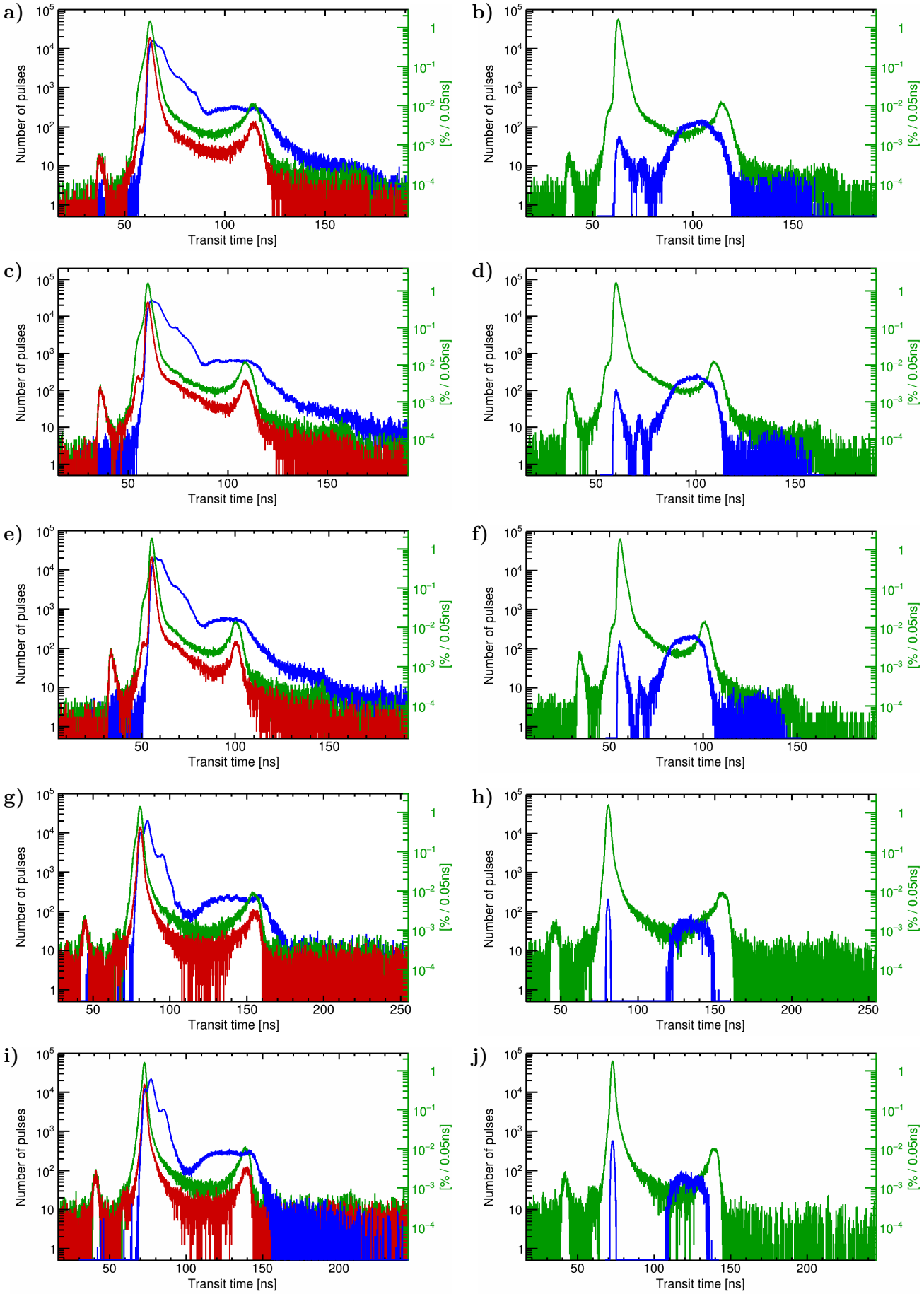


Figure 3.13: Timing results from SPAX: TTD after unshadowing and DCR subtraction. Figures on left from pulse-fitting, on right from classic pulse search for same measurement. Red clean PPs (no pileup, no APs), green PPs, blue all EAPs. Fraction of all PPs per bin on right axis. Absolute transit times plotted. Discussion see text.
a)–f) R5912: a) and b) +1275 V (minimum AP charge 0.06 pC), c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

but less marked than in other PMTs. L1 itself is very pronounced, and even the L1L1 cluster is extremely well visible.

In the LNGS TTD of the R5912 (fig. 3.12b) γ1, γ1L2, and γm are visible, as well as a marked E1 cluster at moderate earliness. σ_T is small, which emphasizes the E1 population. The right main peak flank shows CTS and is strongly affected by EAP pileup: Since the peak region in the primary pulse is very flat, EAPs with short delays can efficiently add to the height of the sum-pulse. This shifts the CF-time backwards and causes the kink in the right flank of the main peak in the TTD. This kink is less pronounced in the SPAX classic TTD (fig. 3.13d), since due to the lower threshold more underamplified pulses are registered in this time range. After the peak Rm/ΛL11 are clearly present and a high ΛL1 probability is found. The L1 peak is rather Gaussian, and the onset of the L1L1 cluster is visible near the end of the window.
In SPAX for the acquisition of the classic TTD a lower threshold was used, for which reason the measured γ1 population is far larger and the number of γ1L2 pulses is also higher. Here the end of the L1L1 cluster is visible due to the longer acquisition window. The region between the main peak and L1 is filled up by setup fluorescence (see p. 476), which extends to beyond L1. This contributes $\mathcal{O}(1\%)$ to the LP probability and has a decay time of about 18.2 ns.
For rising voltages the main peak moves to earlier times and becomes sharper due to a decrease of σ_T, and the lateness and earliness of all TSEs decreases. The γ1 cluster becomes stronger because of the higher pe-threshold, which also affects γm and the E1 cluster.

In the R7081 classic TTD, γ1 lies very early due to the large cathode diameter. The γm content is low. The E1 cluster is quite Gaussian at the left flank, has a short earliness, and strongly overlaps with the main peak due to the large TTS. On the right flank of the main peak, a large contribution from EAP pileup is visible. Rm/ΛL11 is present, the ΛL1 probability is high, and L1 is very broad.

For the R11780-HQE (SPAX only) γ1 and γm are visible, whereas E1 lies very close to the main peak and is merely discernible as a kink in the slope. σ_T is rather small for the PMT size and actually shows a lower value than the smaller R7081. This leads to a rather Gaussian peak, since also only little CTS and pileup are encountered on the right flank. ΛL11/Rm are less marked than in the R5912 and are not distinguishable from the setup fluorescence. ΛL1 is not very strong either, L1 is broad, and some indication of L1L1 is perceptible. For higher voltages the earliness and lateness of all effects decreases, as expected, and the number of γ1 and γm pulses grows a bit.

Fitted primary pulses

Compared to the TTDs for classic PPs, the distributions for fitted PPs of the R5912 show a considerable increase of the E1 cluster. This cluster becomes more discernible with rising voltage although the E1 earliness gets smaller, since the TTS decreases significantly – stronger than for classic PPs. The kink from EAP pileup in the right main peak flank disappears, as EAPs are now resolved.
For the R11780-HQE the E1 cluster also increases significantly but becomes less discernible for higher voltages – here the earliness reduction outweighs the decrease of the TTS.

Clean primary pulses

Comparing the clean primary pulse TTD with the fitted PP TTD provides a wealth of information.
The clean-PP cuts exclude all waveforms with APs or suspected pileup (see p. 488). This eliminates *all* effects based on electronic pulse-branching.
Apart from BX_d1 [31], this excludes all C-, Λ- and B-effects occurring on d2 or later stages [32], sh/anE/Ln/nn effects [33], many stage skipping effects [34], and the majority of multiple site effects. Furthermore, also all photonic APs (γij with i > j) and photonic EPs (γij with i < j) are eliminated. See 2.3.6.3 and 2.3.6.4 for a complete list or branching TSEs which also includes rare effects.
Because the clean-PP cuts exclude such a great number of TSEs, they are helpful to better understand effects by analyzing the differences between the clean-PP TTD (red) and the fitted-PP TTD (green) and comparing them with expectations. This is done in the following.

The γ1 cluster remains nearly unchanged, as do γm and γ1L2. Due to the skipped/low amplification on d1 for these effects virtually no detectable TSEs can be initiated on later stages. Therefore, no APs are produced and the clean-PP cuts remove almost no events.
The pulse selection separates E1 from BE1, sa1E1, sh1E1, X_sE2 and X_sE22. Only ΛE1 and γm are expected with notable probabilities at the same times as E1, which allows to more accurately determine the E1 probability. For the R5912, E1 becomes a well-formed Gaussian which contains a small fraction of the original events of the E1 cluster. For the R11780-HQE E1 is too close to the main peak to allow to observe its shape.
The cuts also isolate the RPs from γk from RPs from the abovementioned effects, which reduces the height of the main peak significantly. Due to the large delays (no pileup) of γn1 and γnk even more events than the pileup probability – which already lies around 65% (see table 3.6) – are removed.
The falling main peak flank shows a slight bump around 4 ns lateness in the fitted PP TTD which disappears in the clean TTD. This indicates that at least the late fraction of CTS, L11 or ΛL11 events has an increased likelihood to create EAPs. This might be due to the altered hit position on d1.
The cuts cannot remove the setup fluorescence, since this consists mostly of delayed γk pulses. Thus, the fluorescence is reduced by same factor as the main peak, because the same secondary effects can occur.
The L1-cluster should also shrink by the same factor as the main peak, which is in agreement with observations.
For ΛL1 a difference is observed which increases with rising lateness, since later pulses have a higher impact energy and thus can create more APs through branching or the creation of photons.
At very late transit times the growing relative fluctuations due to lower bin contents appear to reduce the vertical distance.
All observations match predictions from TST.

[31] BE1, BE11, BL11, BL1
[32] For instance, for the second dynode: E2, ΛE2, BE2, E22, ΛE22, BE22, L22, ΛL22, BL22, L2, ΛL2, BL2
[33] shnEn, sanEn, shnEnn, sanEnn, shnLnn, sanLnn, shnLn, sanLn
[34] E24, L23, s(h/a)nEn(n+2), s(h/a)nLn(n+1)

Transit time distributions for afterpulses

In the TTD of fitted EAPs, the large population of EAPs at the main peak center is remarkable. This consists of the RPs of branching effects which produce EPs (e.g. BE1).
Subsequent EAP populations roughly correspond with the DD (fig. 3.19), since RPs are the most frequent PPs and are rather localized in the TTD. These populations grow with rising voltage, since more particle are present in the cascade and the backscattering coefficient rises.
Since DCR events must occur in the PP-TTD before APs can be detected, the DCR is reduced during the first nanoseconds in the EAP-TTD and then rises to its usual plateau.

In the TTD for classic EAPs high minimum charge and charge fraction cuts were required to eliminate APS oscillations. Therefore almost only γnk (regular charge) and secondary γnk are observable.
Fascinatingly, some of the AP-RP are actually visible around the main peak center although no pulse subtraction or fitting was used. This is only possible if the preceding PP is a very small EP with sufficient earliness and can be understood as follows: While small EPs have the same pulse length as a γk (judging from the APS fitting results), they reenter quickly into the noise band after passing the threshold and only exceed it during a short time around their peak. Therefore, the pulse end is registered prematurely (shortly after the peak), which allows to find subsequent RPs at the main peak time.
Both the γnk and AP-RP populations grow with rising HV.
For the 5912 the second AP peak around 65–75 ns originates from the second, largest APS maximum and is artificial. Since the pulses are not subtracted in the classic analysis, a cut on the minimum AP charge is required to eliminate fake APs resulting from APS oscillations. Due to its size the second APS peak, however, is not eliminable without loosing a large fraction of the true EAPs.

Properties

The transit time of the main peak center T_0, often simply called "transit time", is the time between photon incidence and pulse output of the RPs, meaning γk. Its values are virtually identical for classic PPs, fitted PPs and clean PPs; for the latter (not listed) it is only 0.023–0.093 ns larger than the fitted PP values. As expected, T_0 falls with rising voltage, since the higher fields accelerate the electrons to larger velocities, and rises with increasing cathode diameter (cf. fig. 2.35), mostly due to the growing k–d1 distance x_{k1}.
On the manufacturer datasheet of the R5912 a T_0 value of 55 ns is specified for $g = 10^7$, and values of 64.8 ns at 1275 V, 58.5 ns at 1425 V and 50.2 ns at 1725 V can be extracted from a graph [470]. The value at 1425 V is close to the measured one; however, the observed reduction of T_0 with growing voltage is far slower. For the R11780 a value of 70 ns at $g = 10^7$ is listed [439], but the measured value is significantly higher. This shows the variation of T_0 between samples of the same series.

The mean transit time $\overline{T}$ of the TTD is a related quantity and differs from T_0 due to the presence of EPs and LPs. For all PMTs $\overline{T}$ lies above T_0, because the LPs have larger probabilities and temporal distances from T_0 than the EPs. For fitted

PPs the value is lower than for classic PPs, since more EPs are found, but still lies above T_0. The difference between $\overline{T}$ and T_0 roughly rises with larger diameters and falls with increasing voltage. Classic values for SPAX and LNGS differ slightly due to the different threshold.

The transit time spread values T_{FWHM} and σ_T of the unshadowed, DCR-subtracted TTD still contain the smearing from the light source ETD and the other contributions of the setup jitter. Since no reliable unfolding routine could be implemented, instead the dominant sources of the setup jitter and the main peak were approximated with Gaussians (see p. 480). In this case the total TTS is a quadratic sum of the standard deviation or FWHM of its components (see eq. (2.74)) and the actual TTS can be obtained by inverting the equation. The values which were obtained using the Gaussian subtraction are denoted by a subscript s. For the LNGS the corrected values are almost unchanged due to the low jitter of the laser, while notable differences occur for SPAX due to the use of an LED. The resulting values for the R5912 at 1425 V from LNGS and SPAX differ by only 47 ps, where the remaining small difference could occur due to the different thresholds. This shows that the used procedure is reasonably accurate within the used approximations (all effects treated as statistic variations). Unfolding would be preferable in the future, however.
The classic $T_{\mathrm{FWHM},\,s}$ is smallest for the R6594, closely followed by the R5912 and then the R11780-HQE, while it lies above 3 ns for the R7081. The expected strong decrease with HV is observed. For fitted PPs the values differ somewhat due to the omission of pileup and the addition of EPs. If clean PPs are used (subscript c) the values drop notably, which may be helpful to better study the effect of the CTS. The manufacturer value of 1.5 ns for the R6594 (see table 3.1) is better than measured (1.92 ns). The literature values of the R5912 ($\approx 2.9 \pm 0.3$ ns, manufacturer 2.4 ns, $g = 10^7$) on the other hand exceeds the measured values (2.03 ns for $g = 1.35 \cdot 10^7$) by far. For the R7081 literature data (2.8 ± 0.2 ns, $g = 10^7$) matches the measurement (3.05 ns, $g = 1.35 \cdot 10^7$), while the manufacturer value of 3.4 ns is higher. The R11780-HQE performs better (2.13 ns) than in literature (3.04 ± 0.33 ns). This great variance once again shows the large fluctuations between individual PMTs and the importance of measuring a sufficiently large number of samples of a series to be able to properly establish its properties.
Using $\sigma_{T,s}$ – the standard deviation of a fitted Gauss (3-Gauss fit of the left flank of the main peak in the TTD) – produces lower values than the FWHM. Even after a multiplication by the σ–FWHM conversion factor of 2.35 (eq. (2.68)) the values are significantly smaller, however. This is due to the markedly non-Gaussian right flank of the main peak, which most likely results from CTS and raises the FWHM. $T_{\mathrm{FWHM},s}$ thus is a more realistic estimate of the timing uncertainty. $\sigma_{T,s}$ is smallest for the R6595 and R5912, followed by the R11780-HQE, and is largest for the R7081. It shows the same voltage behavior as $T_{\mathrm{FWHM},\,s}$. The difference between SPAX and LNGS amounts to 52 ps, which might be due to the 3-Gauss fit model not being complex enough.

The standard deviation σ_{TTD} of the *complete* TTD gives a better estimate of the total timing variation, since not only the main peak center but also EPs and LPs are taken into account. This of course depends on the threshold and may be higher than the effective variation in an experiment. The values basically show the same dependence on PMT and HV as $T_{\mathrm{FWHM},\,s}$. The extremely low classic value of σ_{TTD}

for the R11780-HQE at 2092 V is due to a low EP and LP probability, which cause $\overline{T}$ to lie close to T_0.

The newer versions of the R5912 and R7081 with improved, fully-tapered VDs have far shorter TTS and rise time values, slightly shorter T_0 and far better linearity but higher dark current than the ones used in the measurements with the old VD (compare [213] and [315]). Other properties are unchanged. For the R6594 and R11780-HQE the newest models were measured here.

For the R5912 at +1425V the fraction of γk events in the TTD was 79.3% (fitted PPs), 84.9% (SPAX classic PPs), and 87.2% (LNGS classic PPs). It is larger for classic PPs, since less EPs and LPs are found, and even higher for LNGS data due to the larger threshold, which further reduces the EP and LP probabilities.

Charge–time distribution

The charge–time distributions (CTDs) which were recorded with SPAX are shown in fig. 3.14 for both a pulse-fitting and a classic analysis.

Fitted primary pulses

The CTD of the R5912 for fitted PPs was already discussed at length in 2.3.6, so only a few additional comments shall be made.
It appears that for the γk events the bins with maximum content lie slightly above the charge of the clean PP peak in the SER, as does the L1 maximum. This is likely due to CTS, since then γk with suboptimal δ1 also contribute to the SER and shift the peak position downward. Because pileup is eliminated in the clean PP distributions, no known effect could balance this reduction.
With rising voltage the reduction of the E1 charge center due to a growing δ1 is clearly visible – 1 pe was placed at exactly half height of the plots, so one can directly compare the relative vertical extent of the E1 cluster. The same applies to all other effects which are linked to δ1 (e.g. γ1, γ1L2, γm, Rm), and the corresponding secondary effects which are based on L1. Similarly, the E2 cluster is increasingly suppressed with rising HV, because $\delta 1 \cdot \delta 2$ grows. Furthermore, γk and L1 get far more narrow in time with increasing voltage due to the TTS reduction.

In the fitted-PP CTD of the R11780-HQE the far higher DCR of this PMT can clearly be seen, even after 17 d in dark. E1 is quite close to the main peak and merges with other effects, which becomes even more severe at large voltages. E2 apparently occurs earlier than E1 and as for the R5912 the cluster is suppressed with growing voltages. There is a small cluster between γ1 and E1 at γ1 charges, which could be from γ1L2. If this is the case, the earliness would be smaller than for the R5912, possibly due to a larger d1–d2 distance x_{12}. Rm is visible as well. Altogether, even without a dedicated TST analysis, many of the effects which were discussed for the R5912 in 2.3.6 can be identified for the R11780-HQE as well.

Classic primary pulses

When comparing these histograms to their classic-PP CTD counterparts, the loss of events in the classic plots in the E1 and E2 clusters (also in the secondary clusters before L1) and for Rm becomes apparent. Instead, for a fraction of the events EPs with higher charges are created through the pileup of EPs and RP-APs.

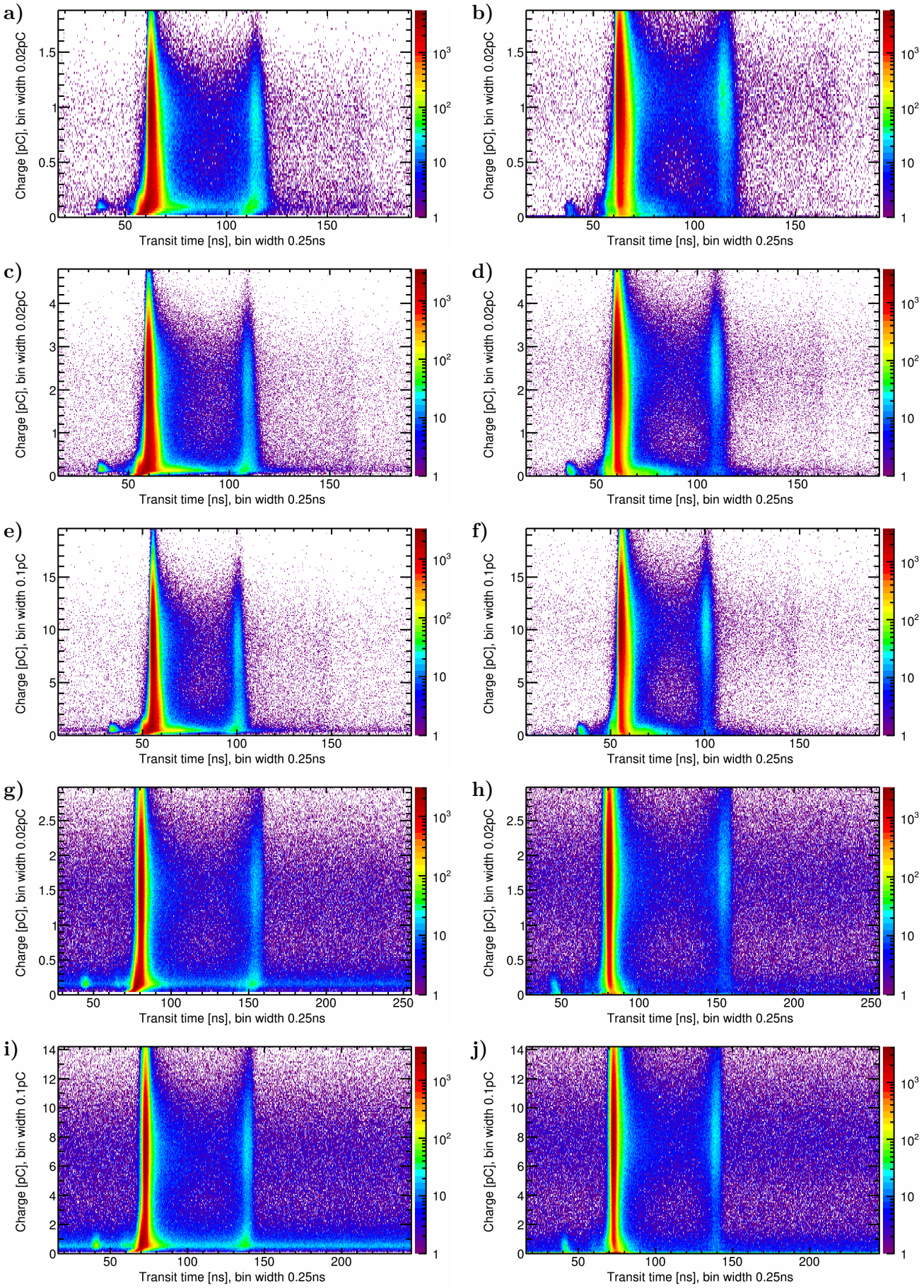


Figure 3.14: Timing results from SPAX: Charge–time distribution (CTD) for PPs. Figures on left from pulse-fitting, on right from classic pulse search for same measurement. Absolute transit times, charge axis from 0 to 2 pe (1 pe equals the peak clean PP charge), the z-axis denotes the number of pulses per bin. Discussion see text.
a)–f) R5912: a) and b) +1275 V, c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

The RP charges get larger due to pileup (note the upward shift of the dark red parts), which also shifts the RPs to later times: The tail towards larger times at regular charges becomes far longer. This effect is also visible at very low charges, where the distribution veers to later times.
The events which lie in the horizontal line at $q = 1/\delta 1$ in the fitted-PP distribution are mostly lowered to charges close to zero in the classic CTD. This effect is best visible for the R11780-HQE due to its high DCR. The reason for this is that the fit restores the charge for all detected pulses using the APS shape. Otherwise, only a few datapoints lie above the noise band, which results in artificially low charges. For example, many $\gamma 1$ events are shifted to too low charges in the classic distributions.

Early pulses

Probability

The total EP probability was determined in two ways: $p_{\mathrm{EP},\,3\sigma}$ is the integral of all pulses with an earliness of up to $3\,\sigma_T$ before the main peak. $p_{\mathrm{EP},\,\mathrm{G}}$, on the other hand, integrates the complete region before the peak after subtraction of a Gaussian fitted to the center of the main peak (3-Gauss fit of the left flank and center of the peak).
In addition, notable effects – E1 and $\gamma 1$ – were studied separately. E1 was fitted both as part of the 3-Gauss fit and individually with a single Gaussian, for $\gamma 1$ the probability and mean time were determined.

The resulting values for $p_{\mathrm{EP},\,\mathrm{G}}$ are always higher than $p_{\mathrm{EP},\,3\sigma}$ (see table 3.3) since also EPs with short earliness values are considered. $p_{\mathrm{EP},\,\mathrm{G}}$ is physically more correct (although CTS pulses are counted as EPs), while $p_{\mathrm{EP},\,3\sigma}$ is more application-focused.

For the classic TTDs, $p_{\mathrm{EP},\,\mathrm{G}}$ is very high for the R11780, shortly followed by the R6594. For the latter the high EP probability is clearly visible in the TTD, for the former due to the small earliness of E1 the large amount of EPs can best be seen in the CTD. While for the R11780 $p_{\mathrm{EP},\,3\sigma}$ is far lower than $p_{\mathrm{EP},\,\mathrm{G}}$ due to the proximity of E1 to the main peak, for the R6594 the value is almost not reduced because of the large earliness. The R5912 shows a moderate, the R7081 a low $p_{\mathrm{EP},\,\mathrm{G}}$. In both cases, $p_{\mathrm{EP},\,3\sigma}$ is not much lower due to a rather large E1 earliness.
Regarding voltage behavior, for the R5912 at 1275 V a comparatively high threshold was used, for which reason the probability values are lower and not comparable. From $g = 10^7$ to $5 \cdot 10^7$ a slight decrease for *both* $p_{\mathrm{EP},\,3\sigma}$ and $p_{\mathrm{EP},\,\mathrm{G}}$ is found, which therefore can not be explained by a decreasing earliness causing EPs to slip below the 3σ threshold. A reduction of $p_{\mathrm{EP},\,\mathrm{G}}$ with growing HV is also observed for the R11780, while here $p_{\mathrm{EP},\,3\sigma}$ rises, because the earliness drops slower than the TTS. Apparently the EP probability *decreases* with rising HV, which is initially surprising but very interesting, as this provides a means to suppress EPs. The reasons for this behavior will be derived on the following page.
For the LNGS measurement a larger $p_{\mathrm{EP},\,3\sigma}$ is encountered despite a higher threshold[35], since the size of the integration window was determined using the fitted σ_T

[35] The 3-Gauss fit was unreliable, since the high statistics of this acquisition revealed that the phenomenological fit model was not complex enough to adequately describe the TTD curve. In view of the about 69 TSEs which are located in this region (see 2.3.6) this is hardly surprising; for the other fits, the larger bin errors gracefully veil the discrepancies.

instead of $\sigma_{T,s}$ to be on the safe side. At the LNGS the setup jitter was smaller and thus the total measured σ_T is lower, so the EP integration range extends closer to the peak. The LNGS value is close to the true value, and the $3\,\sigma$-based probabilities for EPs and LPs are thus higher than shown here for SPAX data. To correct this, instead $\sigma_{T,s}$ might be used or, even better, the value which results after unfolding.
The literature values for the R5912 (0.1–0.5%) are significantly lower than measured here, which raises the question regarding the used EP definition. For the R11780, available references only report a probability of $> 0.5\%$.

For fitted PPs all probabilities are far higher than for classic PPs, since far more EPs are resolved as individual pulses instead of forming pileup. The values are slightly higher for the R11780 than the R5912.
The probability falls with rising HV for both $p_{\mathrm{EP,\,G}}$ and $p_{\mathrm{EP},\,3\sigma}$. This shall be analyzed qualitatively with TST in the following, based on the properties which were calculated for the R5912 at 1425 V.
When comparing the measured total EP probability $p_{\mathrm{EP,\,G}} = 8.37\%$ with the total probability of 57.4% predicted by TST, it becomes clear that only a fraction of EP effects is detected. This can be explained by the small earliness of many effects (compare fig. 2.44).
For the 10 most frequent EP effects the following changes are expected with rising voltage. The combined detectable probability p_c of an effect which was calculated with TST is listed in parentheses after the effect name.

- **sh1E1 (10.3%), sa1E1 (9.0%):** Since $\delta 1$ is at the peak of the SEY, the number of creation trials (equal to $\delta 1$) is constant or decreases slightly with growing voltage. The emission energy should be mostly unchanged and due to the rising U_{12} the probability $\mathrm{u2_{sh/a1}}$ of no scattering at d2 drops. The direction fraction $\mathrm{E1_{sh/a1}}$ should rise slightly but is already at 86%, which limits a possible increase. All in all, the probability should remain about constant.

- **BE1 (7.1%), E1 (2.0%):** The backscattered fraction b1 increases slightly. For BE1 the branching fraction B1 could increase but is already at 80%. The direction fraction (E1 / EB1) will decrease, since U_{12} grows slower than the backscattered energy which is gained from U_{k1}. The unscattered fraction u2 decreases. In sum, the detectable probability should thus decrease.

- **sh2E2 (5.2%), sa2E2 (5.1%):** The number of creation trials ($\delta 1 \cdot \delta 2$) rises slightly, since $\delta 2$ is not at the SEY peak yet. The other probability factors will show a similar behavior as the d1 effects. Thus, a slight increase of the mean number of occurrences is expected. However, because $\delta 1 \cdot \delta 2$ increases, the relative charge drops. Since the individual charge is below detectable thresholds, the effect is only visible if it occurs several times and forms pileup. Therefore, the detectable probability is very sensitive to the threshold and likely *drops*. In addition, it is questionable whether these effects can be resolved at all, as their earliness is very small.

- **BE2 (4.7%), E2 (2.0%):** Except for a strong increase of the backscattered fraction b2, a behavior like BE1 / E1 is expected, so the occurrence probability could actually increase. The detectable probability, however, likely decreases due to the shrinking relative charge.

- **s2E22 (3.4%):** Because $\delta 2$ grows, more s2 are created; the rest of the probability factors is unchanged, since the emission energies likely are similar. This leads to a higher occurrence, but a smaller detectable probability due to the charge suppression by a higher $\delta 1 \cdot \delta 2$.
- **BE33 (3.1%):** Apart from a strong increase of the backscattering fraction b3 and a higher number of creation trials due to a larger $\delta 2$, the behavior is like BE1. But BE33 has a strongly charge suppression ($\delta 1 \cdot \delta 2 \cdot \delta 3$) and is only visible due to pileup. Thus, the observation probability has a strong threshold dependence, which likely discriminates the effect with rising voltage.

These examples already give a picture of the general tendency, which is similar for other effects: The detectable probability is expected to drop for most effects and stay constant for the remainder. This is in agreement with observations.

E1

E1 accounts for a large fraction of the total EP probability. The order of PMT series regarding probability is the same as for $p_{\mathrm{EP,\,G}}$. The values are far higher for fitted PPs than for classic PPs. However, the fit in the TTD also includes other effects in the value, so the classic value is closer to the true probability, since less effects remain at such large earliness values. The predicted drop with HV is found for fitted PPs and for the R11780-HQE also for classic PPs, while the R5912 classic PPs show a slight increase. This could, however, also result from effect bundling – the 3-Gauss fit is only a phenomenological description. SPAX and LNGS values are in perfect agreement.
The E1 earliness decreases with growing PMT diameters and voltages. In the R11780-HQE an increase with rising HV is found for classic values, but not for fitted PPs, which is likely caused by effect-clustering.
The standard deviation of E1 follows the trend of σ_T but has notably higher values. This might be due to different d1 hit positions, which affects the traversed distance to d2. σ drops with rising voltage for fitted PPs and classic R11780-HQE values, but stays approximately constant for the R5912 for classic PPs.

$\gamma 1$

The $\gamma 1$ probabilities reflect the used pe-threshold and are remarkably identical for classic and fitted PPs.
The $\gamma 1$ earliness scales with the PMT diameter, decreases with voltage, and is mostly comparable for fitted and classic PPs as well as between SPAX and LNGS.

Late pulses

Probabilities

As for the EPs, the total LP probability is defined for $p_{\mathrm{LP},\,3\sigma}$ by integrating down to $3\,\sigma_T$ after the main peak and for $p_{\mathrm{LP,\,G}}$ by integrating over the complete LP region after subtraction of the Gaussian-fitted main peak (3-Gaussian fit of left flank).
In addition, another definition was used: The probability $p_{\mathrm{LP,\,p}}$ was determined by integrating the LP region after the end of the main peak. This was defined as the intersection of the linear extrapolations of the right flank and the subsequent LP plateau between the main peak and L1 in a semilogarithmic plot (number of events logarithmic).

For SPAX the contribution from setup fluorescence of ca. 1% has to be subtracted from the integrated values listed in the table.
Furthermore, L1 was fitted separately with a Gaussian. Here the left side of the L1 cluster was used as fit range , since the right flank shows secondary LP effects and CTS.

The resulting LP probabilities are very large for $p_{\mathrm{LP,G}}$ (21–40%), already notably smaller for $p_{\mathrm{LP},3\sigma}$ (8–21%), and far lower than both if only the range after the main peak is integrated ($p_{\mathrm{LP,p}}$ 2.6–4.5%). L1 itself has merely a probability of 1.2–2.7% and thus can only account for a fraction of $p_{\mathrm{LP,p}}$, let alone the other definitions. For $p_{\mathrm{LP,G}}$ and $p_{\mathrm{LP},3\sigma}$ by far the most LPs stem from the late flank of the main peak, mostly from CTS and γMgk. When subtracting the RP-Gaussian of the 3-Gaussian fit from the TTD of the R5912 at 1425 V (result not shown), the peak of the remaining LPs in the main peak is about 2.7 times as high as the contribution which is expected from γMgk. Since the latter already has a large observable probability p_c of 9.7%, additional effects with very high probabilities have to be involved. No effect except the CTS is expected with similar or higher probabilities than γMgk at these times and charges – even when summed, the probabilities of γMmk, L11, ΛL11, E1L2R1 and L1EfR1 are far too low. Therefore, the largest contribution to LPs is in fact expected from the CTS. At later times in the right flank, mainly higher order effects based on γMgk, L11 and ΛL11 contribute.
The probabilities are larger for classic than for fitted PPs due to the forming of pileup, which shifts RPs to LP times.
Comparing SPAX to LNGS, for $p_{\mathrm{LP,p}}$ a different value results for the LNGS even after subtracting 1% setup fluorescence. This is due to a later cut-off point (end time of peak) because of the missing fluorescence. The $p_{\mathrm{LP},3\sigma}$ value is slightly smaller for the LNGS than for SPAX (after subtracting 1%) despite the smaller σ_T. This is likely due to the higher threshold, since many underamplified LPs follow the main peak in the classic CTD (fig. 3.14d).
The LP probability $p_{\mathrm{LP,G}}$ is highest both for classic and fitted PPs for the R7081, followed by the R5912 and R6594, while it is far lower for the R11780. $p_{\mathrm{LP},3\sigma}$ is largest for the R5912, slightly lower for the R7081 and R6594, and again far lower for the R11780. The change in the order results from the different σ_T and the varying threshold. $p_{\mathrm{LP,p}}$ (after fluorescence subtraction) is largest for the R5912 and R6594, closely followed by the R7081, and lowest for the R11780. However, it is difficult to directly compare $p_{\mathrm{LP,p}}$ values, since this quantity strongly depends on the selection of the peak end position, the threshold varies between acquisitions, and setup fluorescence plays a stronger role than for the other definitions for SPAX data. To improve the comparability, SPAX data could be reevaluated with an artificially heightened threshold of 0.2 pe for all measurements.
Regarding the behavior with rising voltage, a qualitative TST treatment of the most frequent effects predicts falling probabilities for the majority of interactions (p_c listed in parentheses after the effect name):

- **CTS ($\approx$16%):** For rising voltages a reduction is expected due to shorter flight times (and thus smaller variation of tk1) and smaller spot sizes on d1 (lower variation of t_{12}).

- **γMgk (9.7%):** The transit time of this effect is completely unaffected by voltage, because only photon flight times are involved. The same applies for

secondary effects at later times. However, σ_T shrinks, so the fraction which is included in $p_{\mathrm{LP},3\sigma}$ grows for larger voltages.

- **L1 (1.3%):** This will have a slightly larger backscattering fraction b1, the direction fraction L1 likely is more or less unchanged, and the unscattered fraction $\mathrm{u1}_{\varepsilon 1}$ is slightly smaller. In sum, the probability should remain mostly constant.

- **ΛL1 (1.0%):** The Λ-fraction Λ1 could decrease, the direction fraction LΛ1 might rise slightly, the unscattered fraction $\mathrm{u1}_{\Lambda 1}$ decreases slightly, and otherwise the behavior is like L1. This leads to a falling probability, which can be confirmed by comparing the TTDs for fitted PPs (R5912: figs. 3.13c and 3.13e, R11780: figs. 3.13g and 3.13i).

- **L11 (0.8%):** The behavior is as for L1 except for a smaller direction fraction L11, which results in smaller probabilities.

- **ΛL11 (0.6%):** This behaves like L11 with an additionally lowered Λ fraction Λ1, which should lead to notably smaller probabilities.

- **Rm ($\approx$0.4% [36]):** The relative collection efficiency $\mathrm{CE_{km}}/\mathrm{CE_{k1}}$ falls, the unscattered fraction um shrinks slightly, and the creation probability cm rises slightly. The behavior of the CE on d1 ($\mathrm{CE_{m1}}$) is difficult to predict, since it depends on the grid voltage but is assumed to remain similar. In sum, the observation probability likely falls. A comparison of the TTD for fitted PPs of the R5912 for 1425 V and 1725 V appears to confirm this.

- **BLf ($\leq$0.4% [37]):** The backscattering fraction bf becomes slightly larger, the branching fraction Bf a bit higher, and the direction fraction LBf increases. Therefore in this one case *larger* probabilities are expected for higher voltages. However, the p_c of BLf is low compared to the other effects.

These predictions can now be compared with data. For $p_{\mathrm{LP,p}}$ a slight to marked decrease is found, except for the R5912 for fitted PPs [38]. This outlier is likely due to variations of the peak end position, for which the fluorescence plays a large role. For the other probability definitions, as expected the values fall with growing voltage for the R11780, but rise for the R5912. Judging from a closer analysis, this is likely due to the 3-Gauss fit producing a slightly too low σ_T for 1725 V, which affects both $p_{\mathrm{LP,G}}$ and $p_{\mathrm{LP},3\sigma}$. For $p_{\mathrm{LP,G}}$ then too few RP events are subtracted, which raises the number of LPs, while for $p_{\mathrm{LP},3\sigma}$ more CTS and γMgk events emerge from the exclusion zone. For the R11780 this effect might be outweighed by a faster decrease of the CTS than σ_T, which leads to a shrinking value.
Literature values of the LP probability quote 3% for the R7081 and $4.3 \pm 0.35\%$ for the R11780-HQE (manufacturer value for the R11780 4%). These numbers are best comparable with $p_{\mathrm{LP,p}}$ values and indicate that similar definitions were used.

[36] Predicted with 2.8%, which was found to be too large by a factor of up to 7.
[37] Predicted with 1.3%, which was found to be too large by at least a factor of 3.
[38] The R5912 for 1275 V is excluded in this comparison, because it had a far higher pe-threshold than the other measurements.

L1

The probability for L1 in general is similar for fitted and classic PPs. It is smallest for the R5912, larger for the R7081 and R11780, and very large for the R6594. For the last-mentioned PMT the E1 and L1 probability both are particularly high (see table 3.3), which can be attributed to a large pk backscatter coefficient on d1 due to the especially high U_{k1} of 719 V (cf. fig. 2.14).
With rising voltage the L1 probability decreases for the R11780 and increases for the R5912. From TST about constant values are expected. The origin of these differences are likely fit inaccuracies. Only a small reliable region is available for the fit, which leads to variations of the included fraction of the complete L1 cluster. The value for SPAX is smaller than for the LNGS, which also likely is due to fit imprecisions.
The L1 lateness is virtually identical for classic and fitted PPs as well as in SPAX and LNGS. The lateness increases with cathode diameter (and thus x_{k1}), and decreases with growing voltage, because $t1k_{\varepsilon 1}$ ($\approx t$k1) falls.
The standard deviation σ in time of L1 is very broad, far larger than for the RPs. The reason for this is that the emission energy variation, the lateral momentum and the CTS are involved *three* times for L1: on the first transit of a pk to d1, on the return path to the cathode, and on the way back to d1 (see L1 section on p. 316). σ is smallest for the R6594 and R5912, higher for the R7081, and highest for the R11780. It is found to scale with σ_T and σ/σ_T lies around 3, as predicted. With rising voltage no change in σ is found for the R5912, and a decrease os observed for the R11780 – again likely due to fit inaccuracies. For both PMTs the FWHM of the L1 peak shrinks significantly with rising voltages, since the right flank rapidly disappears due to the reduction of the triple CTS. The σ values are larger for fitted PPs than for classic PPs, which probably is due to additional L1-EPs on the left flank, which influence the few bins available for fitting. For LNGS and SPAX the values are similar.
The L1 mean charge for clean PPs fluctuates around 1.0 pe within expected errors. For fitted PPs it lies below 1 pe because of L1 pulses which are part of branching effects and have lower charges – these are excluded for clean PPs.

Fraction of primary and early pulses versus threshold

In fig. 3.15 the fraction η_{PP} of PPs which remain after increasing the threshold from zero to a given value is plotted. In the same figure also η_{EP}, the fraction of remaining EPs at this threshold, is shown.
η_{PP} and η_{EP} were obtained by first projecting the CTD onto the normalized charge axis. Following this, the charge range of the resulting histograms was integrated, where only values above the varying charge threshold were counted. To determine η_{EP}, the EP probability definition $p_{EP,3\sigma}$ was used [39]. Since every measurement has a finite threshold and no smaller pulses can be found, the values of η_{PP} and η_{EP} can remain at 100% for thresholds >0 if the detection threshold is large or the x spacing between plotted values is small.

These curves can be used to derive a recommendation for the threshold which excludes as many EPs as possible while preserving as many PPs as feasible.

[39] Using $p_{EP,G}$ would require to fit and subtract a 2D distribution (Gaussian in time axis, scaled Poisson in charge axis) in the CTD, which would introduce errors.

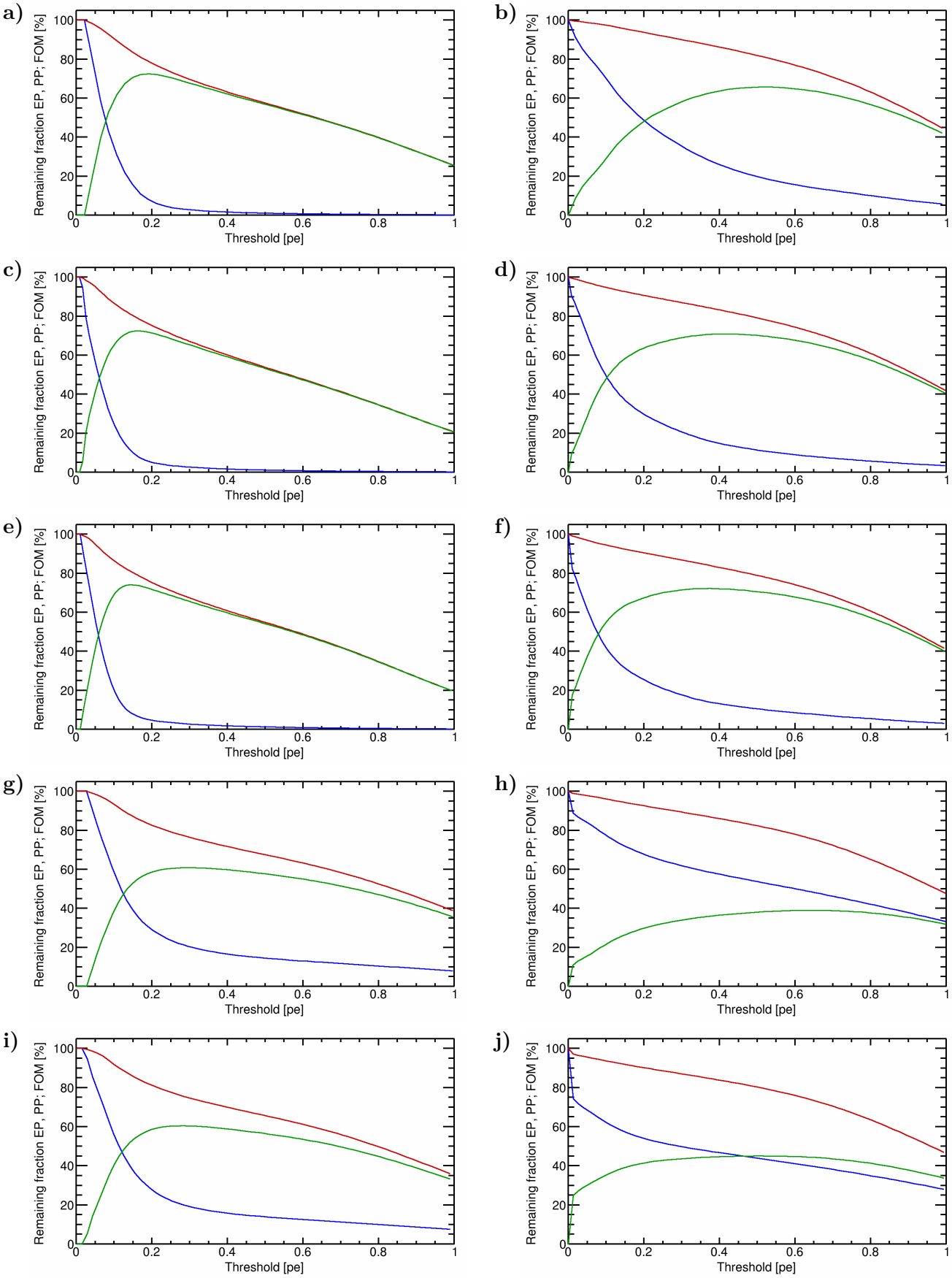


Figure 3.15: Timing results from SPAX: Remaining fraction of PPs and EPs vs. threshold in pe (using γk charge, estimated with the clean PP charge peak $q_{p,\mathrm{PPc}}$). Figures on left from pulse-fitting, on right from classic pulse search for same measurement. Red remaining PP fraction η_{PP}, blue remaining EP fraction η_{EP}, green figure of merit (FOM) $f = \eta_{\mathrm{PP}} \cdot (1 - \eta_{\mathrm{EP}})$. The recommended threshold h_{EP} for EP discrimination is defined as the threshold where f is maximal. Discussion see text.
a)–f) R5912: a) and b) +1275 V, c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

Since both curves are monotonically decreasing, it is difficult to select such a threshold by eye and a figure of merit (FOM, f) which is based on both criteria has to be introduced. This FOM should maximize the PP fraction and minimize the EP fraction and indicate graphically where the best compromise between both is reached. While the ratio $f = \eta_{\mathrm{PP}}/\eta_{\mathrm{EP}}$ is an obvious choice, its values increase monotonically with threshold and reach no maximum within the studied threshold range from 0 to 1 pe [40]. The only conclusion would be to set the threshold as high as bearable, which is not very helpful.
Therefore, instead the FOM definition

$$f = \eta_{\mathrm{PP}} \cdot (1 - \eta_{\mathrm{EP}}) \tag{3.3}$$

was used. This results in a peak near 0.2 pe for fitted PPs (green curves in fig. 3.15), which corresponds to the end of the strong drop of η_{EP} at lower thresholds and reflects what one would choose based on a visual inspection of the graphs. With this definition the recommended threshold h_{EP} for optimum EP discrimination can be defined as the threshold where f peaks.

For fitted pulses, the PP fraction shows an initial drop, since undersized pulses are lost. The drop levels out around 0.2–0.3 pe when most underamplified pulses are discriminated, and the shape of the curve is then determined by the slowly rising exclusion of the regularly amplified spe distribution. Near 1 pe the decline becomes stronger again, since the spe peak is being cut away. For the R11780-HQE the overall decrease is slower than for the R5912, since in this PMT the fraction of undersized pulses is lower. No strong variation with HV is found.
Virtually all fitted EPs are underamplified which results in a steep drop at low thresholds, far faster than for the PP curve. This allows to effectively discriminate EPs with moderate thresholds – h_{EP} lies between 14 and 19% pe for the R5912 and within 28 to 30% pe for the R11780-HQE. In fact, the resulting recommendations are close to the typically used thresholds of around 0.25 pe. For the R11780-HQE less EPs occur but the DCR is large, so the curve likely is strongly affected by regular-sized DCR pulses in the EP-window. This leads to a slower drop of η_{EP} and higher optimal thresholds. The decline of the remaining EP fraction quickens with rising voltage, since the EP probability drops, as was shown above.

For classic PPs far less undersized pulses are recognized and the initial drop is much less marked or even absent. The presence of pileup additionally raises the values at medium threshold, but leads to a stronger drop when approximating 1 pe. Again, the curve lies higher for the R11780-HQE due to the occurrence of less underamplified pulses and more pileup, and no distinct voltage dependence is observed.
Using a classic analysis, most EP effects are not recognized or form larger sum-pulses due to pileup, which leads to a much slower drop of η_{EP} and far higher optimal thresholds (see table 3.3). This is especially pronounced for the R11780-HQE, where the heavy presence of dark noise and the reduced EP probability lead to high values of h_{EP}. The f graphs which are obtained from classic evaluations are thus not suited for selecting the EP threshold, since crucial information about the physical origin of pulses is missing.

[40] Above a threshold of 1 pe likely too many pulses are discriminated for practical use.

3.2.2.3 Pulse shape

The pulse shape can be characterized by a few basic values which are based on certain points (10%, 50%, 90% peak height) on the rising and falling flank. This includes the rise time (10–90% on rising flank), fall time (90–10% on falling flank), length (10% rising to 10% falling flank), pulse width (50% rising to 50% falling flank), and peak width (90% rising to 90% falling flank).
For these shape properties the distribution of values can be studied and the mean value can be extracted from these histograms.
In addition, the pulse shape can be studied using the APS, which itself can also be analyzed with regard to the mentioned shape properties. Here the averaging is already performed during calculation of the APS, so only one value is obtained for each property.

Since the waveform has to be recorded to study the pulse shape, this was only possible with SPAX measurements.
The distributions will be analyzed for PPs, APs, and PPs without pileup. The latter is similar to the requirements for clean PPs, except that here subsequent APs that produce no pileup are *not* used as an exclusion criterion.
The pulse shape parameters which are discussed in the following are the values of the *sum-pulse* (i.e. as for a classic evaluation) even for a pulse-fitting analysis. This is because one cannot disentangle the pulse shape parameters of individual pulses within a pileup-pulse; for a fit with a single APS, the resulting values would all be identical to those of the APS. The classic values, on the other hand, give an indication for the distribution of properties and the influence of pileup. Therefore, the listed values are identical for fitting and classic evaluations, with two exceptions: For APs in a fitting analysis in case of pileup the sum-pulse value of the PP is used [41], while for classic pulses only sum-pulses after the first are taken as APs and pileup APs are not counted. Second, the used pileup cuts are basically ineffective for a classic analysis for the same reason.

The obtained distributions are shown in figs. 3.16 and 3.17, the APS graphs are plotted in fig. 3.18, and all relevant parameters are summarized in table 3.4.

Pulse shape parameter distributions

For the PP distributions of all properties a few common observations can be made. First, there are many events with values which are significantly higher than what would be expected from the spread of the dominant peak. This is a direct result of pileup from various EAP effects, which explains the different peaks and subfeatures. Second, there are also many events with too *low* values, whose provenance is not as intuitively clear. These events stem from small pulses, for which the determination of shape parameters suffers from a systematic problem: If the 10% point on the rising or falling flank lies within the noise band, it becomes impossible to determine its location. To define the time, instead the first datapoint (starting from the peak) which falls below this value is used, which leads to too small values. As a consequence, it is not possible to reliably determine the pulse shape values for small pulses if the

[41] Except for the rare case when the first fit does not find all constituent pulses, which is detected through pileup-cuts on the residual waveform. In this case the range is scanned again for pulses, and if any are found, their shape values are after subtraction of the other pulses using the scaled APS. This occurs for $< 0.35\%$ of EAPs.

Series	**R6594**	**R5912**				**R7081**	**R11780-HQE**	
Voltage [V]	+1670	+1275	+1425		+1725	+1520	+1648	+2092
Setup	LNGS	SPAX	SPAX	LNGS	SPAX	LNGS	SPAX	SPAX
Gain [10^7]		*3.79 · 10^6*	*8.87 · 10^6*		*3.58 · 10^7*		*7.47 · 10^6*	*3.39 · 10^7*
	$9.64 \cdot 10^6$	$6.11 \cdot 10^6$	$1.35 \cdot 10^7$	$1.29 \cdot 10^7$	$5.42 \cdot 10^7$	$1.34 \cdot 10^7$	$1.01 \cdot 10^7$	$4.60 \cdot 10^7$
Baseline RMS [mV]		0.24	0.31		1.29		0.34	1.45
Threshold [% pe]		*11.6*	*6.2*		*5.4*		*13.4*	*10.9*
	20	11.4	6.4	20	6.2	20	12.3	9.7
Fit threshold [% pe]		*2.7*	*1.4*		*1.2*		*2.8*	*2.3*
Voltage divider	6594-091009	E7694-01	E7694-01	E7694-01	E7694-01	E7694-01	E7693-01-MOD	E7693-01-MOD
CF-point [%]		55	55		55		54	55
Occupancy [%]	5.53	3.14	2.92	1.83	3.29	2.91	2.03	2.09
N	$1.73 \cdot 10^7$	$5 \cdot 10^6$	$5 \cdot 10^6$	$1.52 \cdot 10^7$	$5 \cdot 10^6$	$3.19 \cdot 10^6$	$5.02 \cdot 10^6$	$5.02 \cdot 10^6$
Rise time [ns]:								
PP		**3.33**	**3.10**		**2.76**		**3.48**	**3.07**
PP w/o pileup		2.96	2.77		2.46		3.22	2.81
APS		2.80	2.65		2.40		3.05	2.66
Fall time [ns]:								
PP		**5.03**	**4.86**		**4.72**		**5.90**	**5.69**
PP w/o pileup		4.51	4.50		4.50		5.64	5.38
APS		4.31	4.34		4.43		4.88	4.85
Pulse width [ns]:								
PP		**5.56**	**5.40**		**5.00**		**6.93**	**6.66**
APS		5.35	5.15		4.78		7.10	6.98
Length [ns]:								
PP		**10.12**	**9.70**		**8.94**		**11.52**	**10.92**
PP w/o pileup		9.09	8.78		8.23		10.61	9.94
APS		9.15	8.89		8.36		10.98	10.55
Peak width [ns]:								
PP		1.76	1.73		1.46		2.14	2.17
PP w/o pileup		1.62	1.51		1.27		1.75	1.75
APS		2.04	1.89		1.53		3.05	3.05
APS peak to CF-point [ns]		1.93	1.85		1.65		2.25	1.85
Pulse-fitting:								
Mean red. χ^2		*1.22*	*1.24*		*1.34*		*1.32*	*1.24*
Bad fits [%]		*0.06*	*0.18*		*0.53*		*0.22*	*0.76*

Table 3.4: PMT characterization results: Pulse shape.
See table 3.2 for measurement and formatting descriptions. Discussion in text.
The pulse shape values are from classic analyses, for fitted pulses the values would be identical to the average pulse shape (APS) by definition. Mean values for PPs and for PPs without pileup. Rise time t_r from 10% to 90% on rising flank, fall time t_f from 90% to 10% on falling flank, length from 10% on rising to 10% on falling flank, pulse width from 50% to 50%, peak width from 90% to 90%.

underlying height fractions lie within the noise band. This problem is most critical for the 10% points, which affects rise time, fall time and length. For very small pulses also the 50% points (pulse width) are concerned.
Since the AP rates which are found with pulse-fitting lie around 100%, the number of entries in the AP distributions (blue) is similar to that of the PPs (green). The AP curves mimic the PP functions but show more entries for large values, since a large sum-pulse has an increased likelihood to contain several pileup APs. In this case the same pulses is counted several times in the AP distribution. The number of entries for small values is reduced, since small sum-pulses are less likely to contain detectable APs due to their short length. On the other hand, if a higher fraction of APs than PPs has small heights, this can dominate and lead to higher relative numbers for APs at undersized values.

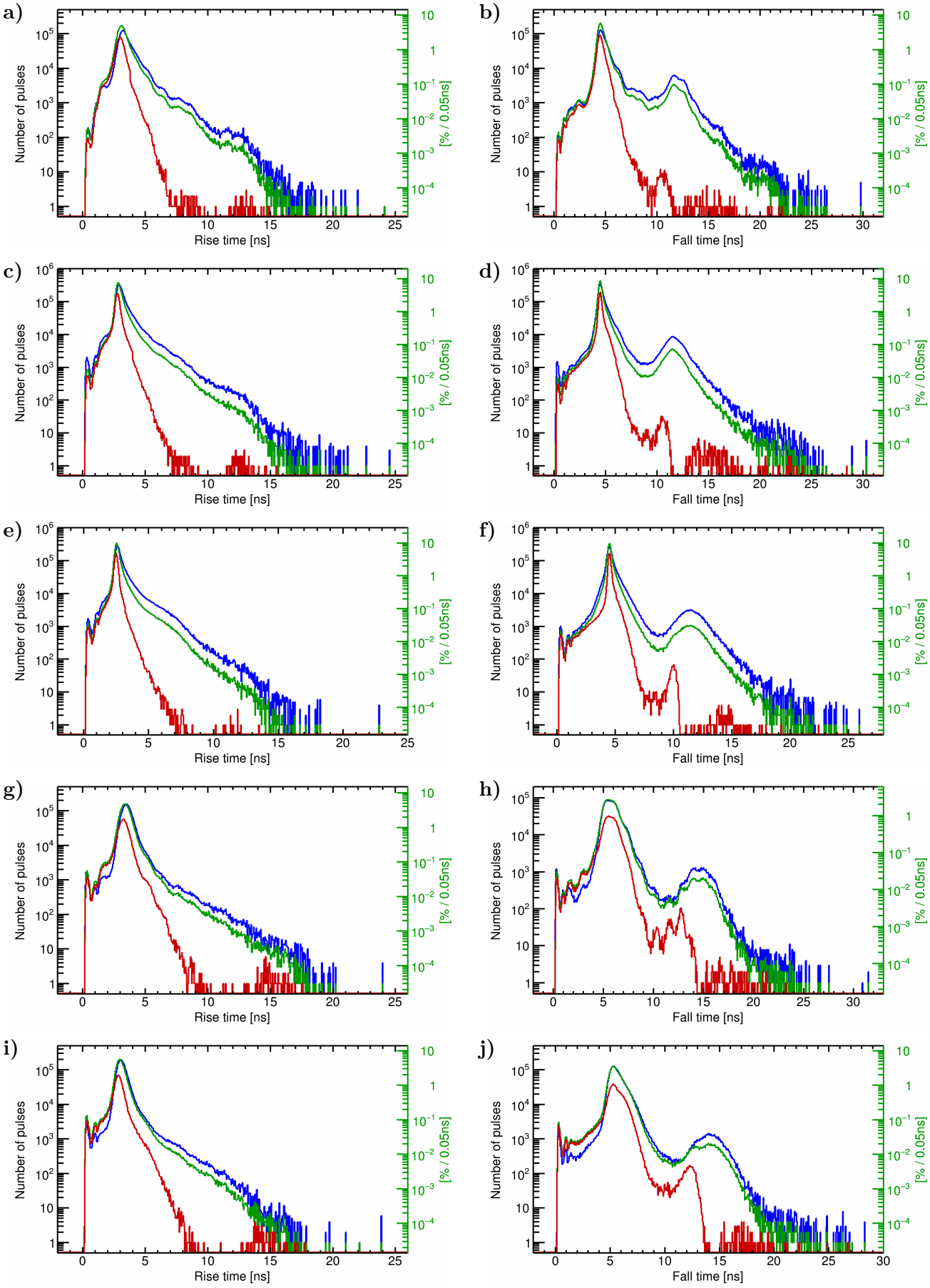


Figure 3.16: Pulse shape results from SPAX for pulse-fitting but using the sum-pulses: Rise and fall time for PPs without pileup (red; pileup information from subsequent pulse-fitting), PPs (green) and APs (blue). For fitted pulses no distribution but only a delta function at the value of the APS would result. Figures on left for rise time, on right for fall time. Fraction of all PPs per bin on right axis. Discussion see text.
a)–f) R5912: a) and b) +1275 V, c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

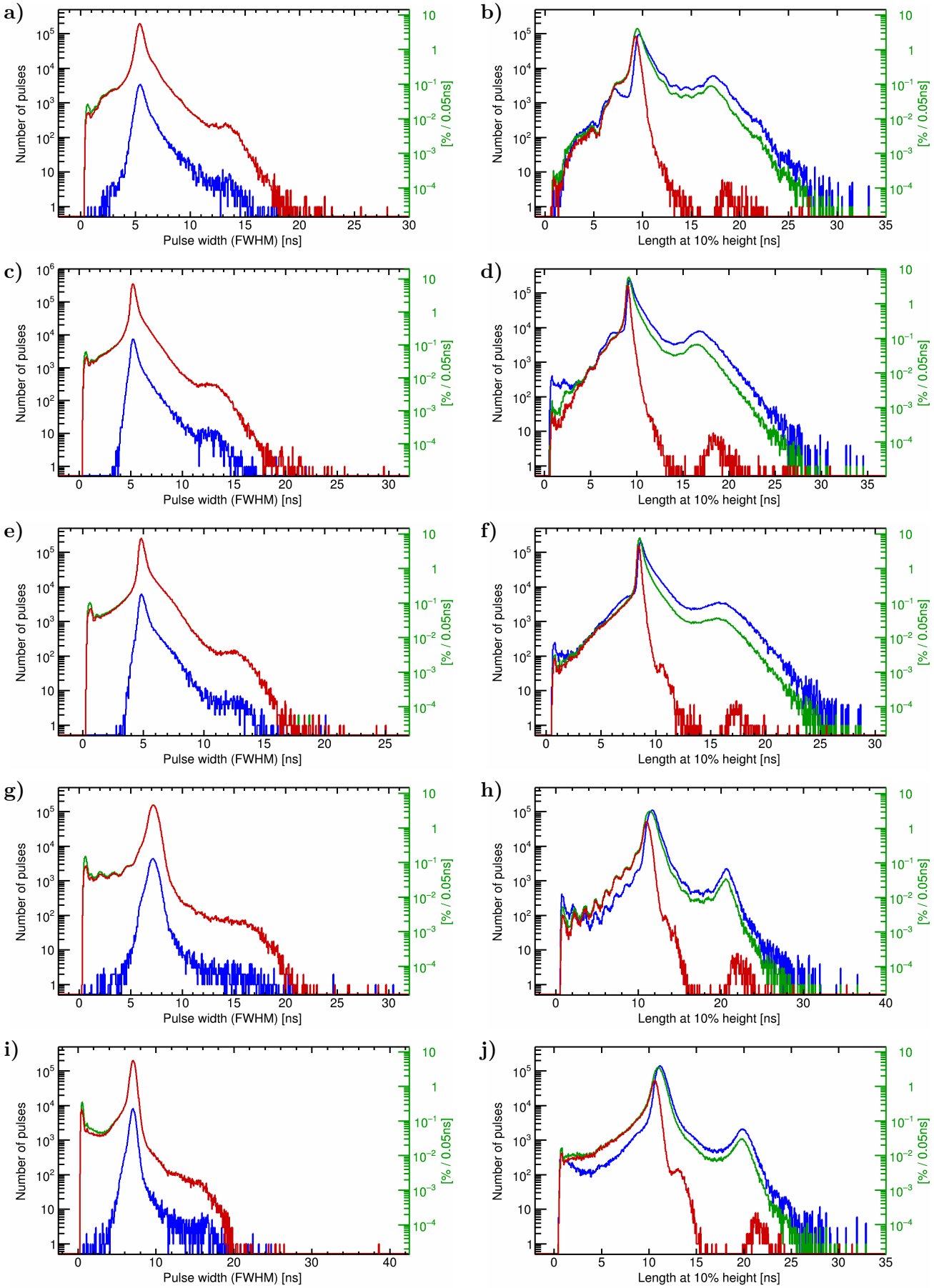


Figure 3.17: Pulse shape results from SPAX sum-pulses: Pulse width (classic analysis) and length (pulse-fitting) for PPs without pileup (red; pileup information from subsequent pulse-fitting), PPs (green) and APS (blue). For fitted pulses only a delta function at the value of the APS would result. Figures on left for pulse width, on right for length. Fraction of all PPs per bin on right axis. Discussion see text.
a)–f) R5912: a) and b) +1275 V, c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

Rise time

For the PP rise time (green in fig. 3.16) the largest observed values are more than six times as big as the peak rise time. Such extreme values result from small PPs which overlap with a large EAP shortly before they would disappear in the baseline noise. Thus, the rise time is the sum of almost the complete length of the PP and the rise time of the AP. For larger areas of overlap between PP and AP, smaller rise times result, which fill the range between the largest values and the peak.

For the R11780-HQE the peak lies at higher values and is broader than for the R5912. With growing voltage the peak moves left and gets more narrow, for the right flank the end point moves to lower times and the substructures change. The undersized rise times strongly depend on the pe-threshold and width of the noise band of the respective measurement.
When all pulses are eliminated for which the pulse-fitting detected pileup (red curves), this effectively eliminates most pulses with too large values, which proves that these in fact originate from pileup.

Fall time

The PP fall time peak position lies at larger values than for the rise time, since the falling flank is longer than the rising, as discussed in 2.4.
In addition to the regular events, a distinct peak is observed at about 11.5 ns (R5912) or 14.5 ns (R11780-HQE), whose position does not notably change with HV. This can be attributed to EAP effects (e.g. BL11, multiple s1L1) with short delays of around 5 ns and medium charges $\gtrapprox 0.2$ pe, whose APs fill up the short first overshoot in the APS of the PP (cf. fig. 3.18). The resulting sum-pulse ends with the overshoot of the EAP. This effect can be see in the waveform in fig. 2.52f.
Even larger fall times result when similar effects with slightly larger delays (e.g. BL11, BL1, s1L1) occur. In this case, the first overshoot of the PP is completely filled up and the first AP minimum is canceled by the second PP peak in the APS. Then the second PP peak also is part of the fall time, which greatly increases its value.
For the R11780, the peak is at higher values and far broader than for the R5912.
With rising voltage the position of the dominant peak does not change much but gets narrower, and the end point of the right flank also move to shorter times. For undersized values and APs a similar behavior as for the rise time is found.
Pileup elimination (red curves) again leads to a strong reduction of large values but is not as effective as for the rise time. The reason for the remaining events above 10 ns are strong noise spikes which occur shortly before the onset of a regular sized pulse and just exceed the threshold. These spikes are not discarded as noise by the algorithm, since due to the rising flank of the subsequent (true) pulse their charge/height ratio lies above the noise cut. Therefore, the pulse peak which is found in a classic analysis is that of the noise spike, and the decrease to 10% of its value does not occur until after the true pulse.

Pulse width

The pulse width was only determined via a classic analysis and not with pulse-fitting [42]. Since the pileup information from fitting is missing, the distribution for PPs without pileup (red) is virtually identical to that of all PPs (green).

[42] The determination of the pulse width was implemented later than the other shape parameters and would have required an extensive reanalysis, since pulse-fitting is numerically costly.

Also far less APs are found in the classic evaluation, but the shape is virtually identical to that of PPs.
All in all, the pulse width shows a similar behavior as the fall and rise time, which is not surprising, since the involved pulse regions overlap.

Length

The length curves are similar to those of the width, only more extreme. As the length is equivalent to the sum of rise time, peak width and fall time, all effects mentioned before are also encountered here. The largest observed lengths lie above 30 ns – more than three times the regular value – and therefore must stem from more than three serially overlapping pulses. The pileup cuts thus prove to be very efficient.

Average pulse shape

In the plots of the APS in fig. 3.18 it is visible that the rise time is generally shorter than the fall time (see also 2.4).
In addition, for the employed combinations of PMT and VD strong oscillations occur after the pulse peak. This is due to the VD design (see p. 149) and possibly to a smaller extent also because of reflections in the readout cables (e.g. the feed-through connector in the dark box). The latter can be basically excluded, since the position and size of the extrema varies with voltage and PMT, while it should stay mostly constant for reflections[43]. These oscillations introduce a series of maxima, which produces multiple fake APs for every pulse. This requires high exclusion cuts for the pulse charge and charge fraction unless the APS is subtracted for each pulse. The second peak is particularly problematic due to its large height. The presence of the first minimum and maximum after the prime peak introduce large variations in fall time, pulse width and length as mentioned above.
The substructures on the falling flank (particularly strong for the R11780-HQE and the R5912 at 1725 V) might also originate from internal oscillations. Alternatively, they could be caused by frequent TSEs and EAPs which lie on the falling flank (see discussion of fall time values below). In any case the distance of these structures to the peak is too short to stem from the setup.
Interestingly, for the R5912 a very small peak is encountered 7–9 ns before the main peak. This is too early to be the Shockley-Ramo pulse (cascade passing anode mesh on its way from d9 to d10, see p. 171), but might be from induction in the anode by the electron cascade when approaching d9. This would match with TST transit time calculations. For the R11780-HQE, instead a foot extending up to 7–8ns before the peak is observed. This probably results from the same effect which for the R11780 overlaps with the broader rising flank of the APS.
With growing voltage the rise time decreases, while the fall time stays approximately constant. In addition, a bump in the right flank appears. Furthermore, the first minimum deepens and the duration of the overshoot increases.
As already became clear from the analysis of the shape parameter distributions, the R11780-HQE has a broader pulse shape than the R5912. This is confirmed by comparing the APS curves for similar gains (for convenience placed side by side).

[43] Barring variations in the actual pulse shape (without oscillations), which alter the superposition with its reflections.

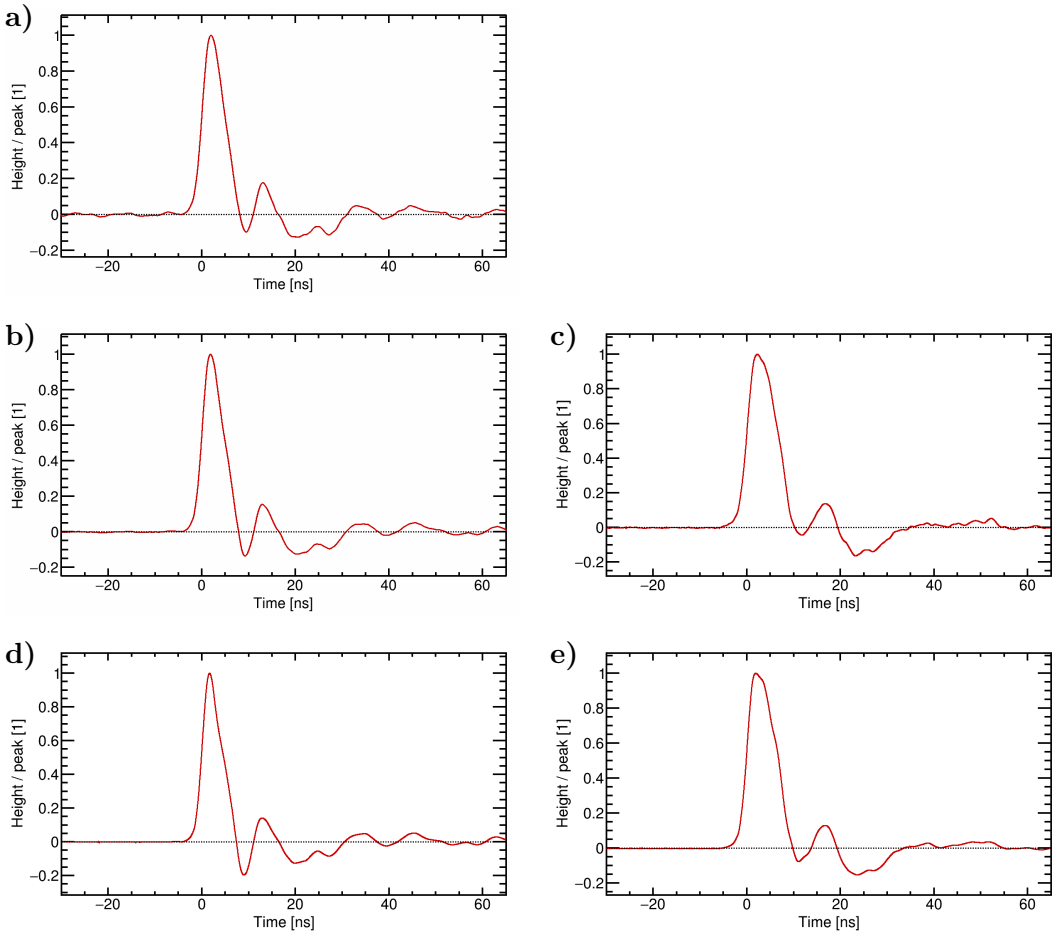


Figure 3.18: Average pulse shape from SPAX measurements. Height normalized to 1, time 0 at CF-point (see table 3.4). Discussion see text.
R5912: **a)** +1275 V, **b)** +1425 V, **d)** +1725 V. R11780-HQE: **c)** +1648 V, **e)** +2092 V.

Pulse shape parameter values

When comparing the mean values from the distributions and the numbers from the APS, it is important to note the differences between the APS and PPs without pileup: The APS was obtained using strict pulse shape cuts to eliminate pileup. On the other hand, PPs without pileup are found without using pulse shape cuts. Instead, the APS is fitted to the waveform and it is analyzed whether overlapping pulses are found (most often) or if after pulse subtraction the residual waveform indicates unrecognized pulses (seldom, for $< 0.35\%$ of pulses). Since different methods were used to eliminate pileup, there are slight differences between the values obtained for the APS and from PPs without pileup.

Rise time

The rise time t_r is smaller for PPs without pileup than for all PPs (as expected) and smallest for the APS. The discrepancy between APS and pileup-less PPs might be from using too rigorous shape cuts for the pulse selection of the APS. As a possible future improvement, pulse-fitting could be included in the APS calculation. Then it would be possible to exclude pileup via pulse-fitting and to loosen the shape cuts, which might allow to achieve even better χ^2 fit values.

The rise times are larger for the R11780-HQE and fall with rising voltage for all definitions.
In the literature, numbers of 3.1–3.2 ns are reported for the R5912 for $g = 10^7$, which agrees with the classic value without pileup removal. The value of the manufacturer (3.6 ns) is a bit larger. For the R11780 the datasheet gives a rise time of 4 ns which also lies a bit above the measured PP value.

In the table also the time between peak and CF-point in the APS is quoted to allow the conversion between CF-times and peak times. This quantity shows a similar behavior as the rise time but is shorter, since only a fraction of the rise time and part of the (smaller) peak width is contained in its range.

Fall time

For the fall time t_f a similar connection between PPs, pileup-less PPs and APS as for the rise time is found.
The values again are larger for the R11780-HQE.

With rising voltage a reduction of the fall time is observed for the R11780-HQE and for all PPs in the R5912, while the numbers remain constant for pileup-less PPs in the R5912 and slightly grow for the R5912 APS.
This can be understood by analyzing the TSEs which cause the larger length of the falling flank compared to the rising flank. As mentioned in 2.4, these effects are X_sL3/4 and SE transit time variations due to varying emission energy and direction. TST predicts a total EAP rate of 586% for the R5912 at 1425 V but "only" 109% are observed. Since the DD is matched well by calculated values at medium and large delays (see fig. 2.54) the unrecognized fraction must occur at short delays. This can be due to two effects. First, EAPs with very short delays can produce pileup which cannot be recognized despite pulse-fitting. Second, EAP effects which occur many times per PP produce a basically invariant contribution to the falling flank and can be considered part of the normal charge response. This statistical averaging even occurs between *different*, independent EAP effects.
A quick exemplary TST analysis for L3 shows that the backscattering fraction b3 will rise with voltage, the direction fraction L3 stays constant or increases a bit, the unscattered fraction $u3_{\varepsilon 3}$ decreases but is large, and $\delta 2$ grows, which leads to more creation attempts. In sum, more L3 pulses are expected for higher voltages. This tendency is the same for L4 and similar effects. Since these interactions have the highest calculated rates at short delays (see fig. 2.54), they can be expected to have the largest influence on the falling flank.
Regarding their influence on the fall time, on the one hand, the lateness of the effects which form the falling flank will decrease with growing voltage. On the other hand, effects on later dynodes also become more influential, since their mean occurrence rises. Both phenomena have contrary effects on the fall time and thus can lead to either an increase or a decrease, depending on which influence prevails.
For the R11780-HQE the lateness decrease dominates.
For the R5912, matters are more complicated. The APS is created by applying strict shape cuts, so the APS shape parameters behave similar to the peak in the respective distributions. For the R5912 the peak position of the fall time rises with voltage, from which it can be concluded that the increasing number of EAPs dominates. This is in line both with the findings previously presented for the R5912 and the growing

bump on the falling flank of the APS. It also becomes visible in the rising fall time values for the APS. The mean fall time shrinks, however, since for a larger HV more small pulses are detectable (more entries at undersized values). For pileup-less PPs both effects cancel and the numbers stay about constant. Finally, for the distribution of all PPs the fall time decreases with rising voltage, since the pileup of the effects on the falling flank is partially discriminated.

In literature fall times between 3.5 and 4.6 ns are reported for the R5912, which is close to but slightly lower than the measured PP value.

Pulse width

The pulse width t_w was obtained from a classic evaluation, which is why the values of PPs without pileup are virtually identical to PPs and are not listed in the table. The APS values are smaller than for all PPs for the R5912, but larger for the R11780-HQE for the following reason. The right flank in the length distribution is actually quite Gaussian for the R11780. Therefore, it was possible to put the upper cut which is used in the APS creation further above the peak than for the rise and fall time. This raises the value for the length and the related width.
The numbers are again larger for the R11780-HQE and, as predicted, a decrease with voltage is found.

Length

The pulse length t_l as anticipated is smaller for pileup-less PPs than for PPs. The APS values are a little higher than for PPs without pileup for the reason already mentioned for the pulse width. The values are larger for the R11780-HQE and decrease with rising voltage.

Peak width

The peak width t_p is defined here as the time between 90% on the rising and 90% on the falling flank. This can be directly calculated from the other values with

$$t_p = t_l - t_r - t_f \tag{3.4}$$

No separate distribution was created, and instead the tabulated values of the other properties were used to obtain the numbers for t_p which are listed in the table.
The resulting values are smaller for pileup-less PPs than for all PPs and largest for the APS. The latter is because the rise and fall times decrease more than the length when comparing the values from PPs to those from the APS.
As always the numbers are larger for the R11780-HQE than for the R5912. With growing HV a decrease is found for the R5912, while the values remain constant for the R11780-HQE.

3.2.2.4 Dark count rate

In table 3.5 the DCR which was measured for the longest available time in the dark is listed.

Series	**R6594**	**R5912**				**R7081**	**R11780-HQE**	
Voltage [V]	+1670	+1275	+1425		+1725	+1520	+1648	+2092
Setup	LNGS	SPAX	SPAX	LNGS	SPAX	LNGS	SPAX	SPAX
Gain [10^7]		*3.79 · 10^6*	*8.87 · 10^6*		*3.58 · 10^7*		*7.47 · 10^6*	*3.39 · 10^7*
	$9.64 \cdot 10^6$	$6.11 \cdot 10^6$	$1.35 \cdot 10^7$	$1.29 \cdot 10^7$	$5.42 \cdot 10^7$	$1.34 \cdot 10^7$	$1.01 \cdot 10^7$	$4.60 \cdot 10^7$
Baseline RMS [mV]		*0.24*	*0.31*		*1.29*		*0.34*	*1.45*
Threshold [% pe]		*11.6*	*6.2*		*5.4*		*13.4*	*10.9*
	20	11.4	6.4	20	6.2	20	12.3	9.7
Fit threshold [% pe]		*2.7*	*1.4*		*1.2*		*2.8*	*2.3*
Voltage divider	6594-091009	E7694-01	E7694-01	E7694-01	E7694-01	E7694-01	E7693-01-MOD	E7693-01-MOD
Time in the dark	9 h	30.8 h	4.7 h	3 h	17.3 h	11.5 h	17.0 d	18.0 d
Temp. [°C]	23	21.8	22.4	23	23.1	23	22.0	21.8
N	$1.73 \cdot 10^7$	$5 \cdot 10^6$	$5 \cdot 10^6$	$1.52 \cdot 10^7$	$5 \cdot 10^6$	$3.19 \cdot 10^6$	$5.02 \cdot 10^6$	$5.02 \cdot 10^6$
DCR TTD [kHz]		*0.42*	*0.61*		*0.72*		*6.24*	*7.84*
	4.95	**0.51**	**0.77**	**1.51**	**1.12**	**2.46**	**6.32**	**9.06**
DCR/A_k [Hz/cm^2]	45.5	1.35	2.04	3.98	2.95	4.64	8.70	12.5
DCR/A_c [Hz/cm^2]	52.1	1.81	2.73	5.33	3.95	6.48	10.3	14.7
DCR iAP DD [kHz]	5.18			1.69		2.66		
DCR/A_k [Hz/cm^2]	47.5			4.44		5.03		
DCR/A_c [Hz/cm^2]	54.5			5.95		7.01		

Table 3.5: PMT characterization results: Dark count rate. See table 3.2 for measurement and formatting descriptions. Discussion in text. A_k cathode area, A_c cathode cross-section.

The values were obtained from the unshadowed TTD and iAP-DD. Although for SPAX the DCR was also measured with a scaler before and after the measurements, using the measurement itself provides several benefits. No extrapolation or interpolation is required and from the measured distributions it is exactly known on which pulses the resulting value is based. This way noise becomes visible and can be excluded.
The SPAX iAP measurements were not used in the table despite having longer times in the dark than the timing measurements in some cases (see table 3.6), since the PP and AP threshold for iAP measurements is far higher. This results in too low DCR values.
In the LNGS measurements the iAP-DD fit contains more datapoints, which lowers the fit error, but still contains higher-order iAPs due to the comparatively short acquisition window of 30 µs. Therefore, higher values than from the TTD DCR fit result, which is why the TTD value is more reliable and was selected as final result.

It is difficult to obtain comparable values for the DCR since this property strongly depends on the illumination history (see 2.5.1). It is also questionable how relevant values after e.g. 24 h are; the DCR can take weeks to cool down more or less completely – for the R11780-HQE initially values above 1 MHz were measured and only after weeks of cool-down usable rates were achieved. Furthermore, the speed of decrease can vary between PMTs [336], which makes it difficult to estimate at which value the DCR of a specific PMT will finally level out. In an experiment the PMTs will be in darkness for months at the same time. Therefore, the value which was measured after the longest time in the dark is listed as most reliable estimate of the achievable DCR. The results for $g \approx 10^7$ are used as final values for the series selection.

The classic TTD DCR increases with PMT diameter with the exception of the R6594, whose values are almost as high as for the R11780-HQE. The latter exhibits a large DCR even after 17 d in the dark, likely due to its HQE cathode (see p. 56). The DCR rises with growing voltage, since more underamplified pulses slip above the threshold.

The values from SPAX are lower than from LNGS data despite the far lower threshold in SPAX. This is partially due to a longer time in the dark but likely is mostly caused by a lower amount of illumination in the weeks before the measurement.
Fitted PPs show smaller DCR values, probably because the fit excludes noise spikes. In the literature (see table 3.1) a DCR of 1.3 kHz is reported for the R5912, which is larger than measured in SPAX but in line with LNGS data. The manufacturer quotes a value of 4 kHz. For the R7081 2.2 ± 0.5 kHz is given, which agrees with data; the manufacturer value of 7 kHz again is higher. For the R11780-HQE 4.4 ± 1.9 kHz is reported, which is in agreement with the measured value, while the datasheet lists 15.5 kHz, which corresponds to shorter times in dark.

The PMTs can be better compared by calculating the DCR per area, since the influence of the cathode diameter is eliminated. Here, the DCR per cathode area is the most realistic value. As expected, this value is highest for the R6594 by far, followed by the R11780-HQE and the R7081, while the R5912 shows the lowest values. The number increases with rising voltage for the same reason as the DCR.
In addition, the DCR per cathode cross-section was calculated. This shows the same behavior as the DCR per cathode area, but yields higher values since the cross-section is smaller than the surface of the bulbous photocathode.
Before the R6594 is excluded because of its large DCR per area, however, further measurements should be performed at longer times in the dark and for multiple specimens – unfortunately neither literature nor manufacturer provide a value for this PMT.

3.2.2.5 Afterpulses

The results for EAPs, iAPs and LAPs are summarized in table 3.6 and will be discussed together with the respective distributions and graphs in the subsequent sections.

Early afterpulses

Early afterpulses could only be measured with SPAX. Attempts which were carried out earlier at the LNGS were not successful due to problems with noise, the PP length and APS oscillations.
The distributions which best characterize EAP behavior are shown in fig. 3.19 (delay distribution, DD), fig. 3.20 (charge–delay distribution,CDD) and fig. 3.21 (charge fraction–delay distribution, FDD).

For the classic SPAX evaluations, hard cuts on the minimum charge, charge fraction and charge–height ratio of APs were needed in order to eliminate fake APs from APS oscillations. Otherwise, for regular sized pulses every oscillation maximum would be registered as an AP. This severely limits the detectable AP charge and most undersized EAPs (e.g. γn1) become undetectable. The charge fraction cut also affects γnk effects (charge fractions around 0.5) stronger than BL1 (fractions above ≈ 0.6). This raises the fraction of large size EAPs from BL1 compared to the corresponding fitted analysis, where γnk has a similar or larger prevalence (compare plots in fig. 3.21). In addition, these cuts cause a vertical axis offset in the classic CDD and FDD figures compared to the fitted distributions. Note the

Series	**R6594**	**R5912**				**R7081**	**R11780-HQE**	
Voltage [V]	+1670	+1275	+1425		+1725	+1520	+1648	+2092
Setup	LNGS	SPAX	SPAX	LNGS	SPAX	LNGS	SPAX	SPAX
Baseline RMS [mV]: EAP		0.24	0.31		1.29		0.34	1.45
iAP			1.41 / 1.40		5.52		1.40 / 1.44	5.50
Gain [10^7]		*3.79 · 10^6*	*8.87 · 10^6*		*3.58 · 10^7*		*7.47 · 10^6*	*3.39 · 10^7*
	$9.64 \cdot 10^6$	$6.11 \cdot 10^6$	$1.35 \cdot 10^7$	$1.29 \cdot 10^7$	$5.42 \cdot 10^7$	$1.34 \cdot 10^7$	$1.01 \cdot 10^7$	$4.60 \cdot 10^7$
Threshold [% pe]:								
PP + AP EAP (fits)		*11.6 (2.7)*	*6.2 (1.4)*		*5.4 (1.2)*		*13.4 (2.8)*	*10.9 (2.3)*
		11.4	6.4		6.2		12.3	9.7
PP iAP	20		29.8 / 27.7	20	28.3	20	56.2 / 53.6	43.3
AP iAP	20		34.4 / 34.1	20	32.9	20	68.9 / 68.1	41.6
Occupancy [%]: EAP		3.14	2.92		3.29		2.03	2.09
iAP	5.53		3.47 / 3.45	1.83	2.61	2.91	2.23 / 2.07	2.16
Time in dark: EAP		30.8 h	4.7 h		17.3 h		17.0 d	18.0 d
iAP	9 h		5.1 d / 6.8 d	3 h	21.9 d	11.5 h	9.2 d / 14.2 d	7.1 d
DCR [kHz]: EAP		0.51	0.77		1.12		6.32	9.06
iAP	4.95		0.26 / 0.41	1.51	0.39	2.46	1.54 / 4.72	2.19
Temp. [°C]: EAP		21.8	22.4		23.1		22.0	21.8
iAP	23		22.5 / 22.2	23	22.5	23	22.4 / 22.1	22.6
N: EAP		$5{\cdot}10^6$	$5{\cdot}10^6$		$5{\cdot}10^6$		$5.02{\cdot}10^6$	$5.02{\cdot}10^6$
iAP	$1.73{\cdot}10^7$		$1{\cdot}10^6$ / $1.5{\cdot}10^6$	$1.52{\cdot}10^7$	$1{\cdot}10^6$	$3.19{\cdot}10^6$	$1{\cdot}10^6$ / $1.5{\cdot}10^6$	$1{\cdot}10^6$
EAP								
Delay window [ns]		*0 to 128*	*0 to 170*		*0 to 135*		*0 to 173*	*0 to 172*
		0 to 128	0 to 141		0 to 135		0 to 173	0 to 172
r_a [%]		***95.3***	***108.8***		***111.4***		***102.1***	***105.9***
		1.81	2.22		2.60		2.69	3.88
$r_{a,nq}$ [%/pe]		*92.4*	*105.6*		*107.8*		*100.0*	*103.7*
		1.76	2.16		2.52		2.62	3.78
Clean pe		*144.1*	*179.8*		*184.8*		*127.7*	*140.5*
		1.99	2.53		2.99		2.78	4.12
$r_{a,qn}$ [% pe]		*51.6*	*50.1*		*45.6*		*43.2*	*46.7*
		1.84	2.36		2.82		2.99	4.52
Clean pe		*32.2*	*28.6*		*25.6*		*34.2*	*35.0*
		1.62	2.00		2.37		2.82	4.16
$r_{a,q}$ [% pe/pe]		***49.9***	***48.7***		***44.1***		***42.4***	***45.7***
		1.78	2.29		2.73		2.91	4.40
Clean pe		*48.7*	*47.2*		*42.5*		*42.8*	*46.5*
		1.77	2.28		2.72		2.92	4.42
p_a [%]		***61.5***	***66.3***		***67.2***		***64.0***	***65.3***
		1.82	2.23		2.62		2.69	3.90
Photonic AP r_a:								
all (lower limit)		*14.8*	*16.4*		*17.2*		*11.7*	*13.6*
		0.26	0.32		0.48		0.40	0.71
γnk (lower limit)		*0.29*	*0.29*		*0.31*		*0.46*	*0.53*
		0.20	0.21		0.37		0.27	0.51
γn1 (lower limit)		*11.9*	*13.6*		*14.2*		*9.41*	*10.7*
		0.00	0.00		0.00		0.01	0.01
Pileup probability [%]		***63.3***	***65.4***		***65.0***		***66.7***	***67.8***
		0.14	0.13		0.19		0.36	0.76
Mean charge [pe]		*0.54*	*0.46*		*0.41*		*0.43*	*0.44*
		1.00	1.04		1.07		1.09	1.14
Mean height [pe]		*0.53*	*0.45*		*0.40*		*0.42*	*0.44*
		0.97	1.02		1.04		1.08	1.12
h_{EAP} [pe]		*0.38*	*0.30*		*0.28*		*0.27*	*0.25*
		0.93	0.93		0.93		0.99	0.99
iAP								
Delay min. [ns]	93.5		108 / 100	113	108	147	112 / 100	112
max. [µs]	29.9		95.0 / 99.8	29.9	95.3	29.9	95.4 / 99.7	95.4
r_a [%]	**0.97**		**(1.70) / 5.36**	**6.30**	**(2.87)**	**4.93**	**(2.51) / 12.1**	**(9.27)**
$r_{a,nq}$ [%/pe]			(1.70) / 5.18		(2.87)		(2.52) / 11.9	(9.31)
Clean pe			(1.82) / 5.85		(3.07)		(2.70) / 10.5	(10.5)
$r_{a,qn}$ [% pe]			(3.05) / 10.8		(5.09)		(10.5) / 37.7	(40.7)
Clean pe			(2.77) / 9.10		(4.66)		(9.23) / 48.8	(34.1)
$r_{a,q}$ [% pe/pe]			**(3.05) / 10.5**		**(5.09)**		**(10.6) / 36.8**	**(40.9)**
Clean pe			(2.96) / 9.92		(4.99)		(9.93) / 42.1	(38.8)
p_a [%]	0.96		(1.26) / 4.10	6.11	(1.90)	4.81	(3.40) / 8.86	(9.74)
Mean charge [pe]			1.73 / 1.87		1.69		4.51 / 3.35	4.41
Mean height [pe]			1.72 / 1.98		1.67		4.09 / 2.93	4.08
LAPs observed [yes/no]			**no / no**		**no**		**no / no**	**no**

Table 3.6: PMT characterization results: Afterpulses. — See table 3.2 for measurement and formatting descriptions and used voltage divider. SPAX AP coincidence and simple measurement results separated by slashes. Unreliable values in parentheses. Discussion see text. — Results after unshadowing and DCR subtraction. Gain for SPAX iAP measurements from corresponding timing/EAP measurement (higher accuracy). Clean PP charge of timing measurement used as clean pe. For LNGS relative AP transit time instead of delay. r_a AP rate, $r_{a,nq}$ AP number–charge rate, $r_{a,qn}$ AP charge–number rate, $r_{a,q}$ AP charge rate. p_a AP probability; for LNGS calculated from rate. Pileup probability: EAP p_a fraction lying before the first zero-crossing of the APS. h_{EAP} optimum threshold for EAP discrimination and PP preservation.

large values to which the artifact from the first oscillation maximum extends in fig. 3.20 (spike around 11 ns). A complete exclusion of these fake APs would eliminate all undersized pulses and a notable fraction of regular sized EAPs. Since with the shown cuts the remaining fraction of fake APs is small and is separated in time from the true effects, this is not reasonable.
This shows that pulse subtraction or, even better, pulse-fitting is far more reliable than a classic pulse search when the voltage divider produces oscillations.
The fitted analysis required none of the abovementioned cuts, with one exception. For the R5912 at 1275 V due to the low gain (larger influence of noise spikes) a minimum EAP charge of 0.06 pC (6.4% pe peak) was needed to eliminate artifacts from insufficient APS subtractions. The values in the table were obtained without a minimum EAP charge, since the occurrence of APS artifacts is far smaller than the EAP losses from charge cuts would be. For the plots a minimum charge of 0.06 pC was used to suppress oscillations.

Distributions

Delay distribution

The DD for fitted PPs was already analyzed in 2.5.2.2 for the R5912.
In the plots shown here a separation of first-order EAP effects in the three regions known from 2.5.2.2 – electronic EAPs, γn1, and γnk – can be recognized for all measurements. Please refer to fig. 2.54 for the TSEs which constitute the DD. At higher delays secondary γnk are visible in all plots. The plateau at the largest measured delays consists of APs – the DCR is already *subtracted* in the plots. As was shown in 2.5.2.3, delays above ca. 100 ns stem from a combination of the earliest iiAPs (majority of events) and high-order photonic EAPs, whose distribution in delay is flattened by the multiple auto-convolution.
For the R5912 at 1425 V at large delays small oscillations from suboptimal APS subtractions are visible. This could be eliminated by introducing a small AP charge threshold as was done for 1275 V, but would come at the cost of discriminating a significant amount of undersized EAPs.

For the R11780-HQE the second peak is more prominent than for the R5912. This is either because the first peak is suppressed more by the minimum resolvable delay or because the probability of X_sLn effects for $n \geq 2$ is higher in this PMT.
Also, a more rapid drop of γn1 is observed. This is also visible in the faster decrease of γnk-γn1 effects after the γnk range, which show the same behavior.
In general, the delays are longer than for the R5912 due to larger distances between stages at comparable voltages.
A higher number of γnk is encountered, and the γnk effects are stretched over a far larger delay range, possibly due to the additional contribution of later dynodes. For γn1 interactions, on the other hand, less events are found. This could be because d1 has a higher likelihood to reflect photons from later dynodes onto the cathode or because the solid angle under which the cathode is visible from e.g. d3 is larger. For the newer R12860 the latter is the case: Here an enormous "box" (d1) and a slightly curved "line"[44] (other dynodes) is used [236], which greatly increases the cathode visibility from d2 and d3. If the R11780, whose design is far newer than that of the R5912, incorporates a similar dynode layout, this could also explain why the drop for γn1 is earlier than for γnk.

[44] This resembles a circular and linear focused design, see p. 579

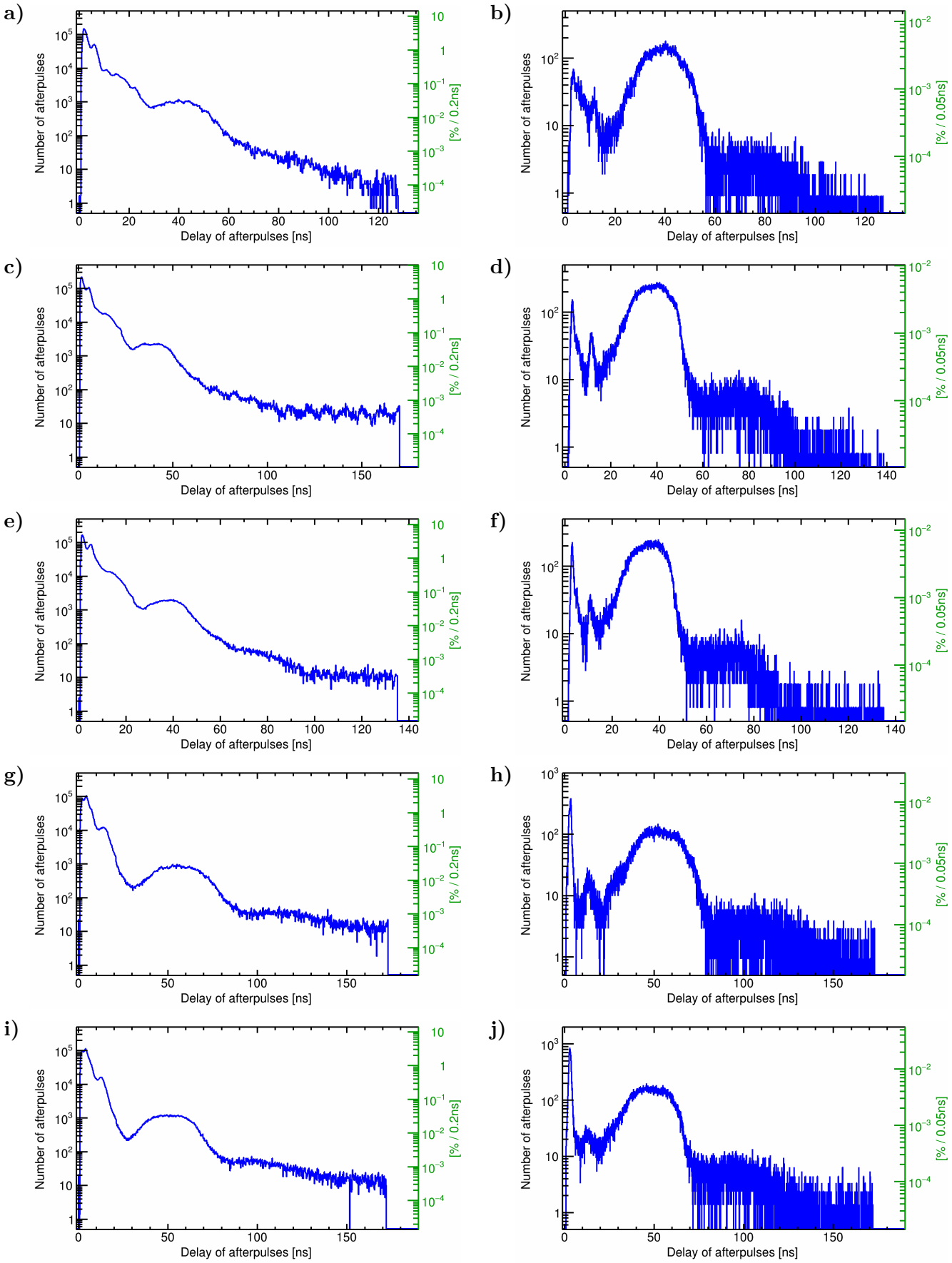


Figure 3.19: EAP results from SPAX: Delay distribution after DCR subtraction. Figures on left from pulse-fitting, on right from classic pulse search for same measurement. The spike around 10–13 ns for classic evaluations is artificial and results from the first APS oscillation maximum (see fig. 3.18). Note the different y-axis scale for fitting and classic analyses. Discussion see text.
a)–f) R5912: a) and b) +1275 V (minimum AP charge 0.06 pC), c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

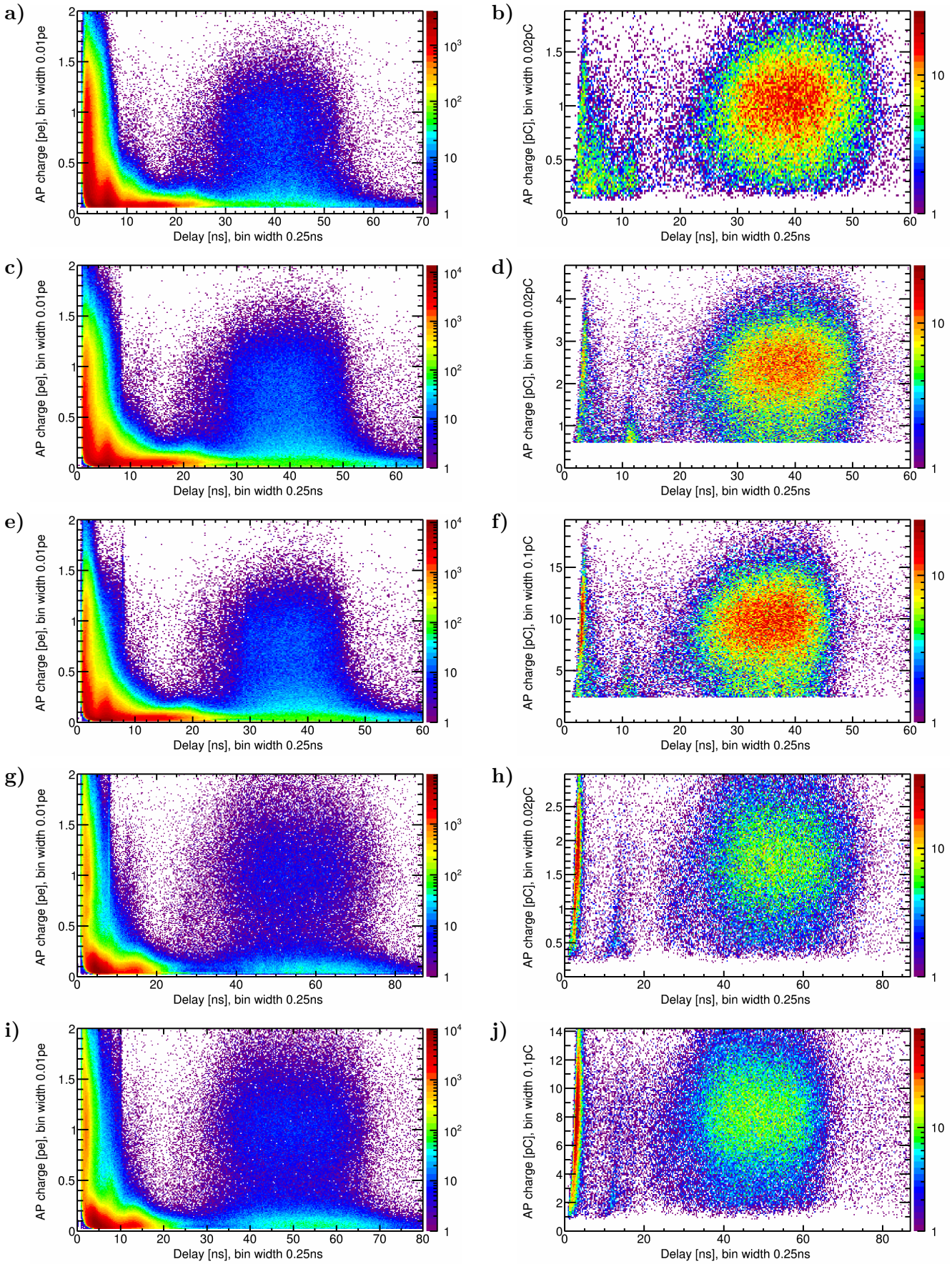


Figure 3.20: EAP results from SPAX: Charge–delay distribution. Figures on left from pulse-fitting, on right from classic pulse search for same measurement. y-axis in pe for fitting (using γk charge estimated with $q_{p,\mathrm{PPc}}$), in pC for classic. x-axis range adapted for fitting, fixed for classic plots. The z-axis denotes the number of pulses per bin. Note the different z-axis range between fitting and classic as well as the higher effective threshold in classic analyses due to cuts required to eliminate APS oscillations. The spike around 10–13 ns for classic evaluations is artificial (see fig. 3.19). Discussion see text.
a)–f) R5912: a) and b) +1275 V (minimum AP charge 0.06 pC), c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

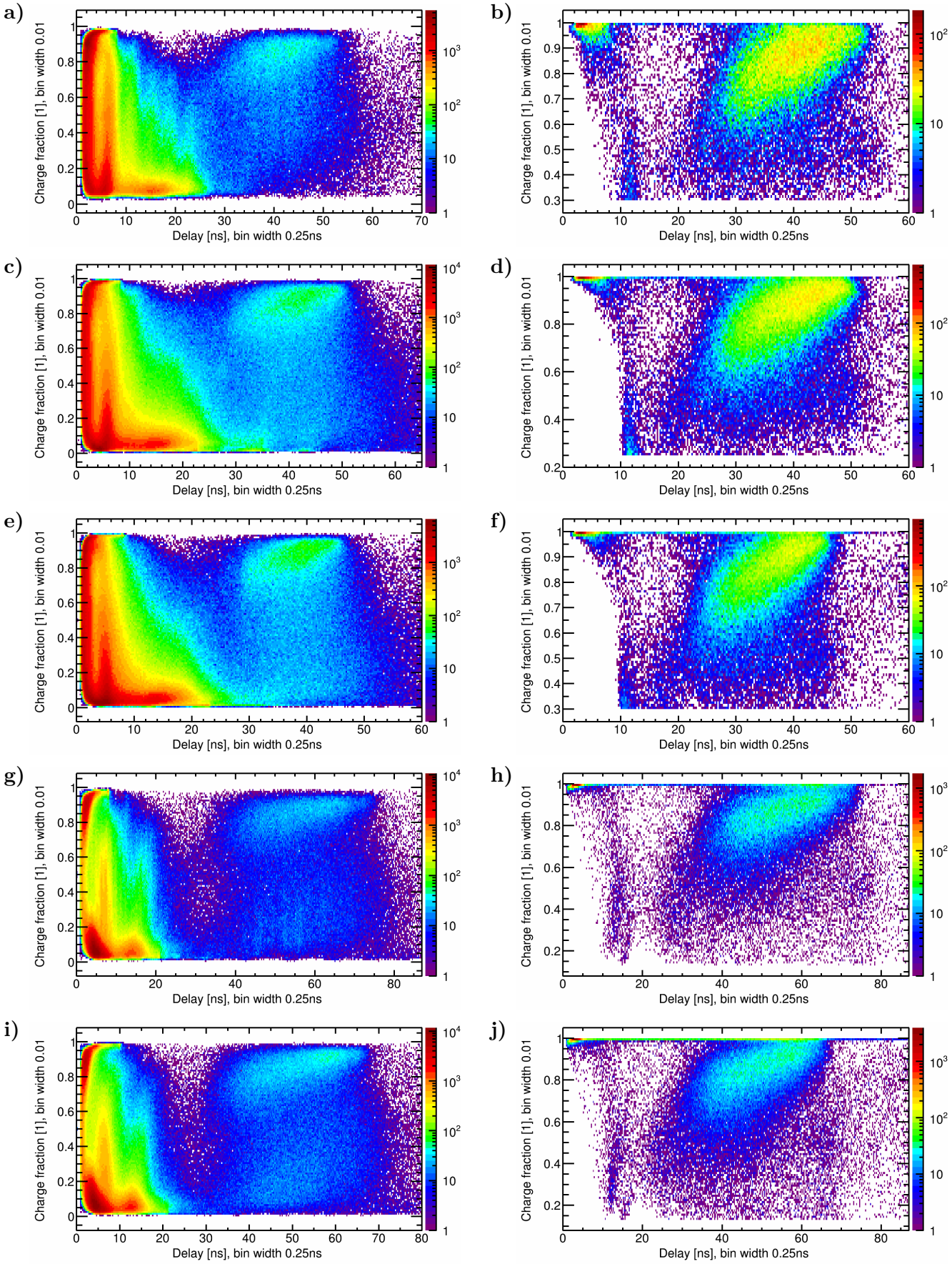


Figure 3.21: EAP results from SPAX: Charge fraction–delay distribution. Figures on left from pulse-fitting, on right from classic pulse search for same measurement. Range of x-axis adapted to voltage for fitting, fixed for classic. The z-axis denotes the number of pulses per bin. Note the y-axis offset due to cuts and the different z-axis scale for classic histograms. The spike around 10–13 ns for classic evaluations is artificial (see fig. 3.19). Discussion see text.
a)–f) R5912: a) and b) +1275 V (minimum AP charge 0.06 pC), c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

The delays of all structures in the DD decrease with rising voltage.
For the R5912 the second peak around 5–6 ns – which consists mostly of L2, BL2, BL4, and possibly also L3 and BL3 – grows with voltage. This is in line with the earlier prediction that the probability of L3 and similar effects will increase with voltage.
For 1275 V a small oscillation is superimposed on the curve in the γnk range (around 40 ns), which likely results from APS subtraction artifacts. However, it could also be caused by a better separation between individual γnk effects, since the delays are longer (due to the low voltage) and γn1 pulses are suppressed more (due to the high threshold). The measurement for 1275 V is difficult to compare to higher voltages due to its larger pe-threshold.
The γn1 and γnk contributions remain rather constant with rising voltage. At first, this is somewhat surprising, since the number of electrons which can produce these effects increases. However, this number rises far slower than the gain, since γn1 and γnk are dominated by the middle dynodes and the gains δn rise most for the last dynodes due to their small $U_{\mathrm{n,n+1}}$ (the SEY curve shows linear behavior in this region). In table 3.6, the pulses which can clearly be assigned to γnk and γn1 (the values in table are only a lower limit, see below) indeed show a small to moderate increase with growing voltage, which confirms this prediction. The secondary γnk region shows a stronger rise with voltage, since its rate is proportional to the square of the single γnk rate.
For the R11780-HQE, the L2 peak rises with larger HV. Also, γn1 and γnk increase notably with voltage (visible in the tabulated values as well), which is due to the different dynode geometry and $U_{\mathrm{n,n+1}}$ values. Correspondingly, the secondary γnk rise more than for the R5912.

In the classic DDs, γnk and second order γnk effects are visible in all curves. However, the number of events is slightly lower than in the fitted DD (note the different bin widths). This is because the hard cuts which are required to eliminate the APS oscillations affect even a fraction of the regular sized γnk pulses.
At the shortest delays, a small fraction of EAPs with delays lower than the pulse length are detectable even with a *classic* analysis for certain circumstances ($\mathcal{O}(10^{-3})$ of the values in the fitted DD). These pulses are also visible in the CDD and FDD, which helps to identify the effects causing these events. This shows that a classic evaluation can detect EAPs with short delays if an AP of sufficient size follows a small PP which only shortly exceeds the noise band (see 3.2.2.3) – longer lengths of the PP would cause pileup instead of two separate pulses. This is the case e.g. for BE1 and BL11. An analysis of the afterpulse transit time distribution (AP-TTD) with overlaid TST curves (not shown) reveals that the APs with delays from 0 to 10 ns in fact originate from BL11 and BE1, possibly also BL1.
Around 11 ns the abovementioned spike from the second APS peak is visible, which could not be completely discriminated without sacrificing nearly all interesting EAP events.
At large delays (for R5912 $\gtrapprox$ 90 ns, for R11780-HQE $\gtrapprox$ 130 ns) less events are found in the classic DD than in the fitted distribution. That is because the higher effective threshold from the charge cut eliminating APS oscillations also suppresses ixRn effects (undersized iAPs produced on the dynodes). This shows that the majority of EAPs at these large delays is in fact from iiAPs. This discrimination better brings out the third-order γnk, which are present at the same delays.

Charge–delay distribution

The fitted CDD was already discussed at length in 2.5.2.2, so here the focus is put on the voltage dependence and the changes occurring for the R11780-HQE.

With rising voltage all delays decrease. For γnk and γn1 an increase of the rate for higher voltages is observed more clearly than in the DD. BE1 and L11 show decreasing rates, as predicted.
The content of the cluster of s1L1 and low-delay BL1 pulses also decreases. This shall be analyzed once more with a qualitative TST treatment (p_f for R5912 at 1425 V in parentheses):

- **BL1 (4.1%):** For larger voltages BL1 will have a slightly higher backscattering fraction b1, the branching fraction B1 will increase to a limited extent, the direction fraction LB1 might rise marginally, and the unscattered fraction $\mathrm{u1}_{B1}$ decreases slightly. In total, this results in an increased probability. However, for small delays the direction fraction will drop severely, leading to a constant or even reduced probability for this part.
- **s1L1 (2.4%):** For s1L1 with higher voltage the creation probability c1 remains mostly unchanged or drops slightly, since δ1 is close to the SEY peak. The direction fraction $\mathrm{L1}_{\mathrm{s1}}$ becomes smaller, while $\mathrm{u1}_{\mathrm{s1}}$ and $\mathrm{c1}_{\mathrm{s1}}$ are unaltered, because the s1 energy distribution should remain mostly identical. Thus, for s1L1 a reduction with voltage is expected.

Since s1L1 should have a larger probability than the low-delay fraction of BL1 in the common cluster, from the above considerations a reduction of events at medium charges is expected, as observed in the measurements.
L2 lies at comparable delays and lower charges and has a voltage dependence which is similar to L3 for most probability factors (see above). However, since δ1 is about constant, in sum only a slight increase with HV is expected. This is confirmed by comparison of the distributions for different voltages (note the changing maximum of the color scale).
Finally, the artifact from the fit range which was mentioned in 2.5.2.2 is visible in form of a line at 8 ns delay.

For the R11780-HQE far less pulses are found at very short delays and medium charges than for the R5912. This is also visible below in the undersized pulses below γnk effects as these events are missing in the secondary EAPs.
BL11 shows a peak for regular charges and ca. 2.5 ns delay, which suggests that the extraction fields are stronger and backscattered pk with low energy are more likely drawn to d2.
There is also a small cluster at 0.65 pe and 2.5 ns, which is not visible in the R5912 CDDs and might result from a higher probability for large charges of BE1 events. This indicates that backscattered pk electrons with high energy are less likely to hit d2.
At low charges and small delays the probabilities which are observed for BE11 and BE1 are far smaller. For BE11 possible explanations are that scattered pk electrons cannot have notable lateness (i.e. t_{12} is more homogeneous over the d1 surface) or that they have a reduced CE_{12} when the part of d1 is hit which is further away from the cathode.

All this suggests a somewhat different dynode geometry than for the R5912, maybe closer to the R12860, as was already suspected above.
s1L1 and the low-delay part of BL1 are situated at smaller delays than in the R5912 – likely due to stronger d2 extraction fields – and in the R11780-HQE are separated in time from L2, which lies at later values.
γn1 shows an even more distinct peak for the middle dynodes than in the R5912 and a rapid decrease of the rate to earlier and later dynodes is observed.

For the classic CDD, γnk and BL1 are visible again. Furthermore, some secondary electronic AP and γn1 appear following γnk at lower charges, the start of secondary γnk is discernible, the spike from the second APS peak is present, and at the lowest delays BL11 and BE1 events are visible.
For the BE1 events the charge rises with growing delays due to the invisibility of pileup in a classic analysis: The larger the AP is (here an RP with reduced charge), the further away the PP must be to avoid the formation of pileup, in which case the AP could not be detected.
The contribution from BE1 and BL11 varies in the individual measurements. For the R5912 at 1275 V far more BE1 are encountered than for the other voltages and the R11780-HQE, where BL1 dominates. This agrees with the earlier finding that BE1 decreases with voltage.
For the cluster of γnk and BL1 the extent and values match those in the fitted distributions (note the different z-axis range). These are the only EAP effects which can be reasonably well detected with a classic measurement – with the mentioned restrictions due to the required strict cuts.

Charge fraction–delay distribution

The fitted FDD was already discussed extensively in 2.5.2.2 for the R5912, as well. Mostly the same behavior as for the CDD is observed.
The FDD reveals that a large fraction of events above 25–30 ns delays is from BL1. However, γnk effects cover a far larger vertical range and thus are dominating at these delays.
With rising voltage the BL1 fraction at large delays increases as predicted before. Otherwise the same changes as in the CDD are observed.

In the classic distributions, mainly BL1 is visible due to the γnk suppression by the cuts. BL1 shows the same bin content and extension as in the fitted distributions. Below 10 ns only pulses with high charge fractions are visible, since for these APs the PP only extends above the noise band for a few datapoints and is registered with a too small charge. Therefore, this delay range is not usable for the analysis of the charge fraction.
The spike from the second APS peak extends to quite high charge fractions.

Properties

The usable delay window varies between EAP measurement due to peculiarities of the different acquisitions. Its length ranges from 128 to 173 ns, but in every case the secondary γnk cluster is fully contained. Therefore, the number of additional APs which are detected by the measurements with longer EAP windows is very small (only iiAPs and 3$^{\text{rd}}$ order γnk events plus secondary effects) and causes only minor differences in the measured AP rates.

The AP rate and probabilities listed in table 3.3 were obtained after applying the occupancy correction, subtracting the DCR from the DD, and unshadowing the PP-TTD. These adjustments are straightforward and introduce no large errors.
Two corrections are problematic, however:
First, the unshadowing of the first-AP DD is unreliable due to correlations between the contained effects. This produces probability values of above 100% for all fitted analyses, which is unmathematical. For this reason the AP probability was calculated from the AP rate (which needs not be unshadowed) after applying the above corrections.
Second, in order to subtract the shadowed PPs from the AP number, it has to be known exactly how many of these PPs are detected as APs and how many produce pileup, which is not counted as extra pulse. This depends on the shape of the TTD and the minimum resolvable delay, which in turn are connected to the pulse heights in a complicated fashion. The true value will lie somewhere between the numbers with and without PP subtraction. In a future next step, the exact fraction of PP entering the DD could be calculated or approximated to resolve this uncertainty. However, the change of the number of APs through PP subtraction is not large, since this is only necessary for waveforms containing multiple PPs, which is efficiently suppressed by the low occupancy used in all measurements. For now, the values with complete PP subtraction are used to arrive at a conservative estimate of the AP occurrence. For classic analyses no PPs were subtracted from the DD, since additional PPs cannot be resolved in most cases anyway [45].

Rate

The fitted EAP rate r_a shows the exceptionally high values which are known from 2.5.2.2. This was explained by the high number of EAP-producing TSEs which can be resolved by the combination of SPAX and PEST, but not by classic measurements. r_a is a little higher for the R5912 than the R11780-HQE, but both PMTs exhibit values above 100% at $g \approx 10^7$. Only the small mean EAP charge – visible in a far lower charge rate $r_{a,q}$ of 43–50% – prevents the device from going into Geiger mode. The charge rate is still quite large and occurs mostly for very short delays, from which it can be concluded that a substantial part of a classic PP consists of pileup from EAPs and higher order EAPs. This is part of the regular charge response, and the classic gain would be significantly lower without the EAPs, as evidenced by the far lower gain values for fitted PPs in table 3.2.
The EAP rate r_a can be ascribed to three regions of first order EAPs. The region for the shortest delays, up to around 10.5 to 12.5 ns, is dominated by electronic EAPs (caused by cascade branching). The vast majority of EAPs – a rate of 80.0–91.0% – lies in this region, mostly electronic afterpulses with a small contribution from some early photonic effects (γ31, γ42, γ52, γ21, γ32, γ11, γ12). Only a few electronic EAP effects also extend to larger delays, most notably BL1 but also BL11 and L1L2L3. Following this, the region at medium delays (up to 27–32 ns) is primarily populated by γn1 and γn2 effects with small contributions from electronic APs and

[45] The large majority of PPs lies in the main peak. Multiple PPs only then do not cause pileup if the first PP is very small, e.g. E1 + γk, but in this case the recognition is suppressed by the low EP probability. Both PPs can also be detected for large transit time differences, for example γ1 + γk, or LPs with sufficient delays like γk + L1. However, these combinations are suppressed by the EP or LP rate as well. In theory *maximally* a fraction of 2-pe PPs which equals $p_{\mathrm{LP,p}} + p_{\mathrm{EP,G}}$ could be detected; using table 3.3 this is about 4–7%.

early γnk effects (γ1k, γ2k). This region contains a still large rate of 11.4–16.5%. Afterwards, at long delays the γnk effects dominate. Notable contributions also come from BL1, which is biased towards large delays, and to a lesser degree from late γn1 (γ91, γ10,1 and γa1) and γn2 (γ10,2 and γa2) effects. Furthermore, the rate in this region is raised by secondary EAPs from electronic effects, γn1 and γn2 which follow the γnk and γn1 at low charges and with short delays. In sum, a rate of 3.7 to 4.9% is observed for the γnk range. Finally, the region after γnk, at the longest delays, is filled with higher order γnk pulses and their short-delay secondaries. After about 100 ns delay, in addition the first iiAPs are present. In this region a small but still noticeable rate of 0.5–0.8% is observed.
The nq-rate $r_{a,nq}$ denotes the number of APs per PP charge and is similar to r_a. The qn-rate $r_{a,qn}$, in turn, shows the AP charge per PP and resembles $r_{a,q}$ (with small differences since the mean AP charge is smaller than the mean PP charge).
For the three rate definitions which involve charge sums ($r_{a,q}$, $r_{a,nq}$ and $r_{a,qn}$), the charge is expressed in pe and not in pC since this is gain-independent and directly relates the quantity to the typical PP charge. The mean PP charge is used by default, but values are also listed using the clean PP peak charge (labeled “clean pe” in the table), which is the best available charge estimate for γk, the standard process of PP generation. Since the peak charge of clean PPs is higher than the mean PP charge, the sum-charge of PPs or APs is lower for clean PPs when using the unit pe. Therefore, $r_{a,nq}$ is higher with the clean pe charge, while $r_{a,qn}$ is smaller. For $r_{a,q}$ both sum-charges are effected equally thus the value is almost unchanged except for the differences in the DCR subtraction, unshadowing and the PP subtraction from the DD.
The R11780 has a smaller r_a than the R5912, $r_{a,nq}$ is larger, $r_{a,qn}$ smaller, and $r_{a,q}$ is also smaller. The later three are the case because for this PMT the APs are on average smaller. For $g \approx 5 \cdot 10^7$ the relations are reversed, since now the mean AP-charge of the R5912 is lower than for the R11780-HQE. Regarding the clean pe values, contrary to the R5912, in the R11780 the clean PP peak charge is lower than the mean PP charge. This raises $r_{a,nq}$ and lowers $r_{a,qn}$, while $r_{a,q}$ is essentially unaffected.
With increasing voltage a rise is observed for r_a and $r_{a,nq}$ (the R5912 at 1275 V is not completely comparable due to the higher threshold). $r_{a,qn}$ and $r_{a,q}$ fall for the R5912 but rise for the R11780-HQE, because the mean AP charge increases with voltage for the R11780, but falls for the R5912. This likely is caused by a faster growth of γnk events in the R11780-HQE.

For classic evaluations the rate values are far lower since only the effects which were discussed for the classic DD are visible.
Again, $r_{a,nq}$ is similar to r_a, and $r_{a,qn}$ resembles $r_{a,q}$. The charge rate itself is in most cases larger than r_a, since almost only BL1 and γnk are detectable, which have (near) regular charges, and their sum-charge is raised above 1 pe by pileup from secondary EAPs.
For clean pe rates the same changes as for fitted pulses are found.
The values are generally larger for the R11780-HQE, because it shows a higher BL1 probability.
The rates rise with growing voltage, since the late part of BL1 increases with HV (see above).
In the literature, for the R5912 an EAP rate of 3.4 ± 0.6% is reported [335] (see table 3.1), which is similar to but larger than the classic rate measured here. This is most likely due to the higher gain of $1.6 \cdot 10^7$ with which the PMTs were measured

in [335]. It could also be due to statistical fluctuations between samples or because the (newer) specimen measured here has a lower EAP rate. The resistance ratios of the used VDs were the same as employed here.

Probability

For the EAP probability p_a, very high values between 62 and 67% are observed. These figures are notably lower than r_a, since the probability cannot exceed 1; r_a is the mean number of APs per PP, while p_a is the probability to observe one or more APs for a PP.
The probability rises with increasing voltage, and is slightly lower for the R11780-HQE, as was already concluded above.
For classic PPs, the values are virtually identical with r_a since the numbers are far below 100%.

Photonic afterpulses

In order to isolate the photonic afterpulse rate for fitted APs, several cuts were employed. The used delay window was limited to the range above electronic EAPs (classic analyses: above the APS spike) and its upper limit was set to include the secondary γnk events (above this the ixRn are present as well). To exclude APS artifacts, a minimum charge was required (in classic analyses this is already covered by the general cuts). Furthermore, the maximum charge fraction was limited to exclude the majority of BL1 pulses without losing too many γnk. The resulting photonic AP rate of 12–17% thus is only a lower limit.
Within this allowed parameter space, also the γnk and γn1 rates were determined by introducing additional charge and delay cuts to separate these event classes from one another and from secondary γnk. The resulting values for γnk (0.3–0.5%) and γn1 (9–14%) are also lower boundaries, since a large fraction of γnk is overlapping with BL1, and γn1 which lie within the electronic AP range are not counted either. The AP rate at γnk delays which is excluded by the charge fraction cut ranges from 1.6 to 5.0%, which is far above the remaining γnk rate and can be attributed to BL1 and overlapping γnk pulses. This shows that far more events are situated at γnk delays than indicated by the cautious γnk cuts, as was already visible in the classic r_a for the complete DD (which also suffered from event reduction by cuts).
The total photonic AP rate for fitted pulses rises with increasing voltage and is larger for the R5912. The fraction of γnk which is included in the cuts slowly rises with growing HV and is larger and increases faster for the R11780-HQE, as was already derived by other means above. γn1 constitutes the large majority of photonic APs and here also a rise with growing voltage is found, while the values are smaller for the R11780.
For classic pulses far lower total rates are found, since only γnk and secondary γnk are resolvable. The γnk rates are similar to the fitted analyses but lower due to cuts. γn1 is basically zero, since almost all small pulses are excluded by the cuts.

Mean afterpulse charge and height

In order to obtain reliable values for the mean charge and height of EAPs, the DCR was subtracted from the summed values of all pulses in the DD. Because no dark noise SER or PHD were recorded, the PP mean values were used as estimate for the mean of dark noise pulses.
For the pe-charge and -height the mean PP values were used.

For fitted pulses, the resulting mean charge ranges from 0.41 to 0.54 pe, meaning that the majority of EAPs are underamplified. The mean charge is slightly smaller for the R11780-HQE. It falls with growing voltage for the R5912 but minimally rises for the R11780-HQE, which (as mentioned above) likely is due to the faster γnk growth for the latter.
Classic mean charge values typically lie above 1 pe, since nearly only BL1 and γnk with their pileup are detectable. Here, values rise with increasing voltage, since the amount of pileup increases. For the R11780-HQE larger numbers are observed, as this PMT is more prone to create pileup.

The mean AP height closely resembles the mean charge and shows the same behavior.

Pileup probability

Pileup designates two or more pulses which overlap and thus produce a sum-pulse with larger charge and possibly height (depending on the delay between pulses and the APS shape). Even in a classic measurement, this can often be recognized by changes of the pulse shape parameters. However, a strict exclusion of all such pulses would lead to a significant reduction of the number of usable pulses – and the effective PDE – as will be shown in the following. Only via pulse-fitting can pileup be eliminated while preserving the PDE.
The following analysis estimates the probability of pileup by using the EAPs which were identified with pulse-fitting. The pileup probability could also be determined via hard pulse shape cuts, as used in PEST to obtain the APS.

The pileup probability was estimated by integrating the EAP probability within the delay range which would lead to pileup with the PP.
For this the DD_1 (delay distribution of first afterpulses) was used *without* DCR subtraction, unshadowing, and subtraction of the shadowed PPs. Dark noise and additional photon pulses can cause pileup as well and thus should be included. Unshadowing, in turn, would convert the probability into a rate and would introduce inaccuracies due to correlations.
To obtain the pileup probability, the integral of the resulting DD was divided by the integral of the complete TTD without unshadowing and DCR subtraction. Here, the unaltered TTD is used, since unshadowing would create PPs which were not recorded, and dark noise pulses can produce EAPs and pileup as well.
In order to determine the integration range, it was assumed that pulses can cause pileup if the point of 10% pulse height of the second pulse occurs at or before the first zero-crossing of the first pulse (see fig. 3.18). Then all pulses whose delays are smaller than the time between the 10% point on the rising flank and the first zero-crossing in the APS create pileup.
The resultant pileup probability can be slightly *higher* than p_a, because the values of p_a were obtained from the AP rate r_a, which was calculated from the distributions *after* unshadowing, PP subtraction and DCR subtraction.

The resulting pileup probability ranges from 63 to 68%, which means that only about a third of the PPs contain *no* pileup and have an unaltered pulse shape. Pulses without pileup occur for all TSEs which do not produce EAPs. Apart from γk, this can be e.g. γ1, E1, L1, or γm. As a consequence, γk itself – the process for which

the PMT is actually designed – contains only a small fraction of pulses (visible in form of the main peak for clean PPs): For the R5912 at 1425 V merely 31.1% of PPs are γk pulses. The high pileup probability also explains the great variation of pulse shapes which is observed for a PMT with an oscilloscope or FADC.
With growing voltage the pileup probability shows a slow rise, since the delays become shorter and the EAP rate increases, which is compensated in large part by the narrowing of the APS.
For the R11780-HQE the values of the pileup probability are slightly higher than for the R5912.

Actually, the pileup probability is far higher than the values calculated here, as already mentioned above: TST predicts an EAP rate of 586% but only a rate of 109% is observed – the rest is pileup. On the physical level, the predicted rate corresponds to actual interactions which cause cascade branching and photonic APs, and actually even higher numbers are expected if later dynodes are included in the calculations as well. These effects, however, quickly become unresolvable due to their small charges and high occurrences, and the border blurs between what is designated an EAP and what should be considered part of the regular charge response and the APS. The pileup probability calculated using the DD and TTD has the advantage that is considers all detectable pulses, which is a logical but threshold- and thus application-dependent solution to this ambiguity.

For practical applications, the relevant pileup probability will be lower, since higher thresholds will be used. The EAP fraction vs. threshold (see fig. 3.22) can give an indication of the threshold dependence and even at high thresholds of 25% pe a fraction of 28–42% of the EAPs remain. Assuming that the same fraction of EAPs will create pileup as for low thresholds, a pileup probability of still about 18–28% will result for this threshold.

Pileup is already part of the classic pulses (see section “Charge” on p. 493) and thus contributes to the currently known and scheduled PMT behavior. This raises the question to what extent the time and charge properties can be improved by pulse-fitting and the knowledge of the underlying time shift effects.
For future experiments with extreme requirements such as JUNO or THEIA, charge and time reconstruction play an ever increasing role. Thus, everything which can improve these properties by means of photosensor development or data analysis should be utilized in order to be able to further push the limits.

Regarding timing, as a first estimator the properties shown in table 3.3 can be used.
For fitted PPs and at typical gains, $T_{\mathrm{FWHM},\,s}$ is smaller than for classic PPs for the R5912 but larger for the R11780-HQE. $\sigma_{T,s}$ is higher for both. This behavior comes from the interplay of pileup removal and improved EP detection. This shows that already through pulse-fitting alone the TTS can be improved, depending on PMT and voltage.
Using clean PPs lowers the TTS, but has the disadvantage of a far lower PDE. $T_{\mathrm{FWHM},\,c,\,s}$ is far smaller than the classic value for the R5912, since many EPs are discriminated. For the R11780-HQE, whose EPs have very small earlinesses, similar TTS values as in a classic analysis are found.

Nevertheless, this is not the limit of the possibilities which are opened by the additional information available through the combination of pulse-fitting and TST: The identification of effects from the properties of PP and APs might allow to apply a transit time correction for the PP and to eliminate APs. This could improve timing and in particular enable the removal of EPs.

Considering the charge response, in the fitted PP SER far more underamplified pulses are recognized, since the pileup is resolved. This deteriorates the charge resolution on the one hand, but also provides a high 1 pe–2 pe discrimination power (see fig. 3.10) on the other hand.
Using the clean PP SER in many cases restores the resolution to the values of classic PPs and in some cases even improves it, but at the expense of losing the pulses with pileup or EAPs.
Effect identification should, however, allow to reconstruct the number of photons with greater accuracy and without losses. This will be most efficient for a low number of pulses, since an unambiguous attribution to effects becomes increasingly difficult with a growing number of pulses in the waveform. In particular, this should lead to an improved efficiency for the 1 pe–2 pe discrimination as evidenced already by the fitted-PP SER.

TST-based effect identification should prove another worthy endeavor for the future, as it would allow to improve the time and charge resolution by waveform analysis. In addition, the amount of pileup in PMTs or other photosensors could be reduced by adapting the sensor layout based on TST analyses.

Fraction of early afterpulses versus threshold

Similar to the EP fraction vs. threshold, in fig. 3.22 the remaining fraction η_{AP} of the AP rate r_a at zero threshold is plotted against the threshold from 0 to 1 pe. η_{AP} was determined by integrating the CDD for values above the current threshold after projection onto the charge axis.
The same definition of the figure of merit $f = \eta_{\mathrm{PP}} \cdot (1 - \eta_{\mathrm{AP}})$ is used here with η_{AP} instead of η_{EP} (compare eq. (3.3)), since $f = \eta_{\mathrm{PP}}/\eta_{\mathrm{AP}}$ would increases monotonically until the end of the plotted range as well. The peak position in f is used as the recommended threshold h_{EAP} for EAP discrimination.

For fitted APs, a steep drop of f is found at low thresholds due to the abundance of undersized EAPs. The slope, however, is not as steep as for the EP fraction in fig. 3.15, since for EAPs more pulses with medium and regular charges are present. After the drop the slope starts to level out, which is more gradual for the R5912 and results in a broad peak region with rather high h_{EAP} values (see table 3.6). For the R11780-HQE, a narrow transition to a nearly flat region is found, because in this PMT less EAPs occur at medium charges, which leads to higher fractions of pulses at low and high charges. This shifts the peak in f to the left and results in lower h_{EAP} values.
h_{EAP} is rather high and lies between 0.25 and 0.38 pe, which is higher than for h_{EP} (0.14–0.30 pe, see table 3.3). This suggests to chose a rather high threshold in the vicinity of 0.25 pe (at least for the two studied PMTs), which again is close to what is typically reachable and chosen. Interestingly, for the R11780-HQE the values for h_{EAP} are lower than those for h_{EP}, likely due to a contamination by dark noise pulses.

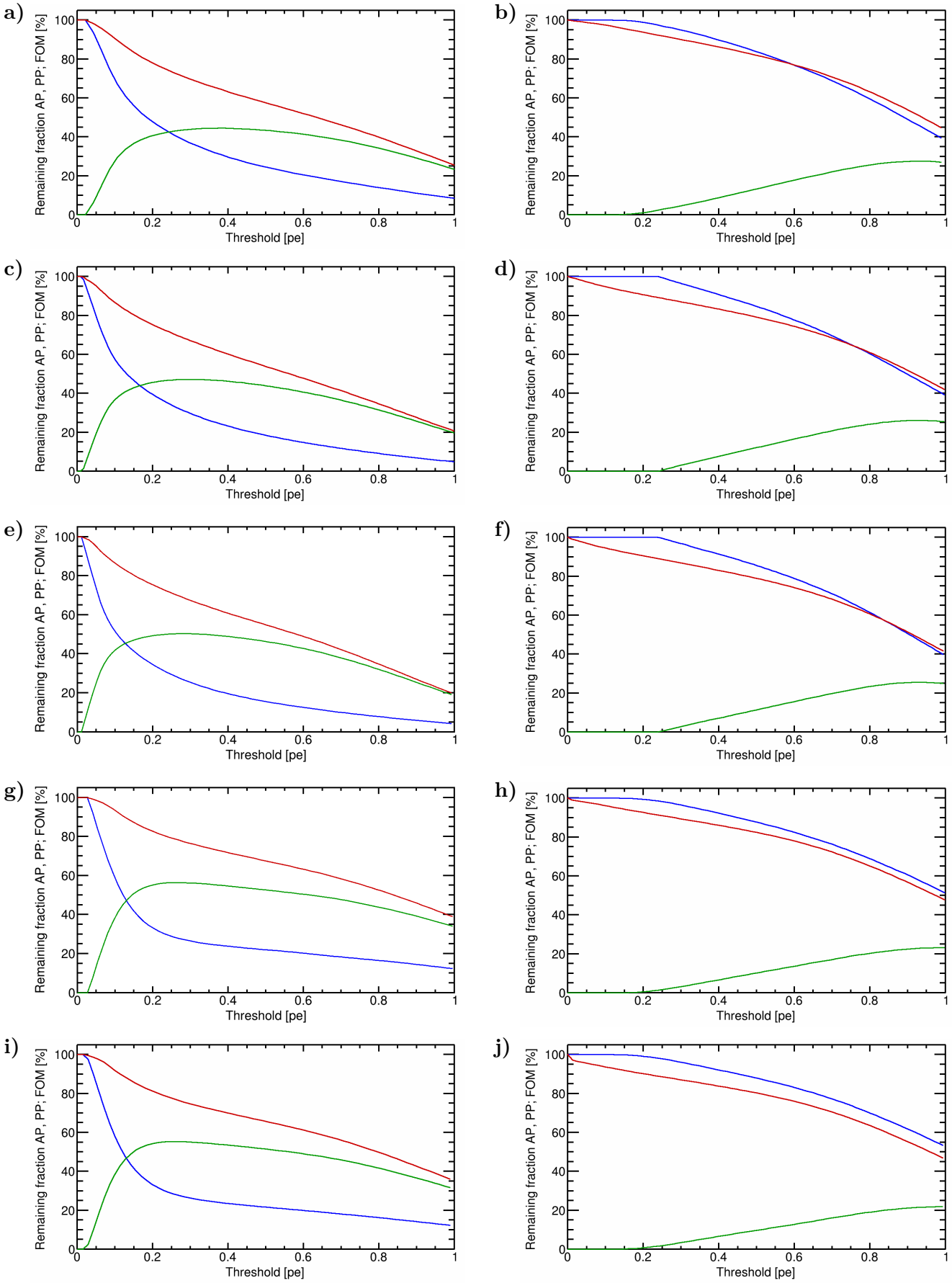


Figure 3.22: EAP results from SPAX: Remaining EAP (blue) and PP (red) fraction vs. threshold in pe (using γk charge estimated with $q_{p,\mathrm{PPc}}$). Figure of merit (FOM) $f = \eta_{\mathrm{PP}} \cdot (1 - \eta_{\mathrm{EAP}})$ in green. The recommended threshold h_{EAP} for EAP discrimination is defined as the threshold at peak FOM. Figures on left from pulse-fitting, on right from classic pulse search for same measurement. Discussion see text.
a)–f) R5912: a) and b) +1275 V, c) and d) +1425 V, e) and f) +1725 V.
g)–j) R11780-HQE: g) and h) +1648 V, i) and j) +2092 V.

The recommended threshold falls with growing voltage. This is because a higher gain allows a better EAP discrimination with pulse-fitting at typical thresholds, the charge reduction of small effects increases due to rising $\delta 1$ and $\delta 2$, and the mean EAP charge shrinks (R5912 only).

For classic pulses, mainly regular-sized EAPs remain (BL1, γk) due to the exclusion of small EAPs by the necessary APS-cuts [46]. Therefore, the EAP fraction curve closely resembles the PP fraction. The FOM f reaches a maximum shortly below 1 pe, since the BL1 charge is smaller than 1 pe, which leads to an optimum threshold for EAP exclusion near 1 pe. This value is unusable in practice – without pulse-fitting no reasonable EAP exclusion is possible.

Ionic afterpulses

Ionic afterpulses could be measured at both the LNGS and with SPAX. For the latter, due to waveform acquisition more data is analyzable, most importantly the charge information. On the other hand, for SPAX hard charge cuts were required due to instabilities of some components of the electronics, which eliminates many undersized iAPs. In SPAX iAP measurements were carried out both with the triple coincidence and without it ("simple" acquisition). In the latter case, due to the small iAP rate most waveforms contain no iAPs, but the iAP rate can be obtained directly from the measurement. This serves as a cross-check of the determination with the scaler.
The measured distributions are shown in figs. 3.23 and 3.26 (DD), 3.24 (CDD) and 3.25 (FDD) and the resulting values are summarized in table 3.6.

Distributions

The following discussion will be based on the analysis for the R5912 starting on p. 392 where the effects which are present in the iAP distributions already were studied in detail. The predicted delays were summarized in table 2.4. The calculations and the effect nomenclature which are used below can be found in 2.5.2.3.

For ik [47] in quadratic fields, the delay is independent of the creation position x_0 and depends solely on the k–d1 distance x_{k1}, the voltage U_{k1}, the ion mass m, and the ion charge Z. Therefore, in PMTs for which the quadratic field dominates the v_{k1} region, the delays of ik effects should be identical to the values calculated for the R5912 after multiplication with a fixed factor which accounts for the different x_{k1} and U_{k1}.
For inRx [48] which traverse several interstage regions, no notable smearing due to a varying start position x_0 was found. Therefore, this variation seems to be of secondary importance and the delay can be approximated for all inRx to be independent of x_0 as well. For inRx also x_{12}, x_{23}, U_{12} and U_{23} alter the delays to some extent, but in the first approximation the delays can be assumed to scale with those in the R5912 with the same factor as used for the ik delays.

[46] The charge cut is visible as the threshold up to which η_{AP} remains at 100%.
[47] Ions created in the vacuum between the cathode and d1 which hit the cathode and create pulses with regular or large charges.
[48] Ions created in the vacuum between the dynodes which hit a previous dynode or the cathode.

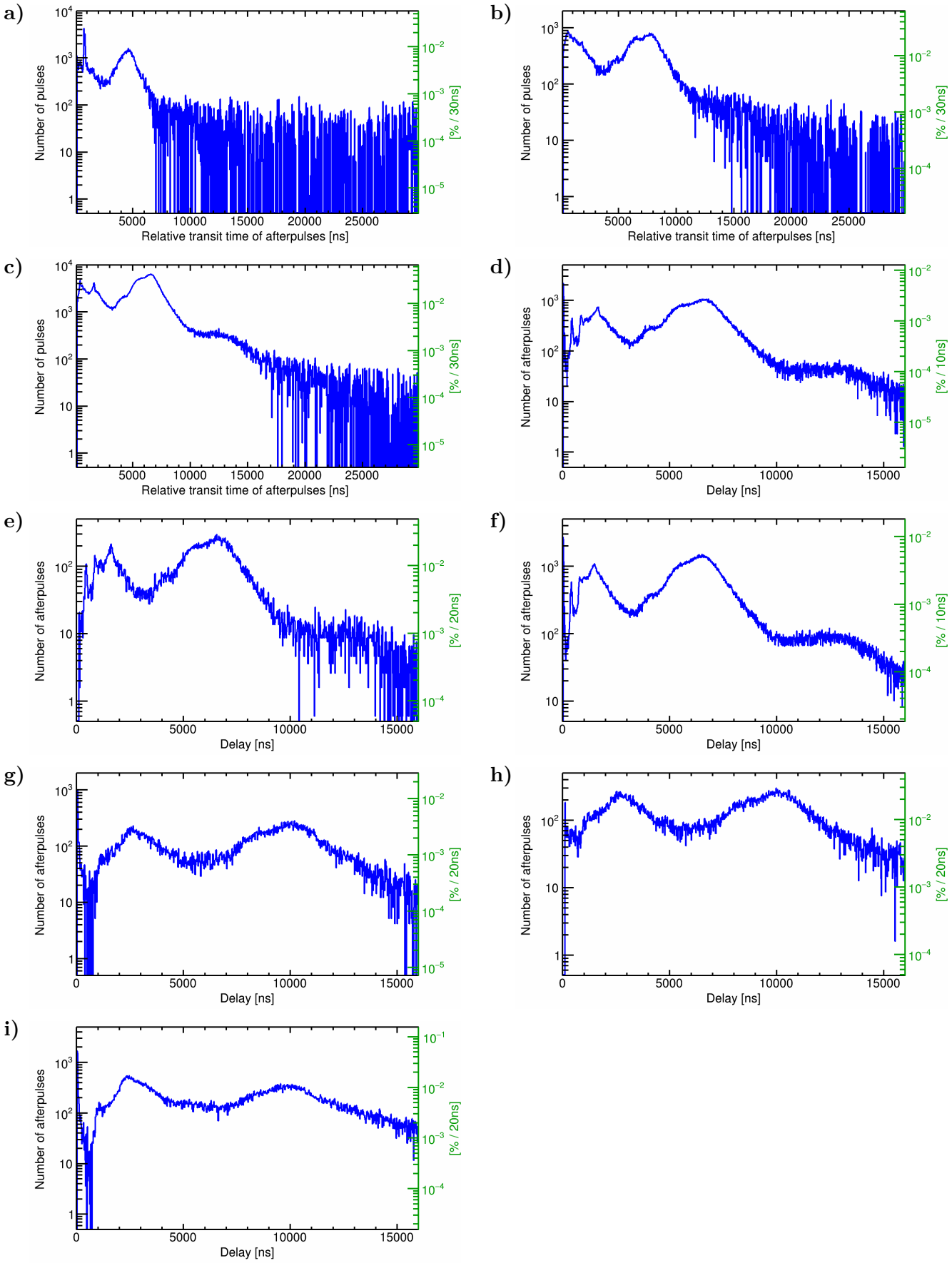


Figure 3.23: iAP results from LNGS and SPAX: Delay distribution after DCR subtraction. AP probability per bin on right axis; note the different binning. Full delay range for LNGS; for SPAX the most relevant section is shown, while the total range is plotted in fig. 3.26. For SPAX classic analyses were used. Discussion see text.
a)–c) LNGS: a) R6594 +1670 V, b) R7081 +1520 V, c) R5912 +1425 V.
d)–i) SPAX: R5912: d) +1425 V AP coincidence, e) simple AP acquisition, f) +1725 V coincidence.
R11780-HQE: g) +1648 V coincidence, h) simple, i) +2092 V coincidence.

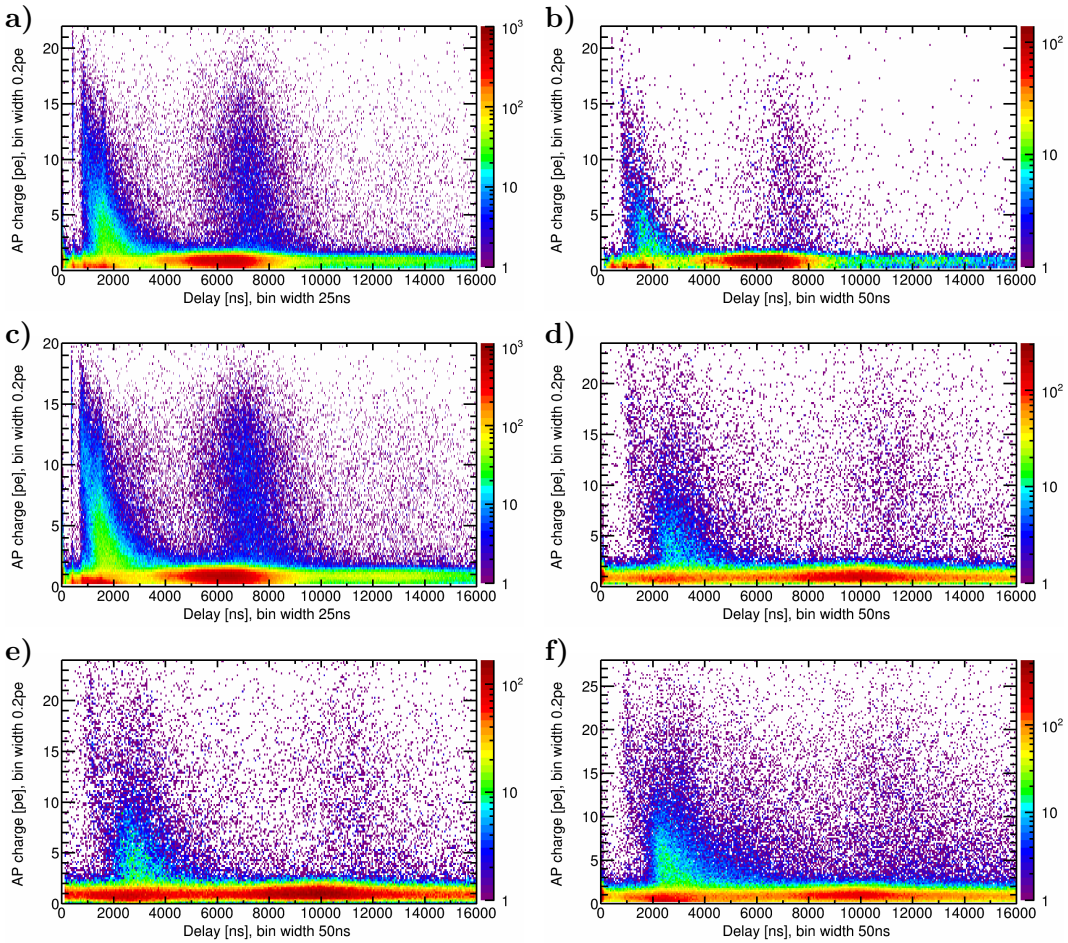


Figure 3.24: iAP results from SPAX: Charge–delay distribution. Same delay range as in fig. 3.23. Charge in pe (using γk charge estimated with $q_{p,\mathrm{PPc}}$). The z-axis denotes the number of pulses per bin; note the different delay bin width. Discussion see text.
R5912: **a)** +1425 V AP coincidence, **b)** simple AP acquisition, **c)** +1725 V coincidence.
R11780-HQE: **d)** +1648 V coincidence, **e)** simple, **f)** +2092 V coincidence.

In the following, thus an identification of iAP effects in the other PMTs is attempted based on the TST calculations already performed for the R5912 at 1425 V by finding a common scaling factor κ to the R5912 for the observed peaks. While U_{k1} and x_1 could be calculated or estimated for the other PMTs, it was found for the R5912 that the calculated values have to be scaled with a fixed factor anyway to arrive at correct delays, as is commonly used in literature. This is likely due to the contribution of the linear field component. Thus, most likely a previously unknown factor has to be introduced in any case and the step of calculating the expected delays was skipped.

For the LNGS measurements, the charge information is not available (no CDD and FDD), which makes it difficult to unambiguously differentiate ixRk [49] and ixRn [50] from the DD alone. However, higher ixRn rates than in SPAX are expected, since the threshold was lower in LNGS acquisitions. Additional SPAX measurements for

[49] Ions created anywhere which hit the cathode and create pulses with regular or large charges.
[50] Ions created between later dynodes which hit a dynode and create pulses which are typically undersized.

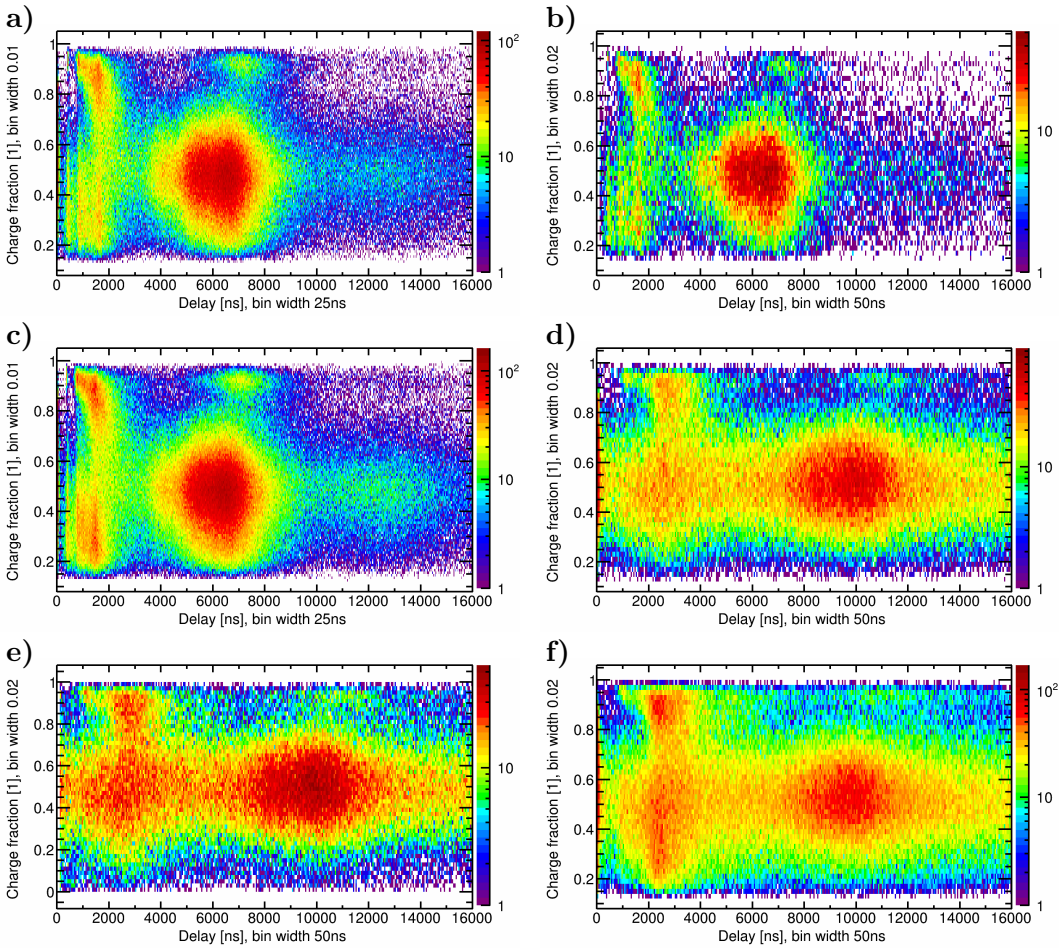


Figure 3.25: iAP results from SPAX: Charge fraction–delay distribution. Same delay range as in fig. 3.23, number of pulses per bin in z-axis (note the different delay binning). Discussion see text. R5912: **a)** +1425 V AP coincidence, **b)** simple AP acquisition, **c)** +1725 V coincidence. R11780-HQE: **d)** +1648 V coincidence, **e)** simple, **f)** +2092 V coincidence.

the R6594 and R7081 would help to confirm the effects which are identified in the following.

Delay distributions

In the LNGS measurements for the R6594, R5912 and R7081 a lower threshold (0.2 pe) could be used than in SPAX (0.33–0.69 pe). From the TST treatment in 2.5.2.3 a higher detected rate of ixRN effects is expected for smaller thresholds. When compared to SPAX measurements of the same or different PMTs, notably more events are observed at small delays in the delay range predicted by TST for ixRn. This confirms the presence of ixRn and the importance of these effects for typical thresholds in the vicinity of 0.25 pe.

R6594

The DD of the R6594 (fig. 3.23a) generally shows a very low iAP rate and the observed peaks are far smaller than for the other measurements.
The dominant features are a sharp early double peak around 755 and 840 ns and a peak at medium delays centered at about 4507 ns. Additionally, a peak around

100 ns, a small distribution surrounding 410 ns, a broad peak around ca. 1500 ns, and a ridge from about 8250 to 10000 ns (better visible for coarser binning) are present. After 15–16 µs delay only residual DCR fluctuations are found.

The delay ratio between the peaks at 755 and 840 ns matches that of He^+2Rk and He^+k/He^+1Rk quite well which results in a scaling factor κ of 0.899 and 0.903/0.894. There might be also a contribution from undersized ixRn. For the peak at 755 ns $O_2{}^+1$ would results in a κ of 0.932, but the presence of free oxygen is unlikely. To the peak at 840 ns Ar^+1 ($\kappa = 0.923$), $CO_2{}^+1$ (0.884), $N_2{}^+2R1/CO^+2R1$ (0.913), and Sb^+2 (0.875) might contribute as well.
The large distribution around 4507 ns resembles that of Hg^+k or Cs^+k for the R5912, which would result in a κ of 0.849 and 0.691. Hg^+k better matches the factor which was found for the helium peaks.
The events at 100 ns stem from higher order photonic EAPs and $H^+2/H_2{}^+2$.
For the 410 ns peak H^+k/H^+1Rk lead to scaling factors of 0.872/0.854; the undersized $CO^+2/N_2{}^+2$ would have 0.854.
The broad distribution around 1500 ns likely is caused by $N^+k/N^+1Rk/N^+2Rk$ with $\kappa = 0.867/0.862/0.980$.
The ridge around about 9125 ns might be from Cs_3Sb^+k as suspected for the R5912 ($\kappa = 0.868$) or from second order iAPs of the peak around 4507 ns. Since the Cs^+ peak is absent or very weak this points to second order iAPs.

All in all there are many remaining ambiguities due to the missing charge information, but each feature has possible explanations. The obtained scaling factors are similar and lie around 0.87. This matches the observation that the delays are smaller for the R6594 than the R5912 (compare figs. 3.23a and 3.23c). This is due to both a smaller x_{k1} and a higher $U_{k1} = 719\,V$ than for the R5912 ($U_{k1} = 514\,V$).
When comparing the relative strength of effects (using the Hg^+k peak) to the SPAX measurement of the R5912 at 1425 V, Cs^+k appears to be absent or far weaker on the left flank of Hg^+k. The helium peaks are very distinct and stronger, while the N^+ distribution is a bit weaker. It might be that the $Cs^+2Rk/Hg^{2+}2Rk$ peaks are present with reduced strength in the valley between helium and Hg.
For this PMT a preference for ik and i1Rk over i2Rk is found for all ion masses except for the lightest (He^+2Rk).

R7081

The shape of the DD of the R7081 (fig. 3.23b) is very similar to the R5912 except that the position of the peaks is scaled up. The total iAP rate is also comparable.
The dominant features are two strong peaks at 7899 and 6269 ns and a large peak around 610 ns which is followed by a broad falling flank with sub-peaks at about 1150 and 1800 ns. Further features comprise a peak at 4645ns and a broad peak around ca. 12250 ns. After this last peak the rate shows a steady decrease and afterpulses are found up to delays of about 25 µs.

The peaks at about 7899 and 6269 ns (fitted with two Gaussians) mimic the structures in the DD of the R5912 and can be identified with the greatest reliability. The 7899 ns peak stems from Hg^+k and the one at 6269 ns from Cs^+k with scaling factors of 1.210 and 1.181. The peak at 610 ns would match H^+k/H^+1Rk, which were present in the R5912, for a larger scaling factor of 1.298/1.271. In addition, the

undersized effects $CO^+2/N_2{}^+2$ ($\kappa = 1.271$) or possibly $O_2{}^+2$ (1.196) could contribute – but O_2 is unlikely to exist in free form.
The pulses around 1150 ns likely are from He^+k/He^+1Rk ($\kappa = 1.237/1.223$); with possible contributions from the underamplified $CO_2{}^+1$ (1.211), Ar^+1 (1.264), $CO^+/N_2{}^+2R1$ (1.250) or Sb^+2 (1.198).
The cluster around 1800 ns most likely originates from N^+2Rk ($\kappa = 1.200$) and maybe also Sb^+1 (1.154).
After delay of 1800 ns, ik, i1Rk and i2Rk populations from N^+, $N_2{}^+$ and Ar^+, possibly also $CH_4{}^+$, H_2O^+ and $CO_2{}^+$, should contribute to the falling flank up to about 3.6 µs.
As for the R5912, on the left flank of Cs^+k a small peak around 4645 ns is observed, which is difficult to identify. Cs^+2Rk would results in a very low κ of 0.997, and Cs_3Sb^+2R1 would provide a better match (1.197) if this compound is produced. Alternatively, the peak might be explained by double ionization through $Hg^{2+}2Rk$ (1.147) or $Cs^{2+}k$ (1.235).
Finally, the broad 12250 ns cluster most likely results from a combination of second-order Cs^+k and Hg^+k following Cs^+k or Hg^+k, with a possible addition from Cs_3Sb^+k/Cs_3Sb^+1Rk ($\kappa = 1.166/1.167$).

Again, many ambiguities in the effect identification cannot be resolved without the charge information, but all features can be explained. The scaling factor is found to decrease with rising ion mass and lies between about 1.2 and 1.3. The delays are thus larger than for the R5912. For the R7081 the same VD is used, but in the analyzed measurement $U_{k1} = 549\,V$ is a bit higher. Since in (2.214) the voltage enters only with its root, the greater x_{k1} dominates and larger delays than in the R5912 result.
Since the R7081 was measured with a smaller threshold than the R5912, here also higher rates of ixRn effects are expected. Comparing the relative strength to the Hg^+k peak with the R5912 at $g \approx 10^7$ measured with SPAX, the whole ixRn region at short delays is stronger. Furthermore, the H^+ peak contains significantly more events – possibly due to the larger amount of underlying ixRn – and the Cs^+ peak is slightly higher.
Effects from v_k and v_1 (ik and i1Rk) seem to be preferred over v_2 (i2Rk) also for light ions in this PMT, which indicates that the transport factors from inside the dynode chain to the cathode are smaller than for the R5912.

R5912

For a detailed discussion of the R5912 DD please consult p. 392.
When comparing the DDs of the R5912 at 1425 V measured with SPAX (figs. 3.23d and 3.23e) and at the LNGS (fig. 3.23c), in the latter the region below about 1250 ns is filled with far more events and the H^+ and He^+ peaks are less pronounced compared to their surroundings. This is likely due to the lower threshold for the LNGS measurement, due to which a higher fraction of undersized ixRn are recognized, for which most effects have delays of lower than 1.6 µs.

If the rate of the late peaks is compared and the bin widths are considered, the simple AP acquisition with SPAX agrees with the values from the LNGS measurement within the differences expected from the different threshold. However, the SPAX coincidence measurement is off by a factor af about 5 and the total iAP rate r_a (see table 3.6) is lower by a factor of 3.7 than the LNGS measurement. Except for the

different number of events, the iAP DD, CDD and FDD are completely identical for the simple acquisition and the coincidence measurement as well as above 1 µs for the LNGS and the coincidence measurement. This leads to the conclusion that the AP rate determination for the coincidence with the scaler was compromised. Either a too low number of APs or a too high number of PPs was registered. This might be due to an unstable scaler or logic channel (AND2, cf. fig. 3.3), a too low PP threshold in CFD1 which then also would trigger on noise, or an unstable AP-discriminator (CFD3). Therefore, the distributions which were obtained with the AP coincidence can only be used qualitatively and for a relative assessment of the features in the distributions. The absolute rates are wrong and have to be determined from the simple AP measurements instead, which unfortunately were not recorded for $g \approx 5 \cdot 10^7$.

With higher voltages (SPAX coincidence at 1725 V) the delay of all effects in the DD decreases, while the relative strength of effects appears to remain similar. No statement can be made regarding the change of the overall iAP rate due to the rate determination problem.

R11780-HQE

The DDs of the R11780-HQE at 1648 V (figs. 3.23g and 3.23h) again are similar to the R5912 but show far larger delays.
The most distinct features are the double peaks at 10077 and 7986 ns and a peak around about 2700 ns. Apart from these features, a spike at 100 ns, a peak at ca. 1225 ns, a bump in the right flank of the 10 µs peak at about 13.5 µs and a weak peak around 17.7 µs are present.
After these peaks the iAP rate decreases steadily but extends up to ca. 70 µs (cf. fig. 3.26e) due to higher order iAPs.

The peak at 10077 ns is from Hg^+k with a scaling factor of 1.544, and the one at 7986 ns can be attributed to Cs^+k ($\kappa = 1.505$).
The distribution around 2700 ns likely stems from N^+k/N^+1Rk (1.561/1.552), which was identified for all other PMTs as well and confirms the value of the scaling factor, possibly with contributions from Sb^+2R1 (1.429).
Around 100 ns higher-order photonic EAPs and the earliest ixRn – likely H^+2 or ${H_2}^+2$ – are encountered.
The structure at ca. 1225 ns can be attributed to He^+2Rk (1.476). Contributions from undersized ixRn could be from $CO^+1/{N_2}^+1$ (1.612), H_2O^+2R1 (1.633) and maybe ${O_2}^+1$, whose scaling factor of 1.512 would best match those of the large peaks but which is doubtful to occur.
The fitted peak at about 13.5 µs might consist of Cs_2Sb^+k/Cs_2Sb^+1Rk (1.491/1.493) or Cs_3Sb^+2Rk (1.469), but is more likely from secondary Cs^+k events following Cs^+k, where the mismatch of the delays could result from fit uncertainties.
Lastly, the distribution around 17.7 µs likely is from secondary iAPs from Hg^+k which follow Cs^+k or vice versa.

All in all, the effects show a similar scaling factor of around 1.5, which shows that they have delay ratios which are comparable to the R5912 and match theoretical expectations. The delay is far larger than for the R5912 due to the larger x_{k1}. $U_{k1} = 595$ V is higher than in the R5912, which compensates this increase slightly.

Cs^+ shows a similar relative strength to the Hg^+ peak as in the R5912, N^+ lies a bit higher, and He^+ is similar. The H^+ peak is less pronounced and no clear line is visible in the CDD.
As for the R5912, a preference of i2Rk for light ions, and of ik for heavy ions is found.
The simple AP acquisition (fig. 3.23h) shows the same structures with comparable relative sizes.
With rising HV (2092 V, fig. 3.23i) the delays decrease and the effects around 5 µs become stronger. These likely stem from secondary iAPs following the previous APs, since this increase is not observed in the DD_1 (not shown). The absolute rate could not be determined due to the abovementioned problem.

Charge–delay and charge fraction–delay distributions

The iAP CDD and FDD are shown in figs. 3.24 and 3.25. Both distributions were already discussed at length in 2.5.2.3 for the R5912, which needs not be repeated here.

When comparing the CDD and FDD for simple and coincidence measurements of the R5912, the same features are observed at matching delay and charge positions and with identical relative strengths, but with lower statistics for the simple acquisition.
With growing voltage the delays decrease and the charge centers of effects which have a charge of over 1 pe rise. This also becomes visible in form of rowing charge fractions.

For the R11780-HQE the high DCR covers the effects at regular and small charges, which therefore are not well discernible. Essentially the same structures as for the R5912 are visible.
In the simple and coincidence measurements identical features, delays, charges and relative strengths are found as well.
For rising HV the same behavior as for the R5912 is observed.

Properties

The lower delay limit was set around 100 ns for all measurements to prevent that LPs of 2-pe pulses are counted as afterpulses. The upper delay limit was close to 30 µs for the LNGS. For SPAX a higher limit of around 95 µs for the coincidence and close to 100 µs for simple iAP measurements was used to allow to search for LAPs.
For SPAX measurements the DCR which is obtained from the iAP DD is lower than that from timing measurements which were carried out after similar times in the dark (cf. table 3.3). This is due to the higher threshold in the iAP measurements.
The resulting rate and probability values from SPAX coincidence measurements cannot be used due to the compromised rate determination (see above) and are put in parentheses in table 3.6. The mean charge and height on the other hand are usable, since the charge and height distribution of pulses is not affected by the total rate.

The iAP rate r_a at $g \approx 10^7$ is lowest for the R6594, followed by the R5912 and R7081, and highest by far for the R11780-HQE – as expected from an HQE PMT (see p. 65). There are considerable differences between the PMT series, and values range from below 1% to over 12%. For the R5912 the simple SPAX measurement produces a similar value as that from the LNGS DD; the difference of 0.9% likely is due to the higher number of undersized ixRn events in the LNGS measurement.

The charge rate $r_{a,q}$ is notably higher than r_a which is because the mean iAP charges are larger than 1 pe (compare mean iAP charge values in table 3.6). For the R11780-HQE, $r_{a,q}$ is especially large (37%), which leads to a significant pollution with iAP charge of the first ca. 20 µs after a PP. This has to be considered in a detector. Another conclusion is that a large fraction of the dark current is from iAPs, which also should affect aging (see 2.7.3).
The nq-rate $r_{a,nq}$ is similar to r_a, while $r_{a,qn}$ resembles $r_{a,q}$.
If the clean pe charge is used instead of the mean charge, for $r_{a,q}$, $r_{a,nq}$ and $r_{a,qn}$ the same changes as for the EAPs are observed (see discussion there).

The iAP probability p_a was calculated from r_a for LNGS runs, for SPAX it was obtained from the DD_{1u} [51] after DCR subtraction. Therefore, for LNGS data p_a is close to r_a since the values are far lower than 100%. For SPAX, however, notable differences are found, which shows that a substantial amount of correlated pulses are present in the DD: the higher-order iAPs.

The mean iAP charge lies significantly above 1 pe for all measurements. This shows that the oversized ixRk outweigh the undersized ixRn. It would be interesting to see how this value changes for lower thresholds, since significantly higher ixRn contributions are expected.
For the R11780-HQE measurements the mean charge is far higher than for the R5912, which probably is due to the higher pe-thresholds which result from the flat pulse shape of the R11780.
A slight decrease of the iAP mean charge with rising voltage is found when comparing the values from SPAX coincidence measurements. This could be because the ion-induced secondary electron yield on the cathode levels out (like $\delta 1$), while the rising $\delta 2$ produces an increasing number of undersized ixRn from v_2.
The difference between the values from the coincidence and the simple acquisition might be due to the varying threshold. This affects the DCR subtraction, which uses the mean PP charge as estimate for the dark noise pulse charge, and thus influences the mean iAP charge.
The mean height shows similar values and behavior as the mean charge, but is systematically lower for the R11780-HQE. One possible explanation might be that for ixRk pulses the t_{k1} variance due to the emission angle of the sk electrons is large enough to reduce the pulse height. This effect should be strongest for ixRk pulses with medium charges; for 1 pe no variation occurs, since only 1 sk is produced, while for high charges many sk electrons are created, which causes statistical averaging.

Breakdown

When the charge rates from EAPs (fitted) and iAPs (classic) are summed up, a value of 59.2% pe/pe results for the R5912 at 1425 V. For the R11780-HQE an even higher figure of 79.2% pe/pe is found.
This means that for each pulse – be it in response to an incident photon or a dark count – about an additional 60 or 80% of the mean pulse charge is produced in form of APs (including higher order APs). The latter number is already quite close to a value of 1 which would lead to breakdown: In this case an initial pulse creates a

[51] Unshadowed delay distribution of the first APs

self-sustaining chain of cascades, which continues until it is quenched (not foreseen in normal PMT operation), the sensor is turned off, or the device is destroyed. Since a PMT produces dark count pulses, this would be the case as soon as the sensor is turned on. When the voltage is increased further – above the range of operation – this is likely exactly what happens, as indicated by the strong rise of the iAP rate with voltage in fig. 2.28. This is confirmed by [170; pp. 264, 438] [173, 180, 332], which state that breakdown occurs due to light feedback (i.e. photonic afterpulses) [52]. In addition, [171; p. 9] notes that operation in the breakdown region permanently damages the device. The allowed HV range which is listed by manufacturers thus probably is the region where the device is not in breakdown yet.

As a consequence, the high total AP charge rate of the R11780-HQE is an additional factor which contributes to the exceedingly large DCR of this PMT. Likely, HQE PMTs in general will show heightened total AP charge rates due to their increased iAP rates and possibly larger photonic afterpulse rates.

A further reduction of the EAP and iAP rate should thus be a worthwhile goal for manufacturers – not only because of the lower AP rates themselves, but also since it offers the prospect of a DCR reduction, higher reachable gains (and thus better TTS, etc.) and an increased sensor lifetime.

Late afterpulses

The SPAX iAP measurements were recorded with longer delay window than at the LNGS to enable the search for LAP presence (see 2.5.2.4). For the coincidence measurements the maximum gate width of 100 µs (Gate2) limited the recordable delay window to about 95 µs (see table 3.6). In order to determine the AP charge, height and pulse shape parameters without significant errors from undersampling, a minimum sampling rate of 1 GS/s was required. Due to the available size of the FADC memory, therefore also the simple iAP acquisitions were limited to a delay range of 100 µs.

The delay distributions after DCR subtraction are shown for the full recorded delay range in fig. 3.26.
Even after subtracting the mean bin content from the DCR, its statistic fluctuations remain and cover the region for large delays. This can only be reduced through a large binning, as was done here. For larger bin widths than used in the plots, the higher-order iAPs around 20 µs would not be discernible anymore and no additional features appear, except for a slowly decaying amount of higher-order iAPs.
The only measurements in which notable pulse populations were observed above 30 µs are the coincidence runs of the R5912 at 1725 V, where a cluster between 30 and 40 µs is found, and the R11780-HQE at 1648 V, where a decaying chain of higher-order iAPs is observed up to about 70 µs. No clusters were observed at delay scales of 70 µs or more as reported in [154, 321].

For the R5912 the cluster around 35 µs would still be in the delay scale reachable with third-order Cs_3Sb^+k or the hypothetical iLk of the same ion. If the first case

[52] See also pp. 134, 350, 379, 408 and 418.

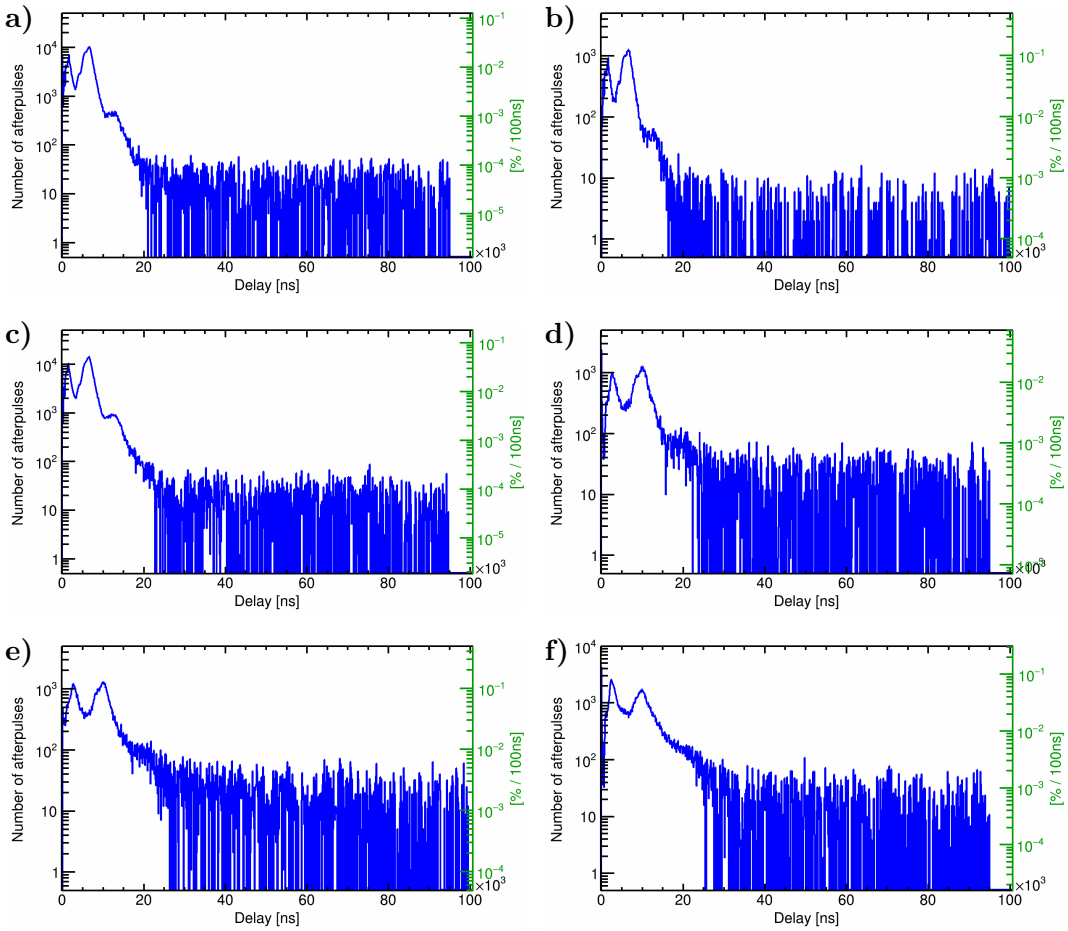


Figure 3.26: LAP results from SPAX: Delay distribution after DCR subtraction for full delay range. AP probability per bin on right axis. No LAP features are visible for the shown or coarser binnings above 70 µs delay, but in some measurements higher order iAPs are present beyond the range shown in fig. 3.23 and at 1725 V the R5912 shows an excess around 35 µs. Discussion see text. R5912: **a)** +1425 V AP coincidence, **b)** simple AP acquisition, **c)** +1725 V coincidence. R11780-HQE: **d)** +1648 V coincidence, **e)** simple, **f)** +2092 V coincidence.

is assumed and the events around 24 µs are interpreted as the corresponding second-order population, this predicts a lower number of events for the third-order distribution than observed, however. iLk, in turn, could not be verified for any ion and should have particularly low probabilities for the backscattering of a compound. The only other possible LAP effect listed in 2.5.2.4 which can produce a cluster (and not an exponential decay) at such high delays is the excitation of metastable levels in the cathode. This is the process which was concluded to be the most likely explanation for the LAPs observed in literature at delays above 70 µs. The time scale might be shorter in the R5912.

The fact that no clusters were observed at large delays might be due to the rareness of these events and that spe illumination was used here. In order to verify that the cluster at 35 µs is not a statistical effect and that indeed no LAPs are present above 70 µs for the studied PMTs, the measurements should be repeated with high occupancy. The region from 100 to 200 µs should also be studied by implementing a digital delay in the setup.

3.3 Selection of optimal candidate

In order to find the PMT series which are most suitable for large LSNDs, an extensive literature research was performed, which was summarized in 3.1. To obtain missing values for promising PMT series, the measurement campaign detailed in 3.2 was carried out. Now, in the present section, the best suited series can be selected based on available information.

Summary of series properties

Table 3.7 summarizes all relevant data by incorporating the results of the measurements, which are listed in tables 3.2–3.6, into table 3.1. The values added by the present work are highlighted through bold text.
For comparability with literature, the acquisitions with $g \approx 10^7$ were selected for the summary table and the values from classic analyses were used unless otherwise stated.
$T_{\mathrm{FWHM},s}$ was used for the TTS, $p_{\mathrm{EP},3\sigma}$ as EP probability. The definition which is used for the LP probability varies in literature[53], so two values are listed: $p_{\mathrm{LP,p}}$ and also $p_{\mathrm{LP},3\sigma}$ in parentheses. For the charge resolution the most realistic variant $R_{F,m,\mathrm{PP}}$ was taken. The peak-to-valley ratio was determined from PPs. $\overline{h}_{\mathrm{PP}}$ was selected as pulse height, for the gain the definition which is based on $\overline{q}_{\mathrm{PP}}$ was used. Rise time, fall time and pulse width are based on the PP values *including* pileup, since these are the numbers which result without pulse-fitting, as in literature. The DCR was obtained from fitting the TTD. For the EAP rate, r_a of the classic evaluation (as literature values) is listed. Its far higher counterpart from pulse-fitting is given in parentheses. As iAP rate also r_a is used.
For the R5912, both values from SPAX and the LNGS were available and in this case for each property the number which is closest to realistic conditions was selected: The LNGS value was used for the TTS (threshold 0.2 pe closer to operating conditions), EP probability (smaller setup jitter and higher threshold), LP probability (no setup fluorescence, higher threshold), P/V (higher threshold), gain (higher threshold) and iAP rate (lower threshold of 0.2 pe than in SPAX iAP acquisitions). SPAX data was taken for the charge resolution (unknown charge scaling in LNGS data) and DCR (longer time in the dark).

Comparison with literature

The values which were measured here can now be compared to literature.

The TTS of the R6594 lies above the value provided by the manufacturer, all other measured PMTs showed lower numbers than manufacturer data (comparing the R11780-HQE with the SQE variant). For the R5912 the measured TTS value is significantly lower than that from independent measurements [335], but the available data is quite old (2002). It could be that the PMT design was adapted in the meantime or a different VD layout was used. Fluctuations between samples could also explain the difference. The result for the R7081 is in agreement with the published value [160] when considering the variance. The measured R11780-HQE

[53] Pulses with lateness above $3\,\sigma_T$ or $5\,\sigma_T$, or pulses after peak; many sources do not quote the applied definition.

Table 3.7: Properties of currently available newer PMT series with large diameters, SQE/HQE cathodes, borosilicate glass and favorable properties. Updated version of table 3.1 (literature research) complemented by values measured in this work. Series starting with R are from Hamamatsu, all others from ETE. Two excluded series are listed for comparison: The R6091 is too small but allows to study the dependence of properties on diameter, and the 9351 is out of production but well studied. All PMTs have hemispherical windows except the R6091 (plano-concave window), and box-and-line (B+L) or linear-focused (LF) dynodes.
Values listed if known, results from this work bold, numbers from manufacturers italicized, figures in red if exceeding requirements (see table 1.1; DCR red if more than 3 times required value). Number of tested samples listed for most significant sources from independent groups; PMTs appearing in several publications of the same collaborations are listed only once. Magnetic field tolerance entries marked with * are detailed on p. 452. Table continued on next page.
Sources: R6091 [189, 213], R6594 [213], 9351 [80, 170, 275, 336, 451], 9354 [201, 263, 309, 342], R5912 [212, 213, 323, 335, 389], R5912 HQE [212, 213, 276, 342, 428, 451], R7081 [160,161,212,213,261,292,301,342,431,452], R7081 HQE [212,213,261,342,452,453], D784 [186,262,306,454], R11780 [190,261,306,439], R11780 HQE [190,261,306,455], R12860 HQE [41, 165, 178, 213, 225, 236, 251, 279, 307, 326, 346, 454, 456–459].

PMT series	**(R6091)**	**R6594**	**(9351)**	**9354**	**R5912**	**R5912 HQE**	**R7081**	**R7081 HQE**	**D784**	**R11780**	**R11780 HQE**	**R12860 HQE**
Number of samples		**1**	2200, 80	4, 1	1625, 324, 1, **1**	255, 110, 1	474, 1, **1**	1, 1	15	7	10, **1**	4620, 145, 1, *260*
Diameter window [cm (inch)]	*7.6* (3)	*12.8* (5)	≈ 20.3 (8)	*20* (8)	*20.2* (8)	*20.2* (8)	*25.3* (10)	*25.3* (10)	*28* (11)	*30.5* (12)	*30.5* (12)	*50.8* (20)
Diam. cathode [cm]	*6.5*	*11*		*19*	*> 19*	*> 19*	*> 22*	*> 22*	*27*	*> 28*	*> 28*	*> 46*
Dynodes	12 LF	10 B+L	12 LF BeCu	12 LF Cs_3Sb	10 B+L	10 B+L	10 B+L Cs_3Sb	10 B+L Cs_3Sb	12 LF Cs_3Sb	10 B+L Cs_3Sb	10 B+L Cs_3Sb	10 B+L
PDE [%] at 420 nm							23.0 ± 1.4					28.1
QE [%] at 420 nm	*26*	*25*	27	23, *26.5*	≈ *25*	37 (408 nm), *35 (390 nm)*	27, *25 (390 nm)*	34.6, *35 (390 nm)*	≈ 26.5	*22*	≈ 32	≈ 28.5
CE [%]							*90*					*95/85* (46/50 cm)
TTS T_{FWHM} [ns]	*1.5*	**1.92**, *1.5*	2.76 ± 0.27	1.58, *2.7*	≈ 2.9, ± 0.3, **2.03**, *2.4*	2.60 ± 0.12, *2.4*	2.8 ± 0.2, **3.05**, *3.4*	2.67, *3.4*	4.66, ± 0.40, *3*	3.23, ± 0.35, *2.6*	3.04, ± 0.33, **2.13**	< 2.65 ± 0.15, *2.4*
Transit time T_0 [ns]	*40*	*45*	70	*65*	**59.9**, *54*	*54*	*62*	*62*	*62*	*70*	**80.6**	*95*
EP probability [%]		**2.67**	0.75 ± 0.35, $\gamma 1 > 0.012$	0.2–0.4	**1.49**, 0.1–0.5		**0.58**, $\gamma 1 > 0.1$			*> 0.5*	**1.63**	0.05
LP probability [%]		**3.41 (14.3)**	7.90 ± 1.26		**2.94 (18.4)**	2.3 ± 0.05	3, **3.16 (18.0)**		4.51 ± 0.74	4.48 ± 0.32, *4*	4.3 ± 0.35, **3.41 (8.62)**	2–3

PMT series	**(R6091)**	**R6594**	**(9351)**	**9354**	**R5912**	**R5912 HQE**	**R7081**	**R7081 HQE**	**D784**	**R11780**	**R11780 HQE**	**R12860 HQE**
Charge res. R_F [%]			106	82–111	71–111, **86.6**		59	67.4	90 ± 25	89 ± 25	103 ± 39, **87.9**	70.0 ± 7.9
P/V [1]		**3.76**	2.10 ± 0.39	2.4–3.2, *2*	3.1 (2.5–4.3), **2.85**, *2.8*	*2.8*	3.2–5.5, **2.79**, *2.8*	4.1 *2.8*	2.32 $\pm$ 0.67, *2*	2.8 ± 0.28, *2.8*	2.24 $\pm$ 0.27, **2.89**	3.8, *2.5*
Pulse height [mV]					**19.8**						**11.5**	6.5
Gain g (max.) [10^7]	*1 (> 10)*	*0.2 (> 2)*, **1.0**	2	1 (*3*)	1, **1.3** (*4*)	1 (*4*)	1, **1.3** (*4*)	*1 (4)*	1 (*3*)	1	1, **1.0**	1 ($\approx$ *3.5*)
Linear range (dyn. range) [pe]							250–300					370 (> 3000)
Rise time t_r [ns]	*2.7*	*3.5*		3.1–3.4, *3.5*	3.1–3.2, **3.1**, *3.6*	*3.6*	*3.8*	*3.8*	*5*	*4*	**3.5**	6.9, *5.2*
Fall time t_f [ns]				4.1–4.4	3.5–4.6, **4.9**						**5.9**	10.2, *8*
Pulse width t_w [ns]				*4.5*	**5.4**				*6*		**6.9**	11.6, *11*
DCR [kHz]		**4.95**	2.48 ± 3.49	≈ 1, *4*	1.3, **0.77**, *4*	5.80 ± 0.78, *6*	2.2 ± 0.5, **2.46**, *7*	5, *8*	18.4, *8*	3.7 ± 5.1, *10*	4.4 ± 1.9, **6.32**, *15.5*	6.4 ± 1.9, 15.5, *18.6*
EAP rate [%]					3.4 ± 0.6, **2.22 (108.8)**	2.7 ± 0.01					**2.69 (102.1)**	
iAP rate [%]		**0.97**	4.94 ± 2.64	2.1–3.2, *5*	< 2, 2.2–2.5, **6.30**	7.1 ± 1.8	2.68 ± 1.46, **4.93**	16		*10*	**12.1**	5
Pressure capability [MPa]				*0.202*	*0.7*	*0.7*	*0.7*	*0.7*	*0.808*	*0.8*		1.25, *0.9*
Magnetic field tolerance [µT]				*, *40*	*		50		> 45*		32	
Radiopurity K_{nat} [ppm], ^{232}Th [ppb], ^{238}U [ppb]| (Bq/kg)				*60* (4.0), *30* (0.3), *30* (1.0)	(0.6), (0.2), (0.9)				*60*, *30*, *30*			

also shows a markedly smaller TTS than reported in [306] and found for the SQE variant, possibly because this was an exceptionally good specimen.
The main peak transit time T_0 was somewhat larger than quoted by Hamamatsu for the R5912 and the R11780-HQE (compared to the R11780 value). In all likelihood this is due to the different HV required to obtain $g = 10^7$ for each sample (see p. 506).
The EP probability is far higher for the R5912 than in [309], possibly because in this publication the end of the EP recording range was further from the main peak than here.
The LP rate definition $p_{\mathrm{LP},\,3\sigma}$ produces far higher numbers than in the literature, since there other definitions are used which result in numbers closer to $p_{\mathrm{LP,\,p}}$ (R11780-HQE: $5\,\sigma_T$ [190], R7081: unknown [160]). Considering the different integration method, the values are compatible with data.
For the charge resolution, the obtained numbers lie within the range observed by other groups.
The P/V agrees with the reported range for the R5912. It is lower than in literature for the R7081 [160]. This might be due to a different VD, since the value agrees with the number given by Hamamatsu, who fabricated the VD employed here. The P/V result for the R11780-HQE is better than the range obtained by [306], but close enough to be due to sampling variations.
For the pulse height, not many numbers are available in literature. Nevertheless, the value for the R12860 in combination with the data reported here indicates that the height falls with diameter, since the pulse shape becomes broader (compare values for rise time, fall time and pulse width). This is probably due to increasing interdynode transit time variations due to a larger size of d1 and possibly also later dynodes.
The rise time of the R5912 is identical to values from literature [309], but its fall time is slightly above the range listed in the same source. Again, a slightly different VD might be the cause.
The measured DCRs are in agreement with published values, when considering the strong dependence on illumination history of this number.
The classic EAP rate of the R5912 is lower than in [335]. which could be due to a different VD (the classic EAP rate is very sensitive to the pulse length) or an altered PMT design.
Finally, the iAP rate for the R5912 is notably higher than in [309, 389] (both Daya Bay). The studied PMT might be an outlier, since strong iAP rate fluctuations between specimens are common (e.g. [336]). If a higher threshold was used by Daya Bay, this would also explain the discrepancy.

Compliance with requirements

Based on these new results the fulfillment of the requirements can be analyzed for previously missing values of the R6594, R5912, R7081 and R11780-HQE.

For the TTS no large changes to literature values were observed. The R5912 and R11780-HQE performed better than expected, the R7081 and R6594 were slightly worse. Therefore, the R7081 and R11780-HQE remain border cases regarding the TTS requirement.
The measured EP probability significantly exceeded the acceptance threshold for the R6594. Too high numbers were also observed for the R5912 and R11780-HQE. In literature far lower values are reported, likely due to more lenient integration

windows. More samples of these series will have to be studied to confirm this excess. Furthermore, it would be beneficial to study how larger $p_{\mathrm{EP},\,3\sigma}$ values affect the event reconstruction in LENA to assess whether a higher limit would be acceptable. Otherwise, this criterion might exclude three PMTs series from the short list.
The LP rate definition $p_{\mathrm{LP},\,3\sigma}$ produces very high values for all measured specimens due to the long trailing flank of the main peak. Further Monte Carlo (MC) simulations are required to see how problematic such pulses with short lateness are for event reconstruction. For now, $p_{\mathrm{LP},\,\mathrm{p}}$ was used as criterion and in this case all studied PMTs pass.
The P/V requirement is easily fulfilled by all PMTs.
The measured DCR of the R6594 was far higher than the permitted value (278 Hz) for this small PMT even when allowing for a factor of 3 due to an extended cool-down period before operation in a detector. To verify whether this series complies with requirements, the available specimen will need to be tested for longer times in the dark and additional samples will have to be studied. The R5912 and R7081 fulfill the DCR demands. For the R11780-HQE, despite the prolonged cool-down of 17 d the DCR is still higher than 3 times the allowed value. This is of course because of the HQE cathode and could eliminate this candidate. More samples and long term cool-down studies are required for a final decision. Alternatively, the influence of a higher DCR per area on physics goals (CNO and ^{7}Be neutrinos) has to be reassessed.
For the EAP criterion, the classic EAP rate is used, since this is the amount of EAPs visible with a standard event reconstruction without pulse-fitting. In light of the new results regarding PMT theory which are reported here, further studies are needed to appraise the improvements to charge, time and position reconstruction which become possible through pulse-fitting. For these enhancements the fitted EAP rate could play a role and new requirements might have to be introduced. The classic EAP rate requirement is fulfilled by both the R5912 and R11780-HQE. Measurements for the other series will need to be carried out as well.
The iAP rate is very low for the R6594 and close to the acceptance limit for the R7081. For the R5912 the measured value exceeds the limit, but since the number is far higher than in literature, the measured sample could be an outlier. Measurement of further specimens is necessary to evaluate this discrepancy. Hamamatsu lists an iAP rate of 10% for the R11780, and as expected the measured HQE variant shows an even larger value, which significantly exceeds the requirement. However, since the iAP rate strongly fluctuates between samples, further tests are needed before this series can be excluded with certainty.

Selection

Of the parameters which were analyzed in the characterization campaign, the EP probability, DCR, and iAP rate turned out to be the most crucial. While the latter two were already identified as key requirements in 3.1, it was not clear before, that the EP presence is this critical and it might be necessary to loosen this criterion to not exclude series with otherwise good performance.
When comparing the current requirements with the results of the measurements it should be kept in mind that only one sample was measured for each series and that more statistics are required to be able to definitely include or exclude sensors from the candidate list. The results presented here can only serve as a first screening to identify the series with the most potential.

A strict application of the current requirements on the measurement results would exclude the R6594 (DCR, EPs), R5912 (EPs; TTS close to limit) and R11780-HQE (iAPs, EPs; DCR and TTS border cases). Literature data excludes the R7081-HQE due to a too high iAP rate and potentially a too large DCR. This leaves only the 9354, R7081 and R12860 in the short list.
In addition, even the remaining series have not been fully characterized yet or show properties which might be problematic. For the R12860 the DCR value from literature was obtained after long times in the dark [165, 236, 346, 458] and it is questionable whether a further decrease to the required average level of 4.4 kHz is possible. Furthermore, the iAP rate lies exactly at the exclusion limit for this PMT (5%). Moreover, due to the discrepancy between literature and data, for the 9354 and R12860 the EP probability should still be determined. In addition, the EAP rate is not known for 9354, R7081, and R12860 and the LP probability of the 9354 still remains to be determined.
Since only one sample was measured for each series, no series can be excluded with certainty at this point. Therefore, the list of current candidates comprises (in order of ascending diameter): **R6594**, **9354**, **R5912**, **R7081**, **R7081-HQE**, **R11780-HQE** and **R12860**, with a preference for the 9354, R7081 and R12860.

Potentially no single PMT series fulfills all requirements. Two possible solutions present themselves: The PMT properties which were compiled here – ideally after measurement of more samples and parameters – could be implemented in the physics MC to see which series allows to maximize the physics yield. Alternatively, as in JUNO a mixed configuration of two or more series which are specialized on different reconstruction aspects could be used.

For the OM development presented in ch. 5, the R11780-HQE and D784 were taken as baseline photosensors, since this was carried out at an earlier stage of this work. The available information at that time looked most promising for these two types; the R12860 was not yet available then.

Higher gain: benefits and drawbacks

Lastly, the changes of PMT properties with higher gain shall be analyzed to assess whether this might improve detector performance.

First, the advantages of a larger gain (i.e. higher voltage) shall be analyzed. With growing gain the P/V rises slightly and the TTS decreases. Also T_0 goes down, which results in a smaller earliness and lateness for all effects. The EPs and LPs are then possibly less critical for event reconstruction. To this contributes that the EP charge gets smaller, so EPs have less influence on the detector response. The smaller charge also lowers the optimal EP threshold and then even typical thresholds exclude more EPs. The LP probability is likely lower for higher voltages as well. Furthermore, the pulse shape becomes more narrow, which also benefits event reconstruction. In addition, the optimum EAP discrimination threshold falls, so more EAPs can be excluded with standard thresholds.

On the other hand, the linear and dynamic range decrease with voltage, the EAP rate increases, and both the iAP rate and DCR rise significantly.

For some properties it is unclear without measurements whether an improvement or deterioration will result from higher voltages.
The PDE might rise or fall (unless U_{k1} is stabilized). CEk1 rises with increasing voltage, but the behavior of the ME is hard to predict. On the one hand, since the signal-to-noise ratio of spe height/charge vs. baseline noise grows, a smaller pe-threshold can be chosen, which should lower the threshold losses. On the other hand, the backscattering losses on d1 increase slightly with rising voltage. Furthermore, the threshold losses could grow despite a lower threshold, if the falling charge of underamplified pulses (due to the rising δn) pushes them below the threshold. Whether in sum the ME rises or falls is difficult to say, and thus no conclusive statement is possible for the PDE, too.
The charge resolution and EAP charge rate either rise or fall with voltage, depending on the series.

After listing the expected changes for all properties, it can now be concluded whether a higher gain is beneficial or not.
The growing EAP rate might be compensated roughly by the lower optimum EAP threshold. The influence of EPs (which proved to be more critical than originally thought) is reduced by a higher gain. Furthermore, the TTS – which is close to the requirement limit for some series – is lowered. However, the strong increase of the DCR and iAP rate likely *excludes* higher gains and the reduced linearity might also become critical.
The best course of action might be to insert the PMT properties which were measured for different voltages into the detector MC, and analyze if a higher gain can improve the physics yield.

4 Alternative photosensors

Particle and astroparticle physics have always been a fertile breeding ground for photosensor innovations.
Despite a formidable number of new photosensor types which have been developed in the past decades, the classic photomultiplier (PMT) is still the workhorse in large neutrino detectors, however (see 1.2). This is because despite the age of this sensor type (invented in the 1930s) it has many favorable properties [1]. In addition, PMTs have shown a remarkable flexibility in design and a considerable ability to improve core characteristics [2].
The aptitude of current PMT series as photosensors of liquid scintillator neutrino detectors (LSNDs) such as LENA [3] or THEIA was studied in chapter 3 based on measurements and literature research. There it was found that in fact several existing PMT series are viable candidates.

Nevertheless, the question remains if other photosensor types might be better suited for large neutrino detectors.

The past few years have seen particularly many new photosensor developments for astroparticle physics (APP) detectors.
Now for two large future neutrino detectors, novel sensor types are either foreseen in addition to PMTs [4] or are being developed to replace PMTs in a future detector upgrade [5].
Furthermore, the first complete prototypes of the LAPPD (see p. 641) have been produced and are now being tested in ANNIE [471]. The combination of picosecond timing and millimeter position resolution for a large sensor that is capable of single photon counting may prove to become a game changer in the years to come.

Therefore, it is only logical to examine other photosensors than the venerable PMTs for their suitability for large LSNDs as well. This is the subject of the present chapter.

The following content aims to give an overview of the multitude of existing and future sensors which might be usable in LSNDs and water Cherenkov detectors (WCDs).

[1] See p. 45 for a list.
[2] Also detailed on p. 45.
[3] Low Energy Neutrino Astronomy
[4] JUNO (Jiangmen Underground Neutrino Observatory) plans to use a 20” micro-channel plate PMT (MCP-PMT) alongside two PMT series [251, 307].
[5] Hyper-Kamiokande (Hyper-K) develops a 20” hybrid avalanche photodiode together with Hamamatsu (Hamamatsu Photonics K. K.) [165, 279] and aims to improve the timing of the JUNO MCP-PMT [279, 346].

Therefore, the scope of this study was mostly limited to photosensors fulfilling the following criteria:

a) **Sensitive in the visible or near-UV spectrum:** Sensors whose quantum efficiency spectrum allows their use for common scintillators with wavelength-shifters (WLSs) and for the detection of Cherenkov light.

b) **Capable of operation at temperatures close to room temperature**, where the content of an LSND or WCD is still liquid and which are reachable with affordable cooling.

c) **Sufficiently advanced design:** A working prototype exists, or the concept seems feasible with no obvious potential problems which might affect functionality.

d) **Current relevance:** While e.g. the first hybrid photodiodes (HPDs) were created as early as 1965 [472] [6], older devices usually can not stand up to current models. Therefore, the list was limited mostly to developments of the past 30 years.

Only in justified cases exemptions from these criteria were permitted.

First a general overview of the processes involved in photodetection is given in 4.1, which is also used to classify sensor types. This is followed in 4.2–4.4 by a description of the individual sensor types and series relevant here which were found in a thorough literature research. The discussed photosensor types are structured using the classes which were introduced before. For each sensor, the working principle is illustrated, the properties reported in literature are collected with typical and best values (if available), and the advantages and disadvantages for LSNDs are addressed. Sensors which are clearly not suited for LSNDs are only mentioned in passing.
These preparations lead to a comparison with the requirements of an LSND using the example of LENA in 4.5, which concludes the chapter with a selection of the sensors best suited for such a task.

The present chapter is meant as a compendium and review of UV/visible photosensors. It is explicitly intended to aid and induce the development of new photosensor types and the improvement of existing sensors. By compiling all photosensor layouts and types together with their advantages and drawbacks, and through the review and discussion of the individual detection steps, it is hoped that new possibilities for enhancements or the combination of techniques might be revealed to the reader.

4.1 Photosensor classification

4.1.1 Detection steps

An extensive literature research was carried out to identify all previously developed photosensor types. The found sensors were analyzed for common features and differences and grouped to allow a better comparison. To this end, using previous

[6] This publication gives a good overview of the history of HPDs.

photosensor classifications encountered in literature [191, 244, 473–476] [477; ch. 1], a classification was introduced which is based on the particle interactions involved in the detection process.
The complete process of photodetection inside the photosensor can be generalized to:

$$\begin{aligned}\text{Photon transport} &\rightarrow \text{Conversion} \rightarrow \text{Information carrier transport} \\ &(\rightarrow \text{Amplification} \rightarrow \text{Information carrier transport}) \\ &\rightarrow \text{Readout}\end{aligned}$$

Not all sensors use amplification and each step in the diagram may consist of multiple sub-steps which are connected by transport processes.
This itemization of the detection is similar to the processes leading to photoemission (absorption, diffusion and emission; see p. 48) and the fundamental principles of time shift theory (TST; chains of particle interactions on device stages which are connected by transport processes and can lead to amplification, see 2.3.5).

It was found that a selection of these detection steps together with a specification of the employed materials is sufficient for a clear and unambiguous classification of photosensors which consists of:

- **Detection mechanism:** The type of interaction between photon and sensor.
- **Conversion medium:** The material where this interaction takes place.
- **Transport medium:** The medium in which the created (quasi-)particles are transported to the readout or amplification stages.
- **Amplification:** If applicable, the amplification mechanism(s) which are employed in the sensor.
- **Readout:** The type of readout process.

The readout and amplification processes are intrinsically connected to the transport. For example, electrons which are created in the sensor need to traverse a potential difference to arrive at a readout anode or to gain sufficient energy for the production of secondary electrons (SEs) or electron–hole pairs.

The resulting classification of photosensor types which is based on these operation principles is shown in fig. 4.1 for the sensors relevant here ("photosensor tree"). To the author's knowledge no other photosensor classification of comparable completeness has been compiled.
This figure summarizes the largest part of the following chapter, for which reason it is of advantage to introduce it now. The parts and abbreviations which are not clear at the moment will be explained and discussed step by step in the following. For the present, it is sufficient to focus on the general layout of the photosensor operation principles – represented by the branches in the figure (colored text with group labels on the very left) – and not the individual sensor types (black text). While many sensor types differ only in the readout process, this was not included as a fifth distinguishing mark in the diagram, because it would add another layer of differentiation and reduce the readability. Furthermore, for many types readout variants exists, which often could not be listed in the schematic due to lack of space. However, these varieties are included in the depicted position resolution capability of the main type and are discussed in the text.

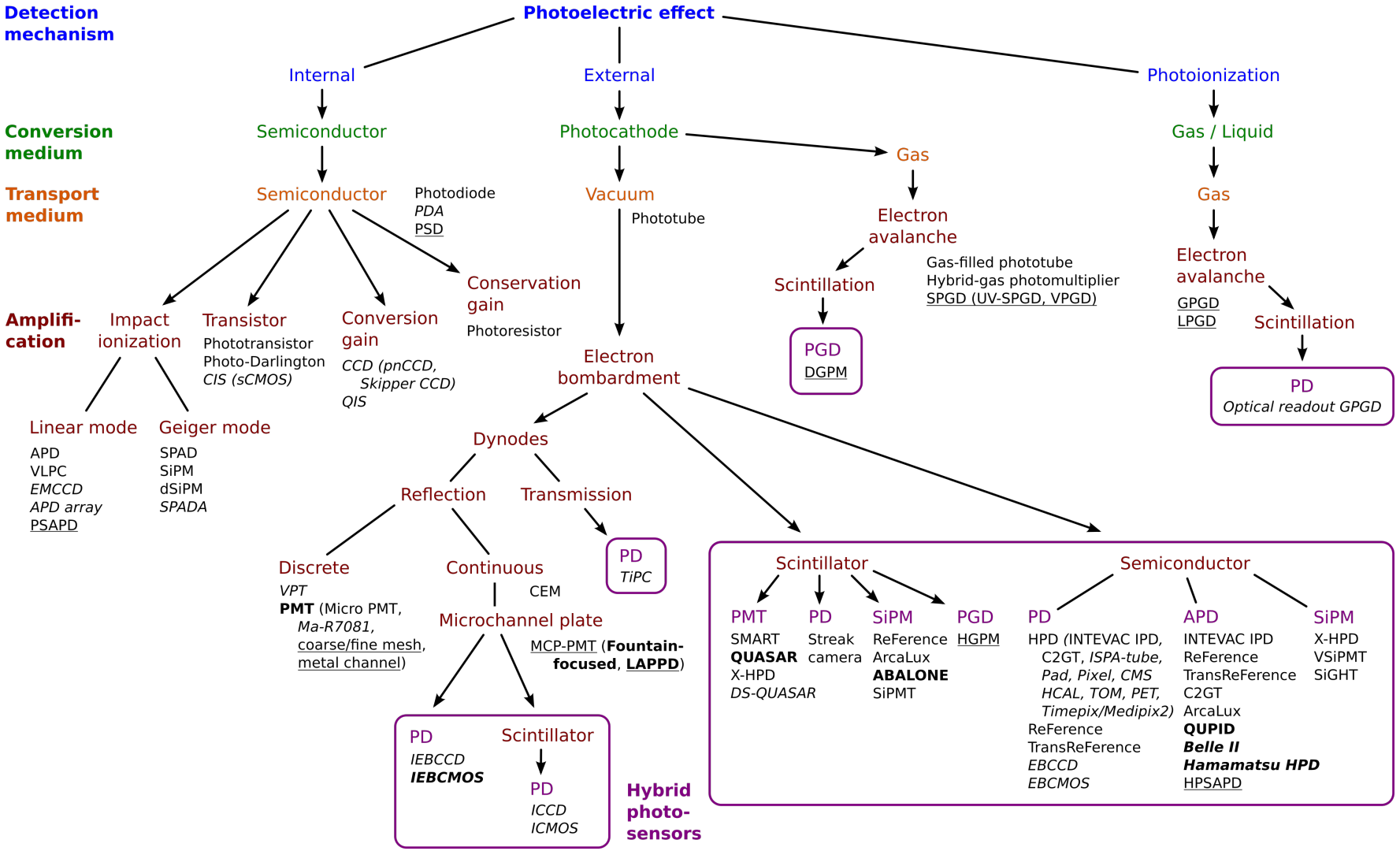


Figure 4.1: Photosensor classification: Photosensor types capable of detecting visible/near-UV light at room temperature which are commercially available, have produced prototypes, or are under development (list likely not exhaustive). Arrows illustrate the progression of the detection process, lines denote subtypes. Sensor names in black, names italic/underlined if image/position sensor models exist (see p. 600). Hybrid sensors (see 4.1.2.1) in lilac boxes, secondary photosensors in lilac letters. Sensor subtypes in parentheses. Sensors ordered by position sensitivity (point sensors, image sensors, and then position sensors), then chronologically by start of development. Only the sensors highlighted in bold font are potentially relevant for liquid scintillator neutrino detectors (LSNDs) at the moment (short list, see 4.5.1). The detection steps, sensor classes, individual sensors, and abbreviations are explained in the text.

The present section attempts to give a *general* overview of the operation principles employed in photosensors so far, and is not bound to the requirements listed above. In fig. 4.1 and the analysis of individual sensors, sensor classes which do not fulfill the prerequisites and thus are not relevant for LSNDs, however, were not included for reasons of clarity and brevity.
In the following, the different forms of the individual detection steps and possible combinations will be discussed.

4.1.1.1 Detection mechanisms

Photoeffect

To the best of the author's knowledge, *all* photosensors for the UV, visible or IR spectrum which were designed so far are based on the photoelectric effect (short photoeffect) in one form or another. There is a single exception that will be discussed subsequently. At the photon energies within the mentioned spectrum (eV scale) Compton scattering is irrelevant and pair production is impossible [478; p. 9].
There are three different subtypes of the photoeffect for the energies relevant here [7] (see also p. 48), where in all cases the photon transfers its complete energy to an electron:

- In a solid, this so-called photoelectron (pe) is excited to a conduction band (CB) and leaves behind a hole (see fig. 2.3). This process is called **internal photoelectric effect** (IPE) [478; pp. 9–10].
- If the created photoelectron has sufficient energy to diffuse to the surface and overcome the surface barrier (also fig. 2.3) it is emitted from the solid, which is then labeled **external photoelectric effect** (EPE) [478; pp. 9–10]. The EPE requires the IPE as an initial process as well as the two additional steps of diffusion and emission.
- In the case of **photoionization** (PI), the photon is absorbed by an electron of an atom or molecule in a gas or liquid, which leads to ionization [196; pp. 1–2]. Thus, PI can be considered the atomic/molecular photoeffect. In a gas the photoelectron is emitted directly into the surrounding medium *without* additional steps. In a liquid, similar to the EPE, the transport and emission processes are important as well.

The EPE emits free electrons into a vacuum, for the IPE the electrons remain in the solid, and for the PI in gases or liquids the electrons are emitted into a gas phase. This affects which amplification and readout methods can be used.

Internal photoeffect

Sensors based on the internal photoeffect can be distinguished by how they use the photoelectrons which are created inside the material to produce a measurable parameter and in this way enable the detection of incident light:

- **Photoconductivity:** For a photodiode (PD) in reverse bias, the created mobile charge carriers in the CB or valence band (VB) increase the conductivity of the material [478; p. 10].

[7] Photodisintegration, the nuclear photoeffect [479], is only possible for high-energy gamma rays.

For intrinsic photoconductivity the electron is excited across the band gap, which creates both a mobile hole and electron [400; pp. 744–745].
Extrinsic photoconductivity occurs for photon energies smaller than the band gap and can only excite electrons from or to impurity levels lying within the band gap [400; pp. 744–745]. For this reason it creates only one type of mobile charge carrier, electron or hole.

- **Photovoltaic effect:** When a photodiode which is operated with zero bias is illuminated with light, the created charge carriers produce a voltage or current [480; p. 467] [477; ch. 15, p. 6]. This principle is used for solar cells. PDs operating in the photovoltaic region produce no dark current, which leads to a much better sensitivity at low light levels (LLLs) than in the photoconductive region, but the time response is worse due to the higher junction capacitance [480; p. 467].

- **Photon-drag effect:** The momentum transfer from a photon to the photoelectron can cause an electron movement, which for a sufficiently high flux leads to a measurable current or voltage [481]. This is similar to the EPE, which uses the kinetic energy gained through the photoeffect for pe emission, while for the photon-drag effect the electron remains within the material. The effect also resembles photoconductivity, where the movement is caused by an electric field. This technique is used only in the IR for large photon fluxes and is not applicable here.

- **Photorefractive effect:** This can occur in a material which exhibits both photoconductivity and the electro-optic effect [8]. The constructive interference patterns of two coherent light beams excite electrons from (immobile) impurity levels in the band gap to the CB, where they can diffuse to the regions of destructive interference (no excitation) and recombine to impurity levels [482]. The redistribution of these electrons causes electric fields, which locally change the refractive index of the crystal through the Pockels effect. This forms a refractive index grating, which imprints the interference pattern onto the crystal and when illuminated recreates the original pattern, thus allowing reversible holographic data storage [483]. Because this method is based on non-linear optics and permits no direct readout, it is not suited for LLL detection in LSNDs.

- **Thermal:** When photon absorption is followed by non-radiative relaxations or occurs for photon energies smaller than the band gap, it produces phonons and thus heats up the material, which can be measured through various means [480; pp. 463, 481–501]. With the exception of cryogenic sensors, this detection mechanism is only capable of measuring high photon fluxes in the IR spectrum, however. Thus, thermal photosensors play no role for the detectors studied here.

  - **Bolometer:** The radiant *power* which is deposited in the material is measured with a temperature-dependent resistance [480; pp. 491–494]. This sensor type is used from mid-IR to microwave wavelengths.

  - **Calorimeter:** For very low temperatures, the heat capacity of a material can be extremely low, in which case the *energy* of a single incident photon can be sufficient to cause a notable change in temperature [9].

[8] Pockels effect: a change of the refractive index proportional to the electric field. Since the refractive index is a consequence of the polarization of the crystal, the photorefractive effect is based on polarization (see p. 570).
[9] Calorimeters measure the deposited energy, while bolometers measure power.

Calorimeter photosensors thus are either cryogenic detectors or can only detect large photon fluxes, so this class is unsuited for LLL detection in an LSND.
In these photosensor types often a superconducting phase transition is employed in combination with nanoscale structures (at least in one dimension). In this case the deposited energy of a single photon is sufficient to either trigger a local phase transition or alter the measured properties to a significant extent. Because one photon affects multiple particles contributing to the measured quantity, this boosts the sensitivity and can allow to detect single photons. This can thus be regarded as a type of amplification.

  - ▷ The temperature variation can be measured by the change of the *resistance* of a temperature-dependent **resistor** [484].
  - ▷ To boost the sensitivity of such a resistance measurement, a **transition-edge sensor** (TES) uses the sharp transition between superconducting state and normal resistivity by operating at a temperature within the transition [485; p. 18].
  - ▷ Alternatively, a **superconducting tunnel junction** (STJ) measures the tunneling *current* of quasiparticles across a Josephson junction, which result from the breaking of Cooper pairs through photons [486; pp. 479–496] [153; pp. 725–726].
  - ▷ The **superconducting nanowire single-photon detector** (SNSPD), in turn, employs thin superconducting wires through which a current close to the critical current flows. [485; pp. 18–21] [478; pp. 41, 42] Incident photons break up Cooper pairs, which lowers the critical current below the applied one, creates a non-superconducting region, and produces a short *voltage* pulse.
  - ▷ A **kinetic inductance detector** (KID), on the other hand, measures the *inductance* change of a superconducting strip placed in an alternating electric field which is caused by the breaking of Cooper pairs through photons [487].
  - ▷ **Superconducting granules** employ the property of type-I superconductors to completely expel a magnetic field from the interior of small superconducting grains. [153; pp. 723–725] These are operated slightly below the transition temperature so that the energy of an incident photon is sufficient to heat the granule above the critical temperature. The resulting change in *magnetic flux* can be measured.
  - ▷ For **magnetic calorimeters**, the absorbed heat causes spin flips of paramagnetic atoms, which changes the *magnetic field* [488].

– **Pyroelectric detectors:** In a pyroelectric material, the heat which is generated by photons gives rise to a voltage through the pyroelectric effect. [489; pp. 558–564] These sensors are used in the IR only, draw their power from the signal alone (passive detector), and can merely detect variations of the incident radiation.

  - **Thermopiles:** A thermopile, on the other hand, uses the *thermoelectric* effect (Seebeck effect) which produces a voltage when different metals are connected and exposed to a temperature difference [480; pp. 483–487]. These also are IR devices.

  - **Pressure:** In a **Golay cell** photons are absorbed by a foil, which heats up a surrounding gas [480; pp. 482–483]. The rising pressure moves a membrane off which light is reflected onto a photodiode. The measured intensity is therefore connected to the elongation of the membrane and indirectly to the incident light flux. The Golay cell is thus an opto-acoustic detector. Its sensitivity can be improved by using laser light and interferometry, but the sensor is sensitive in the IR spectrum only.

- **Optogalvanic effect:** [10] This is based on the resonant absorption of light in a gas discharge which alters the discharge current. If the ionization rate of the excited state is higher, a larger current results, otherwise a smaller. This depends on the lifetimes and ionization probabilities of both states. The optogalvanic effect is used for spectroscopy, where a sample in a gas discharge or flame is probed by a tunable laser. It is also employed to stabilize the frequency and power of lasers, or calibrate their wavelength (e.g. the Laser Galvatron series from Hamamatsu [492]). High photon fluxes are required for the detection and the resulting current change can persist for 10–100 ms, which makes sensors which use this effect unsuited for LSNDs.

- **Photochemical:** In photochemical sensors, the excited states which are caused by photon absorption induce chemical reactions in the material [473]. For example, in photographic film [11] molecules in small silver halide crystals are split into Ag and halogen atoms by incident photons [493; p. 277] [153; p. 727]. In the photoreceptor cells in the retina of the mammalian eye, photons induce the isomerisation of retinal molecules which starts a chemical cascade [494]. These sensors are unsuited for nanosecond timing, however.

Of the plethora of secondary detection mechanisms which are based on IPE, only photoconductivity and the photovoltaic effect allow the detection of UV/visible light near room temperature with acceptable time resolution. Therefore, only these two mechanisms will be considered in the following.

Polarization [12]

There is one other effect besides the photoeffect, which allows to detect photons: **quantum non-demolition measurement** (QND).
This method aims at no less than measuring the number of incident photons *without absorption*. At a first glance, this seems to contradict quantum mechanics, since every measurement has to affect the measured system. However, QND is in fact based on a clever use of Heisenberg's uncertainty principle: For the complementary observables photon number and phase, the quantum noise – which inevitably results from a measurement (quantum back action) – is added solely to the phase, while the number of photons is left intact.

[10] Content based on [490, 491].
[11] And the older photographic plates.
[12] Content based on [478; ch. 3].

This concept has been demonstrated to work using the cross-Kerr effect (which is based on polarization) for classic light constituting of many photons [478; p. 69]. However, single photon measurement is still challenging and is the subject of ongoing studies. Non-destructive detection of single photons was achieved in microwave cavities for *static* cavity photons using qubit-photon CNOT gates, the ac Stark effect and the intrinsic phase shift in Rabi oscillations [478; p. 70]. But the detection of moving single photons – the ultimate goal of QND – is still not possible, which currently excludes this mechanism for LSNDs.
A good summary of the matter and the current state of these promising developments is given in [478; ch. 3].

4.1.1.2 Conversion medium

Since QND is not able to detect incident single photons yet, the only eligible detection mechanism is the photoeffect with its subtypes and secondary detection mechanisms: the internal photoeffect (photoconductivity, photovoltaic effect), the external photoeffect and photoionization.

Sensors which are based on these detection processes can be distinguished by the material in which the conversion from photon to photoelectron takes place.

In semiconductors and insulators (see p. 51), for the IPE is is sufficient if the photon energy is larger than the band gap E_g. For the EPE the energy must exceed the photoemission threshold W_{ph}, which normally is the sum of E_g and the apparent electron affinity $E_{A,app}$ (see p. 51 and fig. 2.3) and usually is larger than E_g. For negative electron affinity (NEA) materials, however, W_{ph} is equal to E_g (see p. 65 and fig. 2.3). The minimum photon energy imposes a limit on the maximum detectable wavelength. For insulators (E_g typically $>4\,\mathrm{eV}$ [495; p. 1312], $5\,\mathrm{eV}$ acc. to [489; p. 527]) the maximum wavelength already lies in the UV spectrum (see eq. (2.5)) and visible light cannot be detected.
In metals (see p. 53), the Fermi level E_F lies in the CB and the minimum photon energy for the EPE is the work function W, which is the difference between the vacuum level E_0 and E_F. For the IPE, on the other hand, in principle no minimum energy is required. However, the mean free path is very short due to the scattering on free electrons, so metals only play a minor role as photocathodes for the EPE and are not used at all for IPE sensors to the author's knowledge.
In a gas the photon energy which is necessary for pe creation is equal to the ionization potential of an atom or molecule.
In liquids the lower distances between the constituents already lead to the formation of thin bands. This reduces the ionization threshold (internal emission like IPE) and to a lesser extent also the photoemission threshold (external emission like EPE) compared to the corresponding gas [196; pp. 2–4] (see p. 48). The emission process here lies in between PI and EPE and as for the EPE transport and extraction processes play a role.

In all studied photosensors which are based on the IPE, a semiconductor was used. EPE sensors *also* mostly employ semiconductors as photocathodes but occasionally subtypes with metal cathodes exists (see pp. 51 and 53). No distinction is made in fig. 4.1, since the differences for other operating principles are more important.

Sensors which are based on photoionization mostly use gaseous photocathodes, but in the 1980s and early 1990s also attempts were made with liquid photocathodes using tetrakis (dimethylamino) ethylene (TMAE), triethylamine (TEA) or ethylferrocene (EF) (all UV-sensitive only). [196; ch. 5] This is achieved by covering a metal plate with a layer of adsorbed gas. For a thin layer on a metal cathode, the photons still can pass the liquid. The adsorbed layer rather affects the work function of the metal and hereby can increase its quantum efficiency (QE) and extend the sensitivity to longer wavelengths (also [196; p. 6]). For a large thickness, all incident light is absorbed in the liquid layer, whose properties now determine the sensor performance. Since the subsequent transport and amplification processes are identical for gaseous and liquid cathodes, both categories were summarized in one group in fig. 4.1.

4.1.1.3 Transport

For the IPE the transport of electrons and holes to readout or amplification stages is achieved by electric fields. These are either applied externally (power consuming) or result from a built-in field which separates charge carriers produced by photons (photovoltaic effect, e.g. in solar cells). Since the transport necessarily has to occur inside the conversion material, the transport medium is the same (i.e. a semiconductor).

For sensors based on the EPE, the photoelectrons can either be emitted into a vacuum or a gas. In both cases electric fields are required to remove the photoelectrons from the vicinity of the cathode, as otherwise the electrons would fall back onto its surface and recombine. In a gas the transport medium itself can be used for amplification, whereas in a vacuum other means are required (see below).
For the EPE, in principle one could also consider the ballistic transport or diffusion of the pe to the surface as an additional transport step. However, since this step is a necessary part of the EPE and is irrelevant for the distinction of EPE sensors, it is omitted here.

For photoionization sensors, the photocathode (meaning the place where the PI occurs) is either the gas which is also used for transport and amplification or an absorbed layer of this gas on a surface. Liquid cathodes are designed to emit into the gas and not the metal substrate to employ gas amplification. Therefore, the transport medium in PI sensors is always a gas, where the electrons are accelerated through electric fields. For liquid cathodes, the application of an electric field across the gas–liquid interface can also assist the photoelectron transport inside the liquid and facilitate the emission from the liquid [196; p. 4].

Vacuum tube focusing types [13]

For vacuum-based EPE sensors, the electron optics of the accelerating field between the photocathode and the next stage in general is designed according to one of three different layouts (see fig. 4.2):

- **Proximity focusing:** As the name implies, the next stage is close to the photocathode and the photoelectron paths are very short. This leads to compact (flat) sensors, high field gradients (linear fields), short transit times (see eq. 2.135),

[13] Content of this section based on [244, 327, 472, 496] [196; pp. 295–296] [153; pp. 297–298].

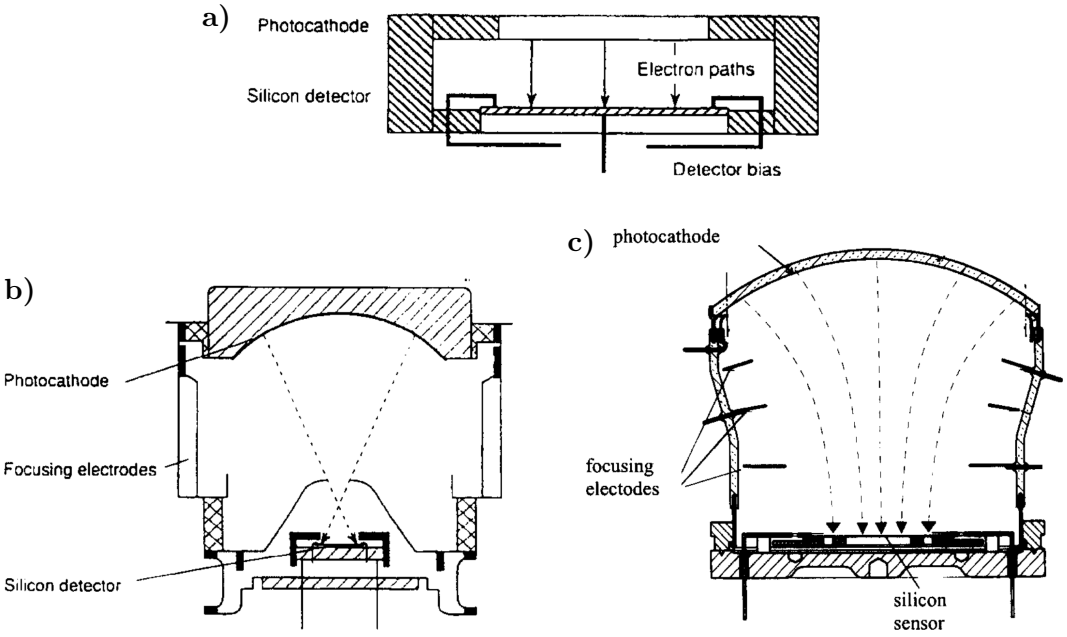


Figure 4.2: Focusing types in vacuum tube photosensors, for the example of hybrid photosensors: a) Proximity focusing. b) Cross focusing. c) Fountain focusing.
All figures from [472]. Discussion see text.

a small transit time spread (TTS) (eq. (2.138)), and high magnetic field tolerance. It also allows one-to-one mapping of points on the cathode to points on the receiver electrode (amplification stage or secondary sensor), since the spot size is small due to the short distance and large fields. This enables to build a position-sensitive photosensor if suitable subsequent amplification and readout methods are used (see below). If proximity focusing is employed in a hybrid photosensor, the secondary sensor has to have the same area as the photocathode (i.e. the primary sensor).
This focusing type is used for example in MCP-PMTs [14] [169] and designs based on them [478,497], mesh and metal channel PMTs [169] and some HPD [472,496] and hybrid APD (HAPD) types [249, 498].

- **Cross focusing:** In the cross focusing design, the electron trajectories run through a focal point before hitting the subsequent stage with mirrored axes (left $\leftrightarrow$ right, front $\leftrightarrow$ back vs. paper plane). This electrostatic lens can compensate the spatial spread of electron trajectories originating from the same emission spot which results from variations of the emission angle and velocity (spot size, see p. 179) to a large extent [15]. While this design results in low spatial resolutions, it shows small pin cushion distortions for large distances from the rotational axis and is sensitive to magnetic fields due to the long required distance between cathode and receiver. This field configuration also allows

[14] With the exception of the newly developed large MCP-PMTs GDB-6201 and GDB-6203 from NNVT (North Night Vision Technology Co., Ltd.) which use fountain focusing due to their size [279].

[15] Compare this to the transit time balancing used in linear focused dynodes (fig. 2.20 and p. 171), which also makes use of cross focusing to compensate the transit time differences due to different start points. Likely in this case the spatial spread from emission velocity for identical electron origins is compensated to a large extent as well.

to map the cathode onto the next stage and build a position-sensitive sensor. With cross focusing, a strong demagnification from the area of the emitter (cathode) to that of the receiver (next stage) can be achieved. Therefore, in hybrid sensors with this design the secondary sensor can be much smaller than the cathode.
Examples comprise PMT variants [172; p. 1-18], HPDs [191, 496, 499, 500] and electron bombarded CCDs [16] (EBCCDs) [244, 501].

- **Fountain focusing:** This focusing type also reduces the cross-section of the electron trajectories, but has a reduced spatial resolution compared to cross focusing, since the electron optics does not correct for the lateral momentum (ballistic spread). Fountain focusing is comparatively simple to realize (yet tricky to optimize) and allows large detection areas. For big sensors usually a cathode geometry which resembles a spherical surface is used to reduce the cathode time spread (CTS). This focusing layout is usually employed where position resolution is irrelevant or only plays a secondary role. It is similarly sensitive to magnetic fields as cross focusing.
This design is used in regular PMTs [169] as well as many (typically large) hybrid photosensors [41, 281, 282, 496, 502–505].

Cross and fountain focusing are sometimes also called electrostatic focusing [153; p. 298] [244, 472].
Older designs also occasionally incorporate magnetic focusing (e.g. for PMTs [170; p. 3] [184; p. 10] and HPDs [506]).

4.1.1.4 Amplification processes

In the following, the amplification processes which are employed in the photosensors considered here (the ones shown in fig. 4.1) will be explained.

Since all relevant photosensor types are based on the photoeffect and thus the production of photoelectrons (and holes or ions), in all materials amplification processes exists which are based on impact ionization: The kinetic energy of the charge carriers [17] is increased by acceleration in an electric field until it is sufficient to ionize an atom or molecule on scattering and thus create additional electrons and holes/ions. The process can be repeated for the newly generated charge carriers and is this manner creates an electron avalanche.
Depending on the transport medium, different names are used for this effect and will be adopted here for distinction: In semiconductors it is labeled simply impact ionization, in vacuum it is dubbed electron bombardment (independent of the type of impact material), while in gases it is called Townsend avalanche.

In semiconductors another amplification process apart from impact ionization can be used. Through the use of structured materials it can be achieved that a single electron or hole manipulates a larger amount of charge carriers and in this manner creates an amplification (transistor, conversion gain; see below).

[16] Charge coupled devices
[17] In sensors based on gaseous ionization detectors, the Townsend avalanche is produced only by the electrons [194; ch. 11].

Semiconductor

Impact ionization

All semiconductor photosensors are based on the internal photoeffect, in which incident photons whose energy exceeds the band gap create electron–hole pairs.
In a photodiode[18] four regions of operation can be distinguished [293; pp. 2–3] [489; pp. 530–533]:

- **Photovoltaic mode:** For zero bias, the diode is in photovoltaic mode and incident photons build up a voltage through the created charge carriers, which can be measured. While this has the advantages of producing no dark current, the time response is slow due to a high junction capacitance, long wavelengths are suppressed, and no amplification takes place [489; pp. 532–533]. Therefore, the sensors relevant here are all based on reverse bias voltages (photoconductive mode), which comprises the other three operation modes (see fig. 4.9c).

- **Photodiode mode:** In the photodiode regime, the applied reverse voltage separates electrons and holes and draws them to opposing diode ends (top left in fig. 4.9c). No amplification occurs and the output current is proportional to the light intensity.

- **Linear mode:** If the applied reverse voltage is large enough, the charge carriers gain sufficient kinetic energy in the field to create additional electron–hole pairs through impact ionization, by raising bound electrons from the valence band to the conduction band. If the applied voltage is high enough, the generated charge carriers themselves can also cause impact ionization and an avalanche is created. For low voltages only the electrons can gain sufficient kinetic energy for impact ionization, since the holes have higher effective masses. Therefore, the avalanche only grows in one direction (center of fig. 4.9c), which limits the amplification. Since for a given voltage the output current is proportional to the number of detected photons (linear gain), this voltage range can be likened to the proportional counting region in gaseous ionization detectors.
  Examples of sensors operating in the linear gain regime comprise avalanche photodiodes (APDs) with gains between about 10 to 200 [194; p. 308]) and electron-multiplying CCDs (EMCCDs, fig. 4.10c). The latter employ impact ionization in their gain registers, but with far lower amplifications of only about 1.013 [244]. This is compensated by the large number of gain stages, which results in a total gain of several 100 or 1000 [244, 507].

- **Geiger mode:** If the reverse bias is increased further, above the breakdown voltage holes can also gain sufficient kinetic energy to cause impact ionization and the avalanche spreads over the complete diode (avalanche breakdown; right in fig. 4.9c). This can be triggered by a single photon and since the avalanche sustains itself, it needs to be quenched to avoid damage to the device. The simplest way to achieve this is by placing a resistor in series with the diode. The voltage drop over the resistor then increases with current and lowers the bias voltage (passive quenching) [478; p. 43]. Because the avalanche spreads over the complete diode, the output charge is the same for every event and

[18] Except for the photoresistor and the photogates in CCDs, *all* semiconductor photosensors consist of photodiodes in one form or another.

the required recharge after photon detection introduces a dead time. All this resembles a Geiger counter, which is why this regime is called Geiger-mode. This is the operation range used in single photon avalanche diodes (SPADs), which form the cells of a silicon photomultiplier (SiPM). For this reason SiPMs are also called Geiger-mode APDs (G-APDs) [19].

Transistor

The charge carriers which are created by the internal photoeffect can be used as the base current of a transistor. This amplifies this current and produces a significantly larger collector current (see fig. 4.9h) [489; p. 536].

Conversion gain

Charge carriers can also be amplified with an built-in charge amplifier. This is achieved by placing them in a very small capacitor, which results in a large voltage change on the capacitor that can be measured with a high impedance amplifier, in this case a field effect transistor [508; p. 923]. The conversion or charge gain g_c (unit V/C) denotes the change of the output voltage U_{out} per input charge Q_{in} and increases with falling capacitance C [509]:

$$g_c = \frac{U_{\text{out}}}{Q_{\text{in}}} = \frac{1}{C} \tag{4.1}$$

This amplification principle is used in charge-coupled devices (CCDs) for the readout. In the quanta image sensor (QIS) the capacitance is minimized (floating diffusion capacitor) in order to increase the conversion gain and surpass voltage noise. Here a g_c of up to about $370\,\mu\text{V}/\text{e}^-$ was achieved, which allows single photon detection [510].

Conservation gain [20]

This amplification mechanism can be used in a junction-less semiconductor with ohmic contacts. If the hole mobility is low and the carrier lifetime is larger than the electron drift time to the anode, the hole still remains in the semiconductor when the electron leaves the anode. To conserve charge neutrality, another electron is inserted at the cathode and again drifts to the anode. This process repeats for as long as no recombination occurs and the hole remains in the semiconductor (typically the holes are trapped in acceptor levels). This thus leads to a current gain which can be maximized by short cathode–anode distances and long carrier lifetimes.
Since no name for this process could be found in literature, it is called conservation gain here.

Electron bombardment

When photoelectrons which are produced by the EPE are emitted from the photocathode into a vacuum, a voltage must be applied in order to collect them on an anode.

[19] It should be noted that the established names tend to be somewhat misleading. G-APD would actually be the proper term for the SPAD (a photodiode using impact ionization in Geiger-mode). The label SiPM also is ambiguous, since it could be applied to a large number of photosensors, and the operating principle of an SiPM has nothing to do with that of a photomultiplier.
[20] Content based on [511; pp. 667–671] [489; pp. 538–540] [194; p. 392].

Otherwise, the photoelectrons would fall back on the cathode and recombine. When the electrons are accelerated by an electric field which is large enough that on impact on a stage [21] the number of particles is increased, this is called electron bombardment (EB) [268; p. 9] [169, 172] [170; pp. 411, 438].
Three different amplification mechanisms can be distinguished based on the produced particle type and the emission direction: Secondary emission, scintillation, and electron–hole generation. These processes and their subtypes will be elucidated in the following and are also shown in fig. 4.1 to provide an overview.

Secondary emission

The impact of the incident electron emits secondary electrons (SEs) from the bombarded solid material (typically a semiconductor or insulator [172; p. 1-13]) into the vacuum. The gain is the number of emitted secondary electrons, the so-called secondary emission yield (SEY). This multiplication mechanism was already discussed extensively for PMTs (see p. 69). A gain stage which employs secondary electron emission is dubbed a dynode, and various subtypes of dynodes exist [195; pp. 181–184] [172; pp. 1-14–1-16] [169; pp. 45–47]:

Reflective dynodes

In the usual configuration, the SEs leave the material from the surface which the incident electron enters (fig. 4.3a). This is similar to a reflective photocathode (p. 65), where the photoelectron is emitted from the surface on which the photon falls.

Discrete dynodes

For discrete dynodes one amplification per dynode takes place [22]. After emission, the SEs are drawn away and focused onto the next dynode or the anode by an electric field. This is the dynode class which is used in PMTs. Discrete dynodes have experienced a vast diversification throughout the decades (see figs. 2.1.3 and 4.3):

- **Linear focused (LF):** Linear focused dynodes (see p. 78 and fig. 2.15a) were developed as early as 1939 but are still commonly used for fast PMTs [170; p. 232] [169; pp. 46, 47] [171; p. 6]. They are designed for transit time balancing (see p. 171 and fig. 2.20), which leads to a good TTS. LF dynodes also have strong extraction fields, which results in short transit times, and very good linearity, since space charge effects are less severe. Furthermore, the design is comparatively compact. Terminations on the dynode sides prevent SEs from hitting the dynode mount and being lost. However, LF dynodes often show poor photodetection efficiency (PDE) uniformity. This is because the shape of d1 [23] trades a good collection efficiency (CE) for a lower TTS. Several of the PMT types which were analyzed in ch. 3 have LF dynodes.

- **Circular focused (CF):** Also called circular cage (CC), this dynode type was developed in 1939 as well and is basically an LF dynode stack which is curled up to a circle for compactness [170; pp. 232–234] [169; pp. 46, 47] [171; p. 6]. Its

[21] Either the anode itself or additional stages at potentials between the cathode and anode.
[22] Unless inelastic backward/intradynode scattering with SE production occurs (BL, BEnn, BLnn effects), see 2.3.5.
[23] dn denotes the n-th dynode.

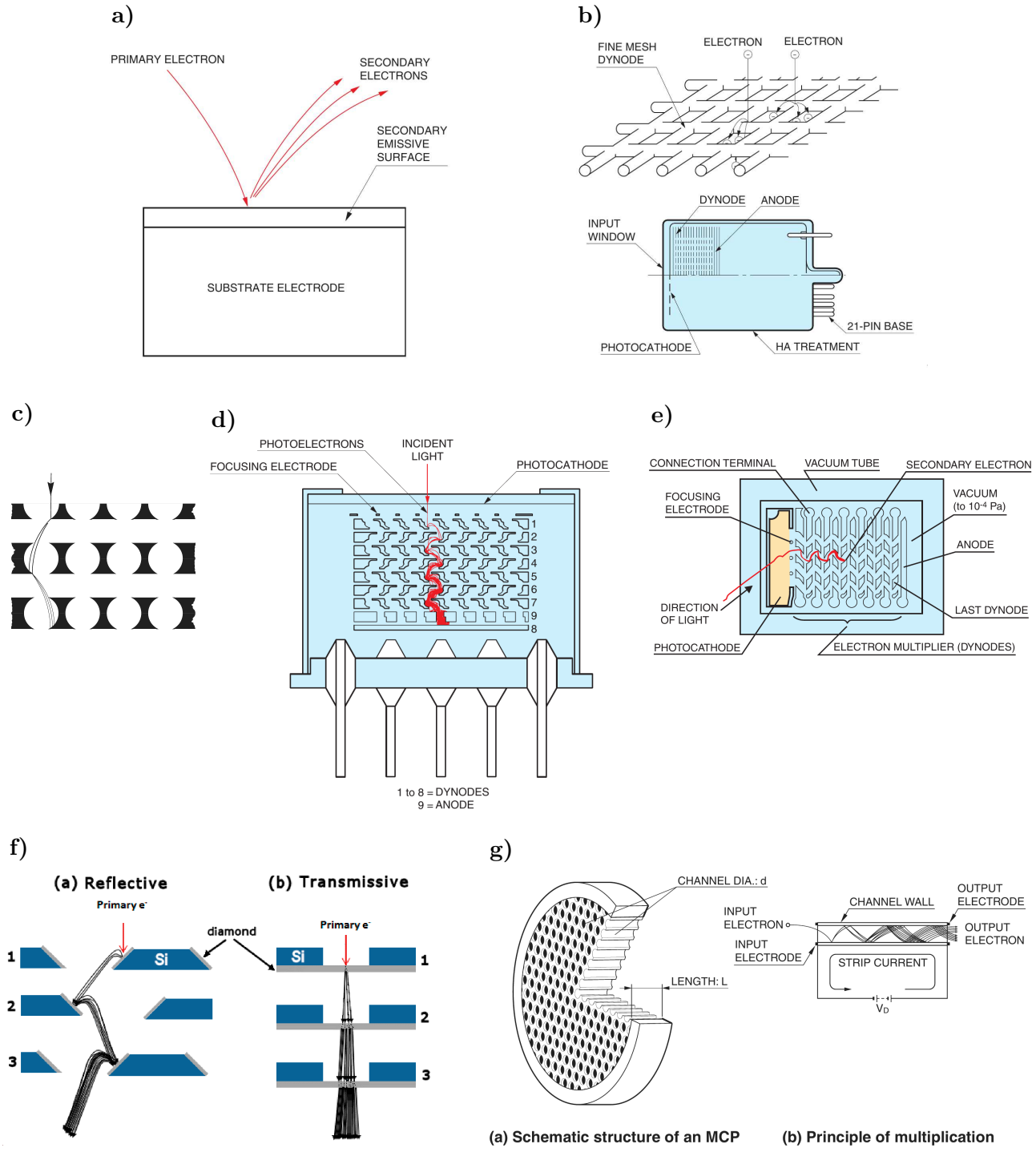


Figure 4.3: Dynode types and their operating principles. See fig. 2.1.3 for additional types. Discussion in text.
a) Reflective dynode, from [169; p. 19].
b) Fine mesh (FM) dynode, from [169; p. 253].
c) Foil dynode, from [172; p. 1-15].
d) Metal channel (MC) dynode, from [169; p. 18].
e) Micro PMT (photomultiplier), from [169; p. 205].
f) Transmissive dynode (tynode), from [512].
g) Micro-channel plate (MCP): structure and principle of operation, from [169; p. 212].

properties are thus similar to LF dynodes: good linearity, good TTS (transit time balancing is used here as well), short transit time, poor PDE uniformity, but also high afterpulse (AP) rates. The latter is likely due to the close proximity of anode and cathode, although a combined ion/light shield is placed between both. CF dynodes are used mostly for small side-on PMTs (not relevant here) and enable compact designs.

- **Box and grid (B+G):** Box dynodes (see p. 78 and fig. 2.15b) have the shape of quadrants on whose entrance aperture a fine honeycomb mesh is placed in order to increase the extraction fields from the previous stage [170; pp. 234, 235] [169; pp. 46, 47] [171; p. 6]. This layout has a very good CE due to the large size of d1, but shows large transit times, very long rise times and large TTS values – despite the meshes the extraction fields are very weak and amount to only about 1/10 of those in LF dynodes. B+G dynodes have low linearity and are very sensitive to magnetic fields but have low AP rates. They are unsuited for the purposes studied here.

- **Box and line (B+L):** Box and line layouts (see p. 78 and fig. 2.15c) combine the LF and B+G designs to utilize the advantages of both [169; pp. 45, 46]. They show shorter transit times, lower TTS values and better linearity than B+G dynodes and have a higher CE than LF. This dynode type is often used in large area PMTs, for which the CE on d1 is a severe issue (see p. 179). The majority of PMT types which were studied in ch. 3 have B+L dynodes.

- **Circular and linear focused (C+L)**: Similarly, CC and LF designs can be combined advantageously to provide a better linearity than with CC dynodes for a design which is still very compact.

- **Venetian blind (VB):** In the venetian blind layout, a dynode consists of multiple tilted slats placed next to each other (see p. 78 and fig. 2.15d) [170; pp. 225–229]. The extraction field is provided by a mesh below the dynode and the slats of the next dynode are angled in the opposite direction. However, in the top half of the dynodes the extraction field is mostly shielded by the slats, so many SEs which are emitted from this region hit the wall of the neighboring slat. This results in a strong edge effect (see p. 84) – for electrons hitting the top of the dynode, the d1 gain[24] $\delta 1$ is only about half of that for a hit close to the bottom. In addition, stage skipping occurs when a pe trajectory has a similar angle as the slats, which leads to a very bad PDE uniformity, although the overall CE is good due to the large d1 area. Both edge effect and stage skipping broaden the single electron response (SER; the charge distribution), which often resembles a quasi-exponential decay with more of a bump in the slope than a single photon (spe) peak. Furthermore, VB dynodes have large transit times and TTS values and are quite large [169; pp. 45, 46] [171; p. 6]. On the other hand, they show very low AP rates, likely since the chevron layout blocks ion and photon paths. Due to the multitude of drawbacks, this design is disappearing from the catalogues.

- **Mesh:** Developed in 1981 based on the VB design, **coarse mesh dynodes (CM)** consist of wires with triangular cross-section [170; pp. 229–230] [169; p. 46] [172; p. 1-15]. The close proximity and alternating wire directions of subsequent mesh layers focus the SEs and preserve the position information.

[24] δn denotes the gain of the n-th stage.

In early designs this focusing was assisted by an axial magnetic field, which changes the SE trajectories into spirals [513]. However, it was found by [514] that the magnetic field can be omitted at the cost of a slight spatial spread, and newer commercial devices use no magnetic field. Gains of up to 3 per stage can be achieved, which for 16–20 layers can result in total gains of 10^6–10^9 [153; p. 300]. In some designs, a layer of guard plates is placed below the triangular gain dynodes with aligned holes and the same voltage as the next triangular layer. This additional layer focuses the SEs onto the subsequent gain layer and attracts ions to reduce the ion feedback [153; p. 301].
In the 1980s the demand for operation in high magnetic fields required to place the wires even closer and led to the development of **fine mesh dynodes** (**FM**, fig. 4.3b) [170; p. 230] [169; p. 46]. Here, two perpendicular sets of parallel circular wires merge to form a mesh. The subsequent mesh is placed with an offset to reduce the passage of light and ions.
Both coarse and fine mesh dynodes show excellent linearity, high magnetic field immunity and short transit times, are compact, and allow proximity focusing. In addition, if a cross-wire anode or multianode is used – optionally with charge division or a summation network – (see 4.1.1.5) the photon position can be measured with a spatial resolution of a few mm [184; pp. 44, 182–185]. However, as for VB dynodes, the mesh layout will be prone to stage skipping, which leads to a poor CE and likely also a bad charge resolution [169,170]. According to [153; p. 300] the PDE uniformity varies by up to factor of 2 over the photocathode surface due to QE and CE variations. Fine mesh dynodes are suited for high magnetic fields in the order of 1 T and show very small transit times and TTS values (see table 2.1) [169,170].

- **Metal channel (MC):** First reported in 1993, metal channel dynodes (fig. 4.3d) consist of thin micro-machined electrodes which are closely stacked [170; pp. 230–232] [169; p. 46]. A focusing mesh above d1 focuses the photoelectrons into the metal channels. This layout leads to short electron paths and thus very low transit times and TTS values (see table 2.1), good magnetic field immunity, and very compact designs. MC dynodes also permit to construct position-sensitive sensors by using a cross-wire anode (p. 605) or multianode (p. 606), possibly with charge division or a summation network [169; pp. 190–196, 200–202]. Spatial resolutions down to about 3 mm were reported [215]. For this dynode type, a high $\gamma 1$ probability (see p. 282) is found, especially for wavelengths above 550 nm. In addition, for off-center photoelectrons stage skipping occurs, which causes a second peak at 1/3 pe charge in the SER.
  Photonis offered so-called **foil dynodes**, which are thin parallel perforated foils (fig. 4.3c) [172; p. 1-16] [170; p. 232]. While the dynode profile is different from the Hamamatsu metal channel dynodes, the concept is the same, and foil dynodes can be considered an early subtype of the MC layout.

If the target it to construct a position-sensitive PMT, this requires that a spatial separation between cascades from separate photocathode origins is maintained. [153; p. 300] This is only the case for CM, FM, MC (and foil), and in principle VB dynodes, since these layouts consist of parallel amplifications channels.
For dynode types where all photoelectrons are focused on a common dynode, position-sensitivity can only be obtained by placing several such dynode chains next to each other – which is basically a repetition of the idea used in the segmented dynode

layouts. This was realized for a large B+L PMT (Hamamatsu R7081, 10”), which was retrofitted to obtain position-sensitivity by including four independent dynode chains [449] (called Ma-R7081 here).
Since in position-sensitive PMTs a segmented anode is used for readout, they are also called multi-anode PMTs (MaPMTs).
The spatial spread of the cascade can lead to crosstalk between neighboring amplification channels, which deteriorates the position-sensitivity. [153; p. 300] The reduction of this spread was one of the reasons to develop FM and MC dynodes.

Continuous dynodes[25]

For continuous dynodes, more than one amplification process takes place per dynode. This can be achieved by coating the inside of a glass tube with a resistive secondary emitter material, which then acts as both voltage divider and electron multiplier at the same time. The primary electron knocks out SEs. which themselves hit the tube walls and produce secondaries. This creates a cascade which grows while proceeding through the channel towards the anode side (right in fig. 4.3g). The SEY of each wall impact is only about 2 (1.2–3), but the total gain rises with the tube length and the gain–voltage dependence follows a power law just as for PMTs (see eq. 2.102). For a fixed voltage the gain depends on the ratio of length (L) to diameter (D). For a higher L/D, the number of impacts increases but the SEY of an impact decreases, since the mean impact energy is lower. The maximum gain is obtained for $L/D \approx 50$. The number of wall impacts varies greatly between individual electrons, which results in poor gain characteristics: If the channel is operated in the linear regime, the total gain for an incident photoelectron follows a negative exponential distribution. However, if a high voltage is applied and the channel is in charge saturation mode, a sharply peaked distribution is obtained. This is basically the dynode equivalent of the Geiger mode for photodiodes. At a sufficiently high gain, all pulses produce the same output charge independent of the number of incident photoelectrons. In this case, the highly resistive coating first has to be recharged before another cascade can be produced, which introduces a dead time.
There are two subtypes of continuous dynodes:

- **Channel electron multiplier (CEM):** In the CEM[26], a single large glass tube of up to several cm length is coated with a semiconducting SiO_2 layer. The CEM can reach gains of 10^7 for 2–3 kV channel voltage[27] and a peak-to-valley ratio (P/V) of 20 or more in charge saturation mode. The channel is usually curved or serpentine to frustrate light and ion feedback (which lead to uncontrolled additional gain and can even cause self-sustaining breakdown) and improve the gain. The dark current is lower than in PMTs. However, the linearity is limited by the charge saturation and the device is insensitive during a long recharge time which can have ms time constants. The excess noise factor (ENF) is > 2. Such single-channel devices are mainly used for X-ray and ion detection.
  Parallel-plate channel multipliers are a variant of the CEM, where instead of a tube a pair of plates are covered with a secondary emitter. Charge resolutions of as low as 16% were reported, but no commercial product emerged.

[25] Content based on [170; pp. 225–226, 237–241] [267; pp. 45–50].
[26] Also called Channeltron after the product name of the Bendix corporation, which invented this design in the 1960s.
[27] [172; p. 1-23] even mentions gains of 10^8 for spiralized devices.

- **Micro-channel plate (MCP)** [28]**:** In a MCP up to several million microscopic channels [29] with diameters of 2–100 µm and lengths typically below 1 mm (0.4–2 mm) are fused together in a lead glass plate (left in fig. 4.3g). Common metal electrodes (usually Nichrome or Inconel) are evaporated onto the top and bottom of the plate. The resistance between the electrodes can reach $10^9\,\Omega$. With a single MCP gains of up to ca. 10^4 can be obtained. When two or three MCPs are stacked and operated in charge saturation mode, the gain is sufficient for spe detection and a very low charge resolution results, which allows photon counting. While this introduces a dead-time after photon detection for individual channels, the large number of channels makes a re-hit of the same pore during the recharge unlikely, which eliminates this drawback for LLLs. This is the same principle as in an SiPM which consists of many SPADs that all have dead time. Furthermore, for MCPs position-sensitive readout is possible. Since the performance depends mainly on the L/D ratio, miniaturization is beneficial, as it allows shorter transit and rise times, improves the spatial resolution, extends the dynamic range, and reduces the recharge time (smaller channel capacitance and resistance).
  Usually, the channels are angled at about 8–10° to the surface to reduce light and ion feedback and to increase the gain for pe trajectories which are perpendicular to the surface. Subsequent MCPs are rotated by 180°, resulting in a chevron pattern (fig. 4.11a), which significantly improves the gain compared to straight channels with the same voltage, since the SE paths are intercepted. This also helps to reduce ion and photon back-flow. Using a Chevron configuration, however, introduces coupling losses and at high voltages the charge saturation causes the gain to level out. If a gap is introduced between the MCPs, the SEs from the first MCP can also hit adjacent channels which further raises the gain.
  MCPs have excellent TTS, transit time and linearity, high magnetic field immunity, and allow position sensitivity, but the open-area ratio of channels on the MCP surface (equivalent to a fill factor for the CE) is usually only 50-60% [169; p. 46] [267; p. 49] (see p. 641 for novel MCP types which allow higher open-area-ratios).
  In newer developments, the area surrounding the MCP pores is covered by a secondary electron emitter material to improve the CE (see p. 644) [251,252,516]. Photoelectrons which hit the region between channels then produce SEs, which can create a cascade, or are backscattered and have another chance to hit a channel. In the latter case, this introduces an additional transit time and thus produces late pulses (LPs) which significantly increase the TTS. Recently, it was proposed to surround the channels with a bowl-shaped high SEY material instead of a flat coating to improve the TTS (see p. 644) [517].

Transmission dynodes [30]

Transmission dynodes (also called tynodes) consist of a thin film and, unlike reflective dynodes, the SEs are not emitted from the surface through which the primary electron enters but from the opposite side (fig. 4.3f compares both dynode types). This design allows proximity focusing and imaging, similar to an MCP or fine mesh dynodes.

[28] Additional content from [196; p. 282] [169; pp. 212, 215] [515; p. 420] [194; pp. 302–303].
[29] Called channels, pores, or capillaries.
[30] Section based on [253] [170; pp. 234, 236–237]

In addition, the very short dynode distances result in an extremely low TTS in the order of a few ps.
To distinguish both modes, one can define a reflection secondary electron yield (RSEY) for reflective dynodes and a transmission secondary electron yield (TSEY) for tynodes. For thin films and large impact energies the primary electron often retains sufficient energy to leave the tynode on the backside. Therefore, also a transmission electron yield (TEY) is introduced, which is the sum of TSEY and the fraction of forward scattered primaries which are emitted from the backside. Similarly, a reflection electron yield (REY) can be defined as the sum of RSEY and the fraction of backscattered primaries which leave the material from the incident surface. In this work, the term SEY is used synonymous with the reflection electron yield and includes both primaries and secondaries.
The requirements a material has to fulfill to have a high TSEY are similar to those for a high RSEY (see 2.1.3): A large escape depth (i.e. diffusion length) is required, which eliminates metals due to the abundance of free electrons which are available for scattering. Furthermore, a low electron affinity (or better a NEA) is needed for a high escape probability at the surface. These requirements are best fulfilled by semiconductors and isolators.
In tynodes the incidence angle of the primary electron is perpendicular to the surface to maximize the penetration depth so that the SEs are produced close to the backside, which maximizes the TSEY. For reflective dynodes, on the other hand, the a slanting incidence angle is better, since then SEs are produced close to the surface, which increases the RSEY.
It was found experimentally that the RSEY is always higher than the TSEY and that the TSEY is largest when the membrane thickness equals the SE escape depth. Since most SEs originate close to the emission surface, the energy of the incident electron needs to be higher for tynodes than for reflective dynodes due to energy losses of the primary electron in the material.
Because the SE emission leaves holes in the material, a sufficient conductivity is important to avoid charge build-up, since this would produce fields which impede SE emission. For reflective dynodes this is easily achieved by placing a thin secondary electron emitter material on a conductive substrate. For thin tynode membranes, this is more challenging and e.g. a thin conductive coating on the incident surface can be used to increase the conductivity and reduce the recharge time (for the last tynode of an spe-sensitive sensor a significant amount of SEs would be emitted). However, metals layers are problematic, since they cause high energy losses for the primary electrons due to scattering on free electrons, and drastically reduce the TSEY.

The concept of transmission dynodes was already studied in the 1960s and 1970s, but only for insulator layers with comparatively large thicknesses of 1–10µm for better mechanical stability. This required large acceleration voltages of several kV and materials with high diffusion lengths. Furthermore, the tynodes showed aging and charge buildup. The measured TEY was about 5 times lower than the REY, and for GaAs for the best sample a TEY of ca. 20 at 6 kV was obtained. In two other publications for GaAs and KCl a voltage of more than about 5 kV was required for a gain of 1, which was likely due to low diffusion lengths and an Al backing layer.

In the last years, new efforts have been undertaken to improve the properties of transmission dynodes to obtain a Timed Photon Counter (TiPC) single photon photosensor. Here, an array of miniature tynodes in form of ultra-thin dome shaped

membranes is embedded in a rigid substrate (see fig. 4.10g), since the area of such thin films is limited to a few mm^2. The curved shape provides higher mechanical stability, which allows to reduce the thickness, and focuses both the incident and the emitted electrons onto this and the next tynode in a proximity focused layout. This dome focusing allows operation in high magnetic fields of around 1 T. Due to the close spacing and the low variation of electron path lengths, a TTS of a few ps is expected. Ion back-flow and photon feedback is blocked from later stages by the tynodes, which prevents degradation of the tynodes and the cathode.
As conductive layer materials, TiN and C have been tested, while graphene might be a future alternative.
Several TSEY materials have been studied, including silicon nitride (SiN), aluminum oxide (Al_2O_3) and magnesium oxide (MgO). The highest TEY (5.5) was found for a 5 nm MgO layer with a 2 nm TiN conductive layer at 1.2 keV electron energy [518].
Diamond and (again) GaAs might be promising materials to further increase the TEY, and a surface termination with H or Cs would reduce the electron affinity and could produce a NEA material.
In addition, the Schottky effect could be used to lower the work function. For insulators the field could even propagate into the material, which would accelerate the SEs to the surface.
Finally, for porous materials high fields can be created in the cavities due to charge accumulation, which could lead to an internal cascade generation and might be used to create an active tynode (trynode).

Scintillation

If the bombarded target is a solid scintillator or a gas, the impact of an electron can also produce multiple photons through excitations and radiative relaxations (see fig. 4.12, 4.16d–4.16f, 4.10f).
For the detection of these secondary particles *another* photosensor is required, which makes a photosensor that employs scintillation as gain process a *hybrid* photosensor. This arrangement effectively constitutes a light amplifier for the incident photons (more accurately an electro-optical preamplifier) [172; p. 1-22] [519].
The amplification is achieved through the conversion of the photon to a photoelectron via EPE and acceleration of the electron in an electric field. Either the gained kinetic energy is converted into multiple photons on impact on a solid scintillator [31] or the pe starts a Townsend avalanche in the glow discharge regime whose emitted photons are used [32]. To further increase the light yield for the first variant, additional electron multiplication steps can be used before the bombardment of the scintillator [33].
If the gain of the first amplification stage is high, this results in a good charge resolution (see eq. (2.65)). For this reason and because also the PDE of the secondary photosensor has to be taken into account, very large acceleration voltages of up to several ten kV are used for the bombardment of the scintillator to produce a large number of photons [34].

Electron–hole generation

Electron impact on a semiconductor not only produces SEs which are emitted from the surface into the vacuum – the gain process used in dynodes. It also creates

[31] For example, in the QUASAR-370 (p. 654), Abalone (p. 661) and HGPM (p. 664).
[32] This is done in the DGPM (p. 706) and optical readout GPGDs (p. 709).
[33] Used in the ICCD (p. 648) and ICMOS (p. 649).
[34] E.g. 25–40 kV for the "SMART" (see p. 653) which creates 30–50 detected photons [172; p. 1-22].

electron–hole pairs through impact ionization (fig. 4.14a), which can be read out or amplified (see above) *inside* the bombarded semiconductor. In this case one uses the electrons which are emitted into the CB and do *not* leave the material and/or the created holes. The relationship between electron-hole generation and secondary emission is the same as that between internal and external photoeffect, only that here the excitation occurs through electrons in place of photons.
Usually a high impact voltage is used, which leads to a very large bombardment gain and – if the overall gain is high enough for photon counting – a good charge resolution [169; p. 46].

Townsend avalanche

Mechanism [35]

In gases the number of charge carriers is increased through avalanche multiplication. Initially, a free electrons is created in the gas through the photoeffect (photoionization in the gas or photoelectron emission from a photocathode) or ionizing radiation. The electron then gains kinetic energy in an electric field and can ionize gas constituents through inelastic collisions once its energy exceeds the ionization potential. This creates an additional electron and ion, which are accelerated in opposite directions and also free electrons through ionization, which results in an exponential growth of the number of electrons and ions. In this so-called Townsend avalanche, electrons move 100–1000 times as fast as the heavier ions, for which reason almost only the electrons create charge carriers. The electrons drift to the readout anodes, where they produce a fast, short pulse, while the ion cloud slowly drifts to the cathode where it induces a long and slow pulse. The combination of diffusion (thermal motion), drift (external field) and the difference in electron and ion mobility leads to a drop-like shape of the avalanche where the electrons are located at the head (see fig. 4.4a) [520].
Naturally, this process of ionization through the acceleration of negative and positive charge carriers is closely related to the impact ionization in semiconductors and the electron bombardment of solids. For EB the acceleration occurs in a vacuum, for a Townsend avalanche in a gas, and for impact ionization in a solid.

Operation regions of gaseous detectors [36]

In gaseous ionization detectors (GIDs, short gaseous detectors (GD)) particles ionize a gas inside an electric field which separates the generated electrons and ions and allows to detect them as currents on the anode and cathode [515; p. 179].

The gain behavior of these detectors depends on the applied voltage and is depicted in fig. 4.4b for a single-wire counter (SWC, also called counting tube). An SWC consists of an anode wire which is surrounded by gas and centered in a cylindrical cathode.
Based on the shown gain–voltage characteristic, generally six different regions can be distinguished for GDs:

- **Recombination region:** In region I electrons and ions can recombine due to the small field. The gain g is < 1, which is why this region is not used for GDs.

[35] Content based on [196; pp. 9–11].
[36] Content based on [521; pp. 113–115], [515; pp. 193–196].

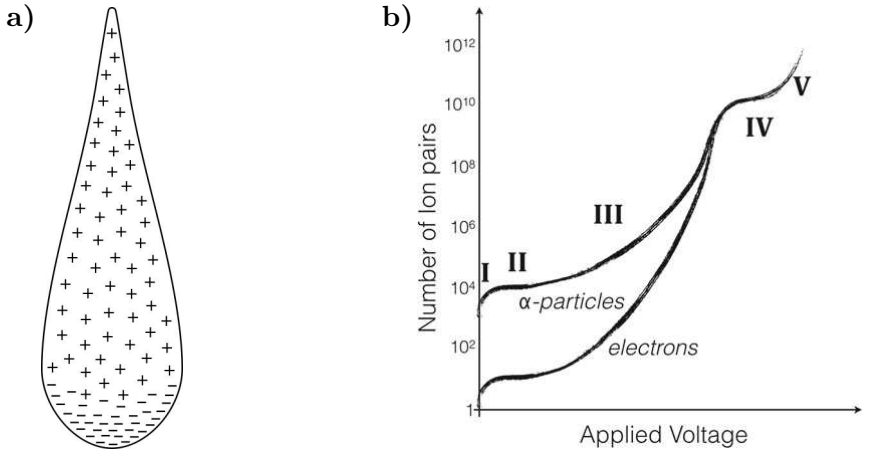


Figure 4.4: Gaseous detectors (GDs):
a) Townsend avalanche: Schematic distribution of electrons and ions, based on [522; p. 38].
b) Operating regions of single wire counters (SWCs) for incident electrons and α particles, from [521; p. 115]. Discussion in text.

- **Ionization chamber region:** In contrast, in region II the field is sufficient to collect all charge carriers. This results in a signal which is proportional to the number of created ion pairs and that is *independent* of the applied voltage. This region without amplification ($g \approx 1$) is used for ionization chambers and GD types derived from them.

- **Proportional counting region:** For higher voltages in region III a Townsend avalanche is started, where the number of secondary ionizations is proportional to the number of primary ionizations ($g \approx 10^3$–10^5). This is because only electrons create ion pairs and only close to the anode wire the field strength is large enough for ionization. Throughout the largest part of the device the electric field is in the ionization chamber region (i.e. just high enough to prevent recombination). Therefore, all electrons which are created outside the avalanche region produce the same number of ion pairs, which only depends on the voltage and sensor geometry. Each ion pair produces a single avalanche, so the output current is proportional to the deposited energy, which allows to use the device as a calorimeter and for particle identification. This operation regime is used for GDs which are based on proportional counters, e.g. multi-wire proportional chambers (MWPCs). In order to construct a gas-based photodetector, the proportional counting region is especially interesting, since it provides high gains and still has a notable dynamic range.

- **Region of limited proportionality:** Between regions III and IV, UV photons start to be created by excited gas particles (glow discharge), which triggers additional avalanches ($g \approx 10^5$–10^8). An ion cloud forms around the anode, which only slowly drifts away and thus builds up space charge which limits proportionality. Therefore, this region is not used to construct GDs. In order to prevent a proportional counter from entering this region, often an organic quenching gas which acts as a photon absorber is admixed (photodecomposition of polyatomic molecules [194; p. 243]).

- **Geiger region:** In region IV the discharge spreads over the entire anode, so the output current is saturated and independent of the number of primary ionizations ($g \gtrsim 10^8$). Since this produces a continuous discharge which would render the detector insensitive and could damage it, the avalanche has to be quenched somehow. This can be achieved through the slowly outwards moving ion charge cloud itself, which starts to shield the fields. Alternatively, external quenching can be used by placing a resistor in series with the anode (exactly as for an SiPM) or by actively turning off the voltage. Due to the saturation, in the Geiger region the energy cannot be determined anymore and it can only be detected whether or not a particle has hit the GD: The detector *counts* the incident particles (multiple hits cannot be registered, however). In addition, a dead time is introduced after a detection due to the recharge. This mode of operation is used in the GD class of Geiger-Müller counters.

- **Discharge region:** For even higher voltages, in region V a continuous discharge occurs even in the *absence* of incident radiation. A conducting plasma pipe connects anode and cathode, which leads to very high signals ($g \approx 10^8$–10^9). Glow, corona, or arc discharges can occur depending on the detector geometry, gas composition, pressure, humidity, and the resistance of the external circuit. This region is rarely used.

Apart from the voltage, the behavior of GDs depends on the design (discussed below), gas type, and pressure.

Gaseous detector types

At first glance, GIDs have nothing to do with the present chapter, which revolves around the detection of photons and not ionizing particles.
However, a GD can be used as amplification and readout stage of a photosensor by combining it with a solid or liquid photocathode or a photosensitive gas (see 4.3.2 and 4.4). Such photosensors have extremely promising properties – among others, the possibility to build very large sensors with an optical coverage close to 1, µm position-sensitivity, ns or sub-ns TTS, magnetic field immunity, high radiopurity, *and* low costs [196; pp. 498–509].
Unfortunately, all these very desirable properties so far are limited to the UV range, using CsI cathodes. Attempts to construct a VPGD – a position-sensitive gaseous photodetector sensitive in the visible range – with a QE comparable to PMTs, long lifetime and sufficient gain for spe detection have not been successful yet. [196] This is due to APs and aging through ion impacts. Nevertheless, the VPGD remains the object of fervent studies and great progress has been made in the last years. A success would constitute nothing less than a game-changer for the field of photodetection.
Thus, GDs cannot be neglected in a comprehensive study of photosensor types. Of particular interest are of course structures which allow position-sensitivity. Therefore, in the following an overview of the various types of GDs which have been used to create photosensors shall be given [37].

Photosensors which are based on a gaseous detector will be called photosensitive gaseous detectors (PGDs) henceforth.

[37] In the names of the GD types the terms “counter” and “chamber” are used interchangeably in literature.

To differentiate between the used cathode type (solid, liquid, or gaseous), also the subclasses of solid photocathode GDs (SPGDs), liquid photocathode GDs (LPGDs), and gaseous photocathode GDs (GPGDs) are introduced.
SPGDs which are sensitive in the visible spectrum – of crucial relevance here – will be denoted by the term visible light PGDs (VPGDs).

For a better understanding of the operation of these devices it is important to note that in GDs in general not the charge of the created carriers but their *induction* in the electrodes is used to create a current [515; ch. 5].
A gas gain of at least 10^5 is required for efficient single photon detection, which for many GD types is already close to the reachable limit due to sparking (see below) [196; p. 34].
The back-flow of ions which are produced in the avalanche to the cathode is problematic. [523] First, this induces space charge effects which limit the gain, counting rate and position resolution in GDs. Second, ion impacts on a solid photocathode damage the cathode and trigger APs[38], which affect the time and position reconstruction and severely limit the gain. Ion back-flow (IBF) is particularly critical for cathodes which are sensitive in the visible range, which is why the reduction of IBF has been in the focus of VPGD development. The use of microstructured anodes and cathodes (see MPGD on p. 590) – often in close vicinity – has proven to be very beneficial and has allowed to raise the gain of VPGDs to 10^5 [523].

The following GD types were identified in a thorough literature research to have been used to create photosensors so far. Their layouts are shown in fig. 4.5.

- **Ionization chamber:** In an ionization chamber no gas multiplication occurs and the charge carriers are collected on the electrodes (e.g. two parallel plates; similar to fig. 4.5a without avalanche formation) where they induce a current [521; p. 121].

- **Single wire counter (SWC)[39]:** An SWC consists of a thin wire (anode) on the axis of a cylinder (cathode) filled with gas. The field is proportional to $1/r$, which creates an avalanche near the anode. SWCs can be operates as proportional counters or as Geiger-Müller counters.

- **Multi-wire proportional chamber (MWPC)[40]:** An MWPC extends the SWC principle to allow position reconstruction by using multiple closely spaced parallel wires (fig. 4.5c). The field lines are perpendicular to the wire planes except close to the wires. The fields thus focus electrons onto the closest anode wire in whose vicinity they produce a Townsend avalanche as for the SWC. The position can then be reconstructed from the wire which produces an output pulse (resolution limited by the wire spacing) and in the other dimension from the output pulses which leave both ends of the wire by calculating the center of gravity from the pulse amplitudes.

- **Parallel plate avalanche chamber (PPAC)[41]:** A PPAC consists simply of two parallel metallic plates that are separated by a gap of a few mm which is filled by

[38] Through the ion-induced secondary electron yield (ISEY).
[39] Content based on [515; pp. 181–182].
[40] Content based on [521; p. 127].
[41] Content based on [196; pp. 22–25, 505].

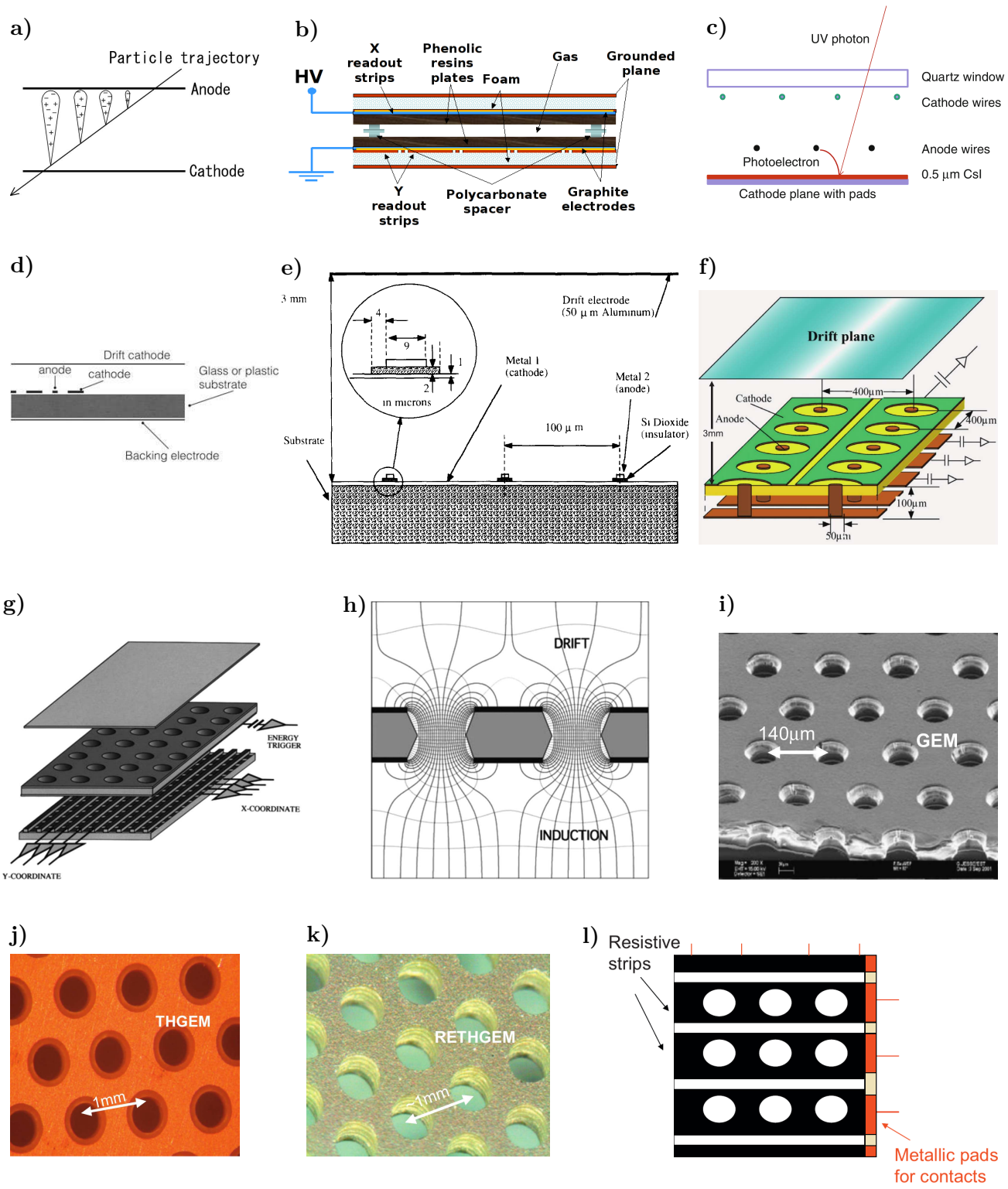


Figure 4.5: Gaseous detectors (GDs). Figure continued on next page. Discussion in text.
a)–b) Plate-based GDs: a) Parallel plate avalanche chamber (PPAC), from [524]. b) Resistive plate chamber (RPC), from [525].
c) Wire-based GDs: Photosensitive gaseous detector (PGD) based on a multi-wire proportional chamber (MWPC), from [194; p. 306].
d)–q) Micro pattern gas detectors (MPGDs):
d)–f) Surface-amplification GDs: d) Micro strip gas chamber (MSGC), from [521; p. 128]. e) Micro gap chamber (MGC), from [526]. f) Micro Pixel Chamber (μ-PIC), from [527].
g)–n) Hole-type gaseous detectors (HTGDs):
g)–i) Gas electron multiplier (GEM): g) Layout, from [528]. h) Field distribution, from [528]. i) Photo, from [529; p. 117].
j) Thick GEM (TGEM), from [529].
k) Resistive electrode TGEM (RETGEM), from [529].
l) Strip RETGEM (S-RETGEM), from [530]. View from bottom, the metallic strips are beneath the resistive strips.

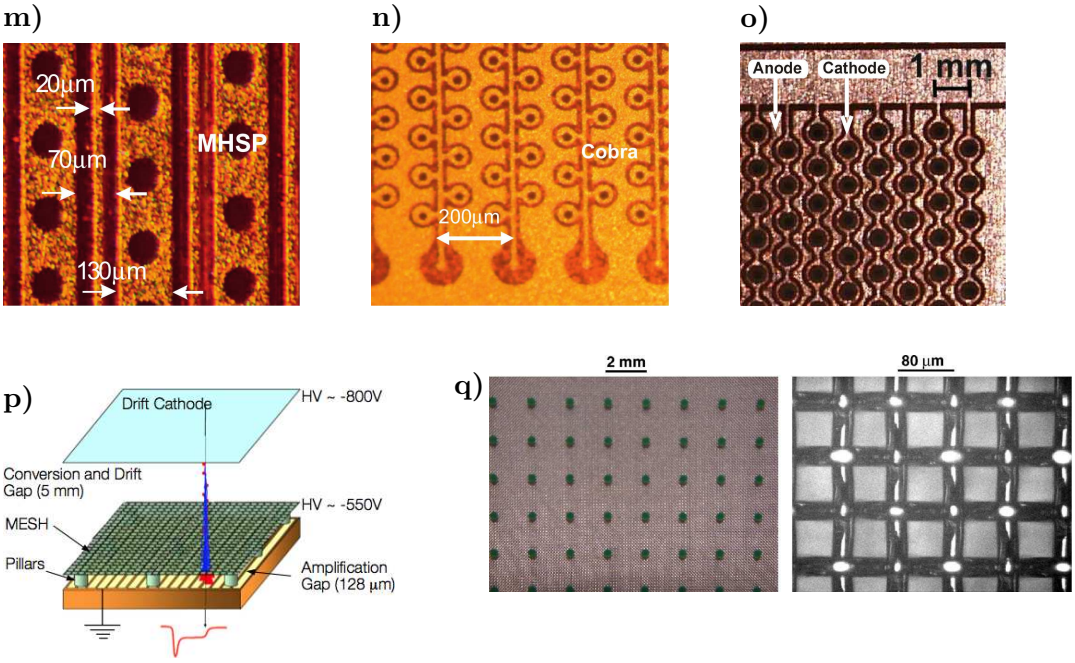


Figure 4.5: Gaseous detectors. Continuation from previous page. Discussed in text.
g)–n) Hole-type gaseous detectors (HTGDs):
m) Micro-hole and strip plate (MHSP), from [529].
n) COBRA, from [529].
o) Thick COBRA (THCOBRA), from [531].
p)–q) Mesh-type MPGDs: Micro mesh gaseous structure (Micromegas): p) Layout, from [532].
q) Photo of bulk Micromegas, from [533].

a gas (fig. 4.5a). The linear field between the electrodes needs to be strong enough to trigger avalanche production. For the electrodes instead of plates also meshes can be used, which allows to use the PPAC as an amplification structure for a detector that consists of multiple GD types. Micromegas (see p. 596) are a PPAC subtype. A position measurement can be realized by using multi-strip cathodes. These are either connected to a resistor array (position calculated from the ratio of output charges, see p. 604) or delay lines (position obtained from the time difference of the output signals, p. 604) [524]. The gas amplification can exceed 10^5 and the spe TTS (σ_T) [42] can reach 0.55 ns (CsI cathode with 20 Torr CH_4).

- **Resistive plate chamber (RPC)** [43]**:** The RPC is basically a PPAC whose graphite plate electrodes are covered with a high-resistivity material, e.g. glass or a resin (bakelite), and kept separated by spacers (fig. 4.5b). The movement of the charges induces a current in the readout electrodes, which are grounded and placed behind the graphite electrodes. This is possible due to the low conductivity of graphite. The readout anodes can be segmented (pads or crossed strips) for position readout. The discharge is limited by the high resistance and leads to a charge buildup on the surfaces, which creates an opposing field and quenches the avalanche. The TTS of an RPC is in the ns-regime and can be as low as 100 ps [521; p. 128]. Single photon 2D position resolutions down to about 30 µm can be reached [196; p. 501].

[42] Standard deviation of the main peak in the transit time distribution.
[43] Content based on [515; pp. 208–214] [194; p. 485] [521; p. 128]

- **Micro pattern gas detector (MPGD):** MPGDs are based on microscopic lithographic structures, mostly holes, meshes or strips. This allows a better temporal (ns range) and spatial resolution ($\approx 30\,\mu m$) than with MWPCs. [194; pp. 247–249] [534].

  - **Surface-amplification GDs:** In this MPGD class, the avalanche is created in the strong fields in the vicinity of strip- or circular-shaped anodes placed on a surface. These anodes are also used to collect the produced electrons. For strip anodes, this layout is basically similar to an MWPC where the wires are placed on a surface. In addition, cathode strips are located between the anode strips to avoid the formation of a uniform equipotential surface. This GD type is sensitive to spark production due to the high voltages needed to produce large gains and the small spacing of anodes and cathodes.

    ▷ **Micro strip gas chamber (MSGC) [44]:** The MSGC (fig. 4.5d), invented by Oed in 1988, was the first MPGD. It consists of a gas layer of a few mm thickness between a cathode plane and an insulating substrate on which thin anode strips are placed. Like wires, the strips lead to high field gradients in their vicinity and induce a Townsend avalanche. To quickly remove the produced ions, cathode strips are located on the substrate between the anode strips. These attract the majority of ions and reduce the amount of ions which slowly drift to the cathode plane. The space charge of the latter would affect the electric field and limit the count rates. Gains of over 10^5 and position resolutions of $30\,\mu m$ (10–20 times better than for MWPCs) were achieved. Furthermore, 2D position reconstruction is possible with additional readout strips on the back of the substrate or when using readout pads instead of strips. However, uncontrolled discharges caused by highly ionizing particles (e.g. Au nuclei knocked out of an electrode) can damage the electrodes. Since this problem could not be solved, it triggered the development of the other MPGD types where the gain is either split up on several steps or sharp edges are avoided.

    ▷ **Micro gap chamber (MGC):** The MGC (fig. 4.5e), introduced by [526], advances the MSGC design by coating the complete substrate with a conductive cathode and placing the thin anode strips on top of insulating strips on the cathode. This further improves timing and the rate capability due to the smaller distance of anode and cathode. Likely, this layout suffers from the same discharge problems as MSGCs, since the field gradient here is even stronger, and a cursory literature search revealed no publications after 2000.

    ▷ **Micro pixel chamber (μ-PIC) [45]:** The μ-PIC (fig. 4.5f), invented in 2001 by [535] [46], consists of a $100\,\mu m$ thick printed circuit board (PCB) onto which anodes and cathodes are etched with photolithography. A pixel consists of a circular anode pillar ($50\,\mu m$ diameter) within an electrode-free circle of $250\,\mu m$ diameter, which is surrounding by a cathode strip. The pixel pitch (spacing) is $400\,\mu m$ in both coordinates. Photoelectrons drift toward the anode dot and create a Townsend avalanche in the strong fields between anode and cathode strip. The anode dots are connected by strips perpendicular to the cathode

[44] Content based on [515; pp. 226–229] [194; p. 247].
[45] Content based on [527, 535].
[46] According to [536].

strips, which allows 2D position reconstruction. Gas multiplications of 10^4 were achieved.
While the other surface-amplification GDs were early designs and have been substituted by other types of MPGDs (see below), the μ-PIC was invented *after* the advent of GEMs and Micromegas. The idea was to avoid the floating structures (wires, meshes, foils) which are required in the other types, to allow to construct large area detectors. For anode dots lower voltage differences to the cathode strip than for anode strips are sufficient for high gains. This reduces the risk of discharges from the cathode edges and the energy of a potential spark is reduced due to the smaller anode area. Nevertheless, for highly ionizing particles discharges still can occur. To reduce the damage by sparks, µ-PICs with resistive cathodes (compare RPC and RETGEM) have been developed using sputtered carbon. [537] These also achieve gains of up to 10^4.

– **Hole-type gaseous detector (HTGD)** [47]**:** HTGDs are based on dielectric sheets which are metalized on both surfaces and in which an array of holes is chemically etched or mechanically drilled (see fig. 4.5g). When a voltage is applied between both surfaces, this creates high field-gradients in and near the holes, which focuses the electrons into the holes and induces the formation of an avalanche (fig. 4.5h). The produced electrons are emitted on the other side of the holes, which allows to stack multiplication stages and increase the total gain (see e.g. figs. 4.16a and 4.16b) [196; pp. 36–38]. Furthermore, HTGDs efficiently suppress IBF and photon feedback, which allows to operate solid photocathodes at higher gas gains than reachable with MWPCs and PPACs. In addition, HTGDs can be constructed from low-background materials. These favorable properties explain the high popularity of these designs and the many subtypes which have been developed. HTGDs are mainly amplification structures and are often not suited for readout, which then has to be carried out by a subsequent stage. This allows to optimize the readout stage for its purpose. Ideally, non-amplifying readout electrodes are used to spatially separate the regions of high field strengths from the readout and thus reduce the influence of discharges [515; p. 230].

  ▷ **Capillary plates (CPs)** [48]**:** Glass capillary plates are very similar to MCPs (cf. fig. 4.3g) and are in fact available as a by-product of MCP production. However, CPs are thicker and the holes have larger diameters (10 µm to 1 mm). Furthermore, while MCPs operate in a high vacuum and the gain results from the SEY of a material covering the inner tube walls, CPs work in a gas at atmospheric pressure and here the strong field inside the holes triggers a Townsend avalanche. CPs are thus metalized on both surfaces to apply the voltage, but not inside the holes.
  To create a GPGD, a CP can be combined with a drift space where a weak field is applied between a cathode (drift plane) and the top CP electrode. Photoelectrons which are created by PI in the drift space then are focused into the holes where they are amplified. The photon interaction volume can be extended by increasing the drift space, which, however, also raises the TTS due to drift time variations.

[47] Content based on [538; p. 6] [196; p. 27].
[48] Content based on [196; pp. 27–28] [538; pp. 6–7].

By placing strip or pad readout electrodes beneath the CP, one can determine the hole in which the avalanche occurred and thus deduce the photon interaction position.
Since CPs are quite expensive, they were mostly used in early HTGD designs and the focus later shifted to the development of GEMs.

▷ **Gas electron multiplier (GEM)** [49]**:** The GEM, introduced by Sauli in 1997 [50], consists of a thin insulating foil (typically 50 µm thick Kapton) which is coated with Cu on both sides and is chemically perforated with holes with diameters of 25–150 µm and a pitch of 50–200 µm (see fig. 4.5g and 4.5i). This comparatively simple structure allows the production of detectors with large areas and varying shapes. When a voltage of ca. 400 V is applied between the Cu layers, this produces a strong field gradient of 50–100 kV/cm in the holes (fig. 4.5h) and entering photoelectrons start an avalanche. Thus, each hole acts as an individual proportional counter. Most avalanche electrons are transferred into the region below the GEM – the GEM is a pure gas amplification structure with no readout. This allows to cascade multiple amplification stages (triple-GEMs are very common) to achieve high overall gas gains above 10^5 while strongly reducing the risk of discharges, since each individual stage can be operated with lower voltages than for one-stage amplification structures. In addition, the photon and ion feedback can be strongly suppressed with suitable fields and sensor geometries. Since only the electrons are used to induce a current in the readout electrodes, the signal contains no ion tail and is only a few tens of ns long [51]. To increase the induced current, a higher field is used between GEM and readout stage than between drift cathode (top plane in fig. 4.5g) and GEM. For the same reason, the distance between the last GEM and the readout anodes should be minimized. The readout stage consists of grounded electrodes on a PCB and is fully decoupled from the amplification stages, which allows to vary the readout layout and protects subsequent electronics. An spe σ_T of 2.1 ns was reached for a GEM with a CsI cathode, 140 µm pitch and $1.2 \cdot 10^6$ gain [196; pp. 505, 508]. With 2D electrodes GEMs can achieve 2D spatial resolutions of 50 µm for single photons [196; p. 501] [52]. However, at high gains discharges which damage the device can still occur. Therefore, the gain of each stage should remain below 100 and during assembly dust free conditions are required.

◇ **Micro-blasting glass plate (MB-GP)** [53]**:** The MB-GP (also called glass GEM) is a subtype of the GEM and consists of a 300 µm thick Pyrex glass sheet, where the holes are created through sandblasting (outer diameter 126 µm, hole pitch 230 µm). This design thus lies in between GEM and CP. Gains of up to about 10^3 could be achieved but considerable IBF was observed.

[49] Content based on [194; pp. 249–250] [515; pp. 229–231] [538].
[50] It was originally conceived as a preamplifier structure for an MSGC to solve the problem of discharges.
[51] In other GD types the ion tail can have a length of many µs long due to the slow ion drift velocity, e.g. typically 100 µs for an MWPC [196; p. 16].
[52] This sensor used a transmissive (on the window) / reflective (on the GEM surface) CsI cathode and a CMOS image sensor for readout and is thus a hybrid sensor [539].
[53] Content based on [196; p. 261].

◇ **Thick GEM (TGEM or THGEM)** [54]**:** When it was found that the maximum achievable gain of GEMs increases with the foil thickness, this led to the development of thick GEMs. A TGEM has a significantly higher thickness than GEMs (typical 0.4–1 mm, but up to 2–3 mm) and larger holes with diameters of 0.3–1 mm at a pitch of 0.7–1.2 mm (fig. 4.5j). Regarding the dimensions, a TGEM thus is closer to a CP than a GEM. Also, the manufacturing process differs from the GEM and is far simpler: A 0.4–3 mm thick PCB is metalized on both sides and then holes are drilled mechanically into it. Therefore, TGEMs can be produced with large areas up to 1 m^2 and the cost per area is significantly lower than for CPs or GEMs. Often a circular rim around the holes is etched away photolithographically to remove sharp metallic edges from the drilling, which could cause sparking. A small or absent rim improves the gain stability, while large rims allow high maximum gains and reduce the probability of discharges but lead to a significant charge-up when switching the device on. TGEMs have significantly higher operating voltages than GEMs (like CPs) but this is rewarded with a 10 times higher maximum gain. Due to the larger holes, TGEMs are less sensitive to dust and micro-particles in the holes and do not require clean-room conditions during production. They can also withstand sparking better. Furthermore, the CE on subsequent stages is higher than for GEMs, since the hole diameter is larger than the electron diffusion range. However, the larger hole pitch also reduces the position resolution. Gas gains of 10^5 and 10^7 can be reached for single and cascaded double TGEMs (also see [531]), respectively, which allows single photon detection. Stable operation at photon fluxes above 1 MHz/mm^2 was reported with sub-mm position resolution and TTS in the 10 ns range. An spe σ_T of 8 ns was reached for a CsI cathode, Ar +5% CH_4 at 1 atm and 0.7 mm pitch; the TTS is higher than for GEMs due to the larger pitch [196; pp. 505, 507]).

◇ **Resistive electrode TGEM (RETGEM or RETHGEM)** [55]**:** A RETGEM (see fig. 4.5k) is a spark-protected version of the TGEM with resistive electrodes in place of conductive Cu electrodes (compare RPC and μ-PIC). To form resistive electrodes, sheets of 50 µm thick carbon-loaded Kapton or printed resistive surfaces are attached to both surfaces of the PCB. The holes have a diameter of 0.3 mm, a pitch of 0.6 mm, and are mechanically drilled. When a voltage is applied to the resistive electrodes, they charge up due to the finite resistance and form equipotential surfaces. This creates the same fields as metallic electrodes and the GD operates like a regular GEM. But at high rates and during a discharge, the current in the spark is limited by the electrode resistance (spark energy 100–1000 times lower) and the behavior resembles a RPC. The RETGEM is thus a hybrid between GEM and RPC. In improved designs, the electrodes are divided into resistive strips with underlying metallic readout strips. The latter pick up the induced signals and are separated from the resistive electrodes by thin dielectric layers of 50–200 µm thickness to decouple them.

[54] Content based on [196; pp. 33–34] [194; pp. 251–252] [538].
[55] Content based on [194; p. 252] [196; pp. 34–35] [538].

- ◇ **Strip RETGEM (S-RETGEM)** [56]**:** S-RETGEMs contain double-layered strip electrodes which – unlike for GEMs, TGEMs and RETGEMs – are also used for *readout* (fig. 4.5l). First, an insulating G-10 plate is lithographically covered with Cu strips with circular openings (width 0.7–1 mm, pitch 1–1.4 mm, circle diameter 0.5–0.8 mm). The surface of the strips is then coated with a 15 µm thick resistive layer. Afterwards, holes are drilled into the double-layered strips, where the hole diameters are smaller than the openings in the Cu strips to create rims. This structure is used on the bottom of the HTGD stage. The electrons which are created in the holes induce a signal on the metallic readout strips, which allow 2D position reconstruction and are spark protected by the resistive layer as in a RETGEM. This GD type is therefore a mixture of MHSP (see below) and RETGEM. In the original layout, the top side was even coated with a UV-sensitive CsI *photocathode*. S-RETGEMs have been produced with areas up to 200 cm^2 and gains of 10^5 and 10^6 were reached with one and two layers of S-RETGEMs, respectively, which enabled single photon detection in the UV with even one stage.

- ▷ **Micro-hole and strip plate (MHSP)** [57]**:** An MHSP (fig. 4.5m) is a hybrid amplification structure which uses the principles of both GEM and MSGC. A GEM-like hole structure and top electrode layout is combined with thin anode and broad cathode strips (around the holes) on the bottom electrode as for an MSGC. This is nothing other than a GEM with a different electrode pattern on the bottom and is similar to the S-RETGEM. In the MHSP, avalanche multiplication occurs in the holes *and* additionally at the anode strips, and a large part of the ions is collected on the cathode strips. Since the electrons are collected on the anode strips, no additional readout stage is needed.
  A breakthrough regarding the IBF reduction was achieved through the use of cascaded GEMs and MHSPs, where for the latter the orientation and voltage polarity was varied. In a flipped reversed MHSP (F-R-MHSP, fig. 4.16a) the polarity is changed (reversed bias; cathode and anode strips change roles), and the strip-covered side is on the top, facing the photocathode, instead of the bottom (flipped). The negatively biased strips then efficiently focus the photoelectrons into the holes and capture most of the ions which drift back though the holes. This allows to reach higher gas gains. For a VPGD with a triple structure of F-R-MHSP, GEM and MHSP (shown in fig. 4.16a) a gain of 10^5 was reached before feedback effects became visible and a CE of nearly 100% was observed.

- ▷ **COBRA** [58]**:** The COBRA (fig. 4.5n) was derived from the MHSP and uses the holes (60 µm diameter) and bottom side layout of a GEM, whereas the top side consists of zigzag-shaped patterned electrodes (similar to the F-R-MHSP). Circular, connected anodes (30 µm wide) are placed around the holes and meandering cathode strips are situated in between the anode strips to attract ions and reduce the IBF. A cascade of a COBRA and 2 GEMs

[56] Content based on [196; pp. 196, 199] [530].
[57] Content based on [194; p. 250] [196; pp. 255–257].
[58] Content based on [523].

produced IBF values of as low as $3 \cdot 10^{-6}$, which is 10^4 times lower than for triple GEMs. However, the CE amounts to only 20%. Gains of above 10^3 were reached for a single stage [540].

▷ **Thick COBRA (THCOBRA)** [59]**:** The THCOBRA is a combination of COBRA and TGEM (fig. 4.5o). Similar to a TGEM, holes with 300 µm diameter are drilled in a 400 µm thin insulating G-10 plate which is covered with Cu on both sides, and 100 µm thick rims are etched around each hole to reduce the discharge probability. On one side, the metal electrode is then etched to create zig-zag anode strips (300 µm width) between round connected cathode strips (100 µm wide) as for the COBRA, but with 5 times larger dimensions. On the other side, either a regular GEM pattern or also a COBRA pattern is used [540]. Depending on polarity, the anode strips either create secondary avalanches (primary avalanche in the hole) or they attract ions and suppress the IBF. Gains of 10^5 for UV photons were achieved.

– **Micro mesh gaseous structure (Micromegas)** [60]**:** The Micromegas was introduced in 1994 by Charpak and is a PPAC with a micromesh cathode and a small amplification gap (fig. 4.5p). Micromeshes are thin metal grids (15–20 µm openings, pitch about 25 µm) and are a standard component of TV screens. They are thus commercially available in various sizes and shapes. The Micromegas consists of a several mm thick drift region (between drift cathode [61] and mesh) and a thin amplification gap of 25–150 µm thickness between the micromesh and readout electrodes (strips or pads printed on an insulator). Insulating support pillars provide a uniform gap distance between the anode plane and micromesh at the cost of a small efficiency loss due to the fill factor (about 1% [534; p. 30]). Because of the small dimensions a comparatively low voltage of 400–800 V is sufficient for operation. The fields in both the drift region (low field of ca. 1 kV/cm) and the amplification region (high field of about 40–80 kV/cm) are homogeneous, and only near the mesh openings deviations occur in form of funnel-shaped field lines. Electrons which are produced through ionization or photoeffect drift through the mesh holes into the multiplication region where they are amplified as in a PPAC. With the proper choice of voltage, most positive ions are collected by the micromesh, which prevents a space-charge accumulation and leads to a fast signal with a short ion tail (50–100 ns length). The gas gain can reach several 10^4 but is lower than in PPACs, which can achieve above 10^5. Due to the small amplification gap, excellent spatial resolutions of down to 12 µm (limited by the mesh pitch) and spe time resolutions down to 0.68 ns σ_T [541] were obtained.
As for GEMs, various subtypes exist [194; p. 249] [534], of which only two shall be addressed briefly:

▷ **Bulk Micromegas** [62]**:** For bulk Micromegas (fig. 4.5q) the amplification region is produced in a single piece. A PCB is first covered with 50–150 µm photoresistive film, then with a woven mesh (19 µm mesh diameter), and finally with another layer of 50–150 µm photoresistive film. The pillars are created photolithographically, extend from the PCB through the mesh, and

[59] Content based on [531].
[60] Content based on [194; pp. 250–251] [196; pp. 24–25, 505].
[61] The drift cathode can also be a mesh.
[62] Content based on [194; p. 251] [542; p. 14].

have a diameter of 200–400 µm and a pitch of 2–4 mm. Bulk Micromegas are robust, inexpensive and allow to produce large GDs.

▷ **Microbulk Micromegas [63]:** In Microbulk Micromegas a thin flexible Kapton foil, which is covered with a Cu layer on both sides, is used to create both the wire structures and the pillars through etching. The pillars lie below wire crossings, either under every crossing or with a pitch of 500 µm (see figs. 1 and 2 in [543]). This procedure results in improved properties such as a flexible structure, a small mass to active area ratio, and a longer lifetime. Furthermore, a high radiopurity can be achieved and the uniformity of the amplification gap width can be well controlled through the thickness of the Kapton film. However, the structures are delicate.

The main GD types which are in use today comprise MWPC, PPAC (including the important subtype Micromegas), RPC, and HTGDs [196].

Photon counting

To allow photon counting, a sensor needs to have a defined single photon (spe) peak in the SER which must be distinguishable from noise, i.e. the pedestal. Furthermore, the peaks of n simultaneously detected photons (npe peaks) have to be narrow enough to be able to reconstruct the number of photons with high reliability.
Whether these prerequisites are fulfilled, depends on the interplay of three factors: the sensor gain, the charge resolution and the noise of the readout electronics or preamplifier, if one is used.
To separate the spe peak from the pedestal, either a high gain or a low noise RMS (root mean square) is required. Thus, a small gain can be compensated with a low-noise preamplifier, and the required gain for photon counting can be smaller than for PMTs (about $5 \cdot 10^6$, see p. 132). However, the smaller the sensor gain is, the closer the preamplifier has to be placed to the sensor output to reduce the input capacitance (eq. (4.1)), since otherwise the charge resolution is smeared. For example, in HPDs (gain a few 10^3) often preamplifiers are placed *inside* the vacuum tube, while HAPDs have higher gains (around 10^5) and so external amplifiers coupled directly to the sensor are sufficient. For PMTs (gain around 10^7), often no preamplifier is needed at all.
A small charge resolution reduces the overlap between spe peak and pedestal (more spe events lie above the threshold) and allows to reliably distinguish the number of detected photons. The noise RMS is also connected to the charge resolution, since the SER is the result of the convolution of the sensor response with the electronics response (see caption of fig. 2.21). The amplifier noise can even dominate the charge resolution for very low sensor gains.

In the following, first the Fano factor and the equivalent noise charge (ENC) are introduced, which determine the dependence of the charge resolution on the EB voltage and the amplifier noise. Afterwards, the gains which are needed for photon counting are discussed exemplarily for important sensor classes.

[63] Content based on [194; p. 251] [543] [544; p. 12].

Fano factor

For many materials, the Fano factor F_F – which describes the conversion of energy to secondary particles[64] – is significantly smaller than 1. This includes semiconductors (e.g. Si, Ge and GaAs) and gases [195; p. 118] [194; p. 440]. When a photoelectron is amplified by EB[65], the charge resolution R_F after the first gain step reads[66] [196; p. 97]

$$R_F = 2.35 \cdot \sqrt{F_F \frac{W}{E}} = 2.35 \cdot \sqrt{\frac{F_F}{N}} \tag{4.2}$$

where W is the mean energy required to create charge carriers (electron–hole pair, ion pair), E is the kinetic energy. Thus, E/W is the number N of created secondary carrier pairs.
Therefore, a small F_F improves the charge resolution.

Equivalent noise charge

The noise of an amplifier is often quoted as the equivalent noise charge (ENC). This is the ratio of the noise output voltage U_n (unit V) and the output voltage U_e which results from a single electron as input charge (unit V/e$^-$):

$$\mathrm{ENC} = \frac{U_n}{U_e} \tag{4.3}$$

The ENC thus has the unit "electrons" (e$^-$) [515; p. 802] and is equivalent to the input charge (i.e. signal) which is required to obtain a signal-to-noise ratio (S/N) of 1 [194; pp. 39–40]. It indicates the achievable signal resolution after amplification. Many photosensors use internal amplifiers, whose ENC usually is quoted as e$^-$ RMS. This can be directly compared to the charge resolution R_σ listed for other photosensors.
The S/N can be obtained by dividing the spe charge q_1 by the ENC.

Hybrid photodiodes

For silicon, at large impact energies $W \approx 3.6\,\mathrm{eV}$, $F_F = 0.1$ [194; p. 30] and the energy losses in the insensitive entry layer of a Si detector typically range between several hundred eV and 1 keV [153; p. 299] [545; p. 139]. Typical EB impact energies for hybrid photosensors with Si diodes lie between 8 and 20 keV [194; p. 303] [169; p. 231] and thus between 1500 and 6500 electron–hole pairs are created [545; p. 139] [153; p. 298] [169; p. 231]. These numbers would result in a charge resolution of only 0.9–1.9%[67], as was already pointed out by [153; p. 299]: Since only a single amplification step with high gain is used, the statistical gain fluctuations are very small [545].
In reality far larger resolutions are measured due to three effects:

- **Backscattering[68]:** Photoelectrons can be scattered out of the Si surface after depositing only a fraction of their energy (on average half the impact energy). Backscattered electrons are either lost (BL1-Tm, only the regular

[64] See p. 96 (footnote) and p. 121.
[65] Kinetic energy which was gained in a field is converted into secondary charge carriers through impact on a material.
[66] R_F can not fall below the Poisson limit, however, since the primary particle is still an UV/visible photon, which can only create one photoelectron (see p. 121).
[67] Compare this to the values for PMTs in table 3.7.
[68] Content based on [153; p. 298] [327, 496, 545].

pulse is detected and the late pulse is lost, see 2.3.5) or are refocused onto the silicon sensors and hit another pixel (crosstalk, similar to BL1). For an impact energy of 13 keV, about 18% of photoelectrons are backscattered. In the case of crosstalk, the dead layer has to be crossed again, which leads to additional energy losses. Thus, even if the charge of both pixels is summed up, the total charge is lower than without backscattering. Losses and crosstalk both reduce the detected charge and fill up the regions between the sharp npe peaks in the SER of an HPD, which increases the charge resolution.

- **Amplifier noise**[69]**:** For HPDs, the gain results only from EB and thus is very low (several 10^3), which requires to use low noise preamplifiers for LLL detection. Nevertheless, the electronic noise of these devices still significantly increases the actual resolution. A detailed treatment of amplifier noise can be found in [194; ch. 2] [153; ch. 17] [546]. If the readout electronics is included in the vacuum tube, this increases the S/N and lowers the required number of feed-throughs through the glass. However, then the electronics must be compatible with the bake-out cycles before cathode deposition and the power which is dissipated by the electronics must be removed from the sensor by a good thermal contact to the tube base.

- **Charge sharing:** For segmented sensors with small pixels ($\leq 100\,\mu m$), charge sharing between pixels can occur at the pixel boundaries, typically within a range of 10 µm. [496, 545] This also reduces the signal height and increases the resolution.

Despite these detrimental factors, the resolution of HPDs is still excellent. This has the consequence that – although their gain is only several 10^3 – HPDs *can* detect single photons. In fact, they have very good photon counting capabilities, which by far surpass those of PMTs.

Hybrid avalanche photodiodes[70]

For HAPDs, the bombarded Si sensor is an APD instead of a photodiode. The combination of EB gain and avalanche multiplication (50 to several 100) allows to achieve a higher gain (over 10^5) than for HPDs, which eliminates many of the problems mentioned above. Alternatively, a lower high voltage (HV) can be used. However, the statistics of avalanche amplification greatly increases the charge resolution and the npe peaks are therefore no longer isolated from each other.

Photosensitive gaseous detectors

For gases, the ionization potential is in the order of 30 eV [194; p. 30] and $F_F \approx 0.1$–0.2 [196; p. 162] [195]. For this reason, the resolution is not quite as good as for the EB of Si, but still better than for PMTs (see below). Usually a value of 10^5 is given as the threshold for spe detection (e.g. [196; p. 34]).

Photomultipliers

The fact that PMT dynodes follow Poisson statistics (see 2.2.2) indicates that they have a Fano factor of 1. W is significantly higher than for Si and about comparable

[69] Content based on [153; p. 299] [545; p. 139] [194; p. 303].
[70] Content based on [153; p. 300] [194; p. 304].

to the value in gases, since the SEY is only in the order of 5–20 for impact energies of several 100 eV. Therefore, due to the large resolution even for gains of 10^7 photon counting is only possible with an uncertainty. An advantage of PMTs is that at such high gains amplifiers often are not necessary.

4.1.1.5 Readout

Each photosensor requires a readout stage to collect or produce a signal. For the sensors relevant here, the output signal is generated mostly through charge collection (photodiode family, electron bombardment[71]) or current induction (GDs); compare 4.1.1.1.

Position sensitivity

The readout stage is intrinsically connected to whether or not a sensor can resolve the photon position. However, a photosensor only then is position-sensitive, if *all* transport, amplification, and readout stages preserve the location information. Three different types of sensor can be distinguished with regard to their capability to resolve the photon hit location:

- **Point sensors** can *not* determine the photon location and thus their position resolution equals the size of the sensor's active area.
  Examples are classic PMTs or SiPMs.

- **Image sensors** have a segmented readout (pads/pixels[72], strips or wires) or consist of multiple sub-sensors[73] (see p. 609) and have a *discrete* position resolution, which is related to the dimensions of their elements.
  This is the case for e.g. CCDs and CMOS[74] image sensors.

- **Position sensors** have a *continuous* position resolution. This can be achieved either with a monolithic detector and several readout channels or with a segmented image sensor with special readout methods.
  Examples comprise the position sensitive diode (PSD), mesh and metal channel PMTs, and MCP-PMTs.

The different readout techniques and their ability for position determination will be detailed in the following section (starting on p. 601). The relation between the number of readout channels and the position-sensitivity is discussed on p. 609.

Contrary to intuition, a position sensor does *not* necessarily provide more accurate position information than an image sensor. This is because the structures of the latter can be made smaller[75] than the position errors of the former. As discussed below, some continuous-position readout methods introduce significant errors (e.g. resistive network). In addition, if multiple neighboring segmented readout structures are hit,

[71] Depends on readout method; there are also many EB sensors which use induction.
[72] A small square electrode is usually called pad if it is larger than 1 mm and pixel if it is smaller [153; p. 495] [515; p. 305].
[73] Called cells or *also* pixels.
[74] Complementary metal–oxide–semiconductor
[75] Extreme example: The QIS has sub-diffraction-limit pixels.

the resulting position accuracy can be *lower* than the spacing between structures, since the errors of the centroid algorithms (see below) are reduced [515; pp. 841–850]. Furthermore, the dynamic range of position sensors often is very limited, because the position reconstruction can become ambiguous if more than one photons are detected simultaneously. This problem can also occur for strip-/wire-based image sensors if they are not designed with redundant elements to compensate this (see Cross strip anode on p. 605).

For all position-sensitive photosensors, the spatial resolution and crosstalk to adjacent channels are crucial properties [169; pp. 191–192]. However, sometimes crosstalk is desired and included by design to reduce the spatial resolution. For most position-sensitive sensors also ion back-flow and photon feedback are of critical importance.

As already discussed on p. 572, in vacuum tube-based sensors some focusing types are more favorable for a good spatial resolution than others. Proximity focusing achieves good position resolutions due to the small distances and high fields which are used. Cross-focusing produces good spatial resolutions, as well, which is due to the compensation of the spatial spread. Fountain focusing, on the other hand, only partially preserves the position information and can not compete with the other focusing methods with regard to position resolution.

Readout structures

In the following, an overview of the multitude of existing readout structures and processes will be given. The layout of the discussed methods is shown in fig. 4.6.
Often the readout consists of several layers: one for the extraction of the information carriers and subsequent layers to combine readout channels and process the signals. The extent to which position determination is possible depends on the combination of all layers.
While a segmented readout may be present, it is not necessarily used for imaging. This is the case in the streak camera, which trades position resolution for superior TTS. Also, some digital SiPMs (dSiPMs) read out each pixel individually but sum up the output without retaining the pixel position. Due to the small device dimensions, the information which dSiPM was hit here can be sufficient for position reconstruction.

The algorithms for position calculation which will be discussed in the following are either based on:

a) The **centroid** of output pulses/currents. This is also called center of gravity, charge division or Anger method.

b) The transit **time difference** of output pulses (delay line). In this case, the pulse amplitude is *ignored* and not used for weighting as in case of the centroid.

In the following discussion, readout structures will be grouped by the photosensor stage which is used for readout (anode, cathode, dynode).
The position resolution of each method is indicated in its title by an abbreviation. This contains the reconstructible number x of dimensions (xD), where 0D means that the only available information is which sensor was hit (point sensor). This is

complemented by whether a continuous position information is available: The suffix i is used to label discrete positions (image sensor, position resolution limited by segmentation; e.g. a CCD is 2Di). The suffix p shows the capacity to determine continuous positions (position sensor; e.g. a PSD is 2Dp). Furthermore, the number y of required output channels is stated (yCh). Here the value x indicates a number equal to all readout segments (pixels, strips, etc.).

Anode

- **Unsegmented**
  - **Single readout channel – 0D, 1Ch**
    - ▷ **Wire anode**
      The anode is a single wire which is bent to a spiral.
      This is employed in gas phototubes which are used as UV flame sensors.
    - ▷ **Reflex anode**
      The SEs which are emitted ("reflected") from the last dynode are collected on an anode mesh in front of it which lies between the last and second last dynode (see p. 45 and figs. 2.15a and 2.1).
      This readout type is used in regular PMTs.
    - ▷ **Bulk anode**
      The anode is a plane surface (Micro PMT, position-insensitive MCP-PMTs; figs. 4.3e and 4.11a) or a bulk material (e.g. PDs, APD, SPAD; figs. 4.9a, 4.9c).
  - **Resistive charge division [76] – 2Dp, 4Ch**
    If a monolithic photosensor is covered with a uniform resistive layer where readout electrodes are placed on opposite ends, a continuous 1D (2 output pins) or 2D position (4 output pins) can be obtained from the output pulses/currents with a centroid algorithm (see figs. 4.9b and 4.6g). This is used in the PSD (fig. 4.9b), PSAPD [77] and HPSAPD [78] (figs. 4.14f and 4.14g). The current which is created in the sensor [79] is split up linearly on the opposing outputs according to their distance to the incidence position. Thus, the 2D position can be calculated as the center of gravity of the output signals.
    If the readout pins are at the corners of a rectangular sensor with lengths L_x and L_y in x and y direction (see fig. 4.6g), the coordinate origin is placed in the center of the rectangle, and the output currents are X_1 (bottom left), X_2 (top right), Y_1 (bottom right) and Y_2 (top left), the 2D position can be calculated by:

$$x = \frac{(X_2 + Y_1) - (X_1 + Y_2)}{X_1 + X_2 + Y_1 + Y_2} \cdot \frac{L_x}{2} \tag{4.4a}$$

$$y = \frac{(X_2 + Y_2) - (X_1 + Y_1)}{X_1 + X_2 + Y_1 + Y_2} \cdot \frac{L_y}{2} \tag{4.4b}$$

[76] Content based on [477; ch. 2, pp. 16, 17] [194; pp. 49–52]
[77] Position sensitive APD
[78] Hybrid position sensitive APD
[79] For PSD photocurrent, for PSAPD in addition avalanche, for HPSAPD from EB and avalanche.

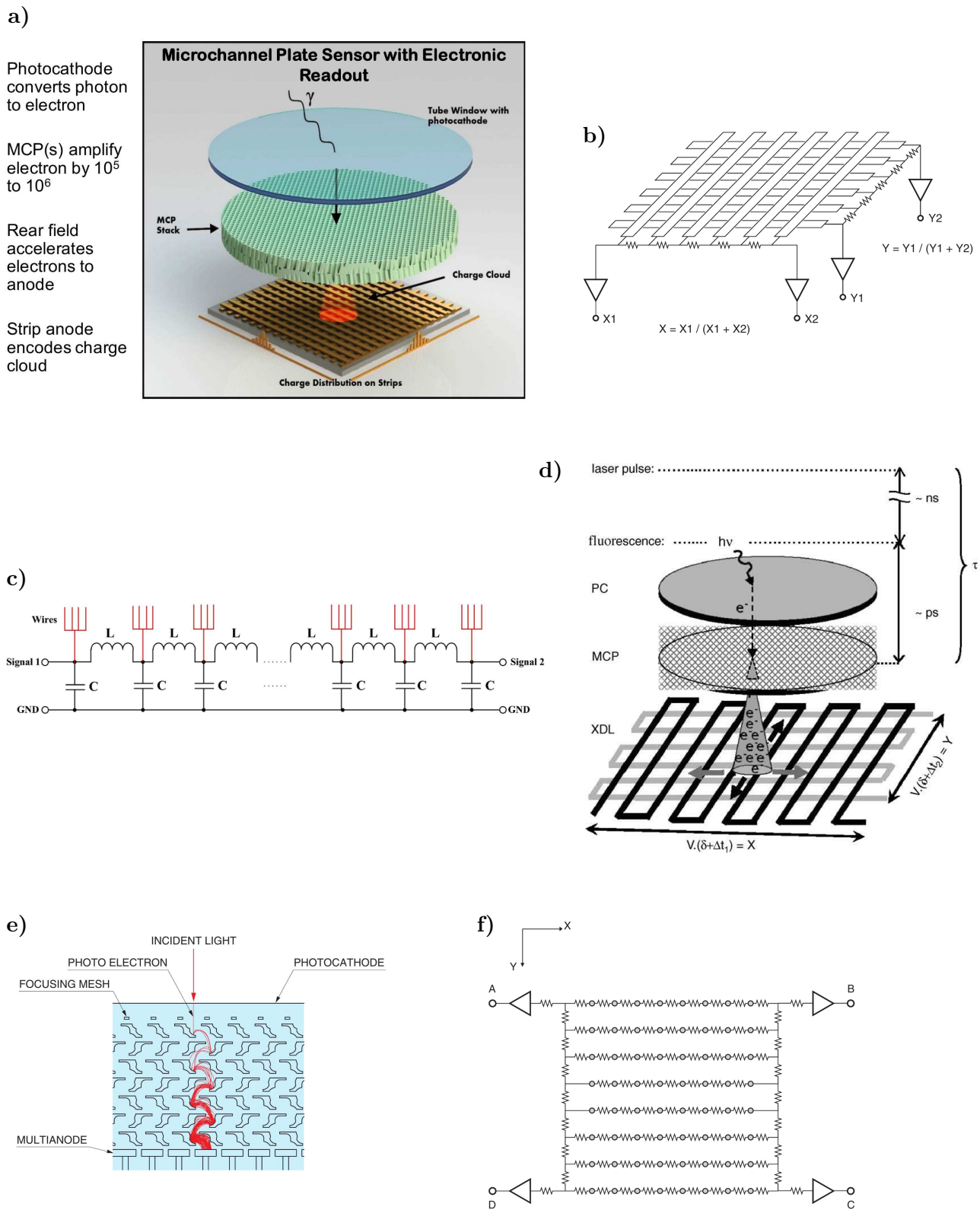


Figure 4.6: Readout types. Continued on next page. Discussion in text.
a) Cross strip (XS) for an MCP-PMT, from [547].
b) Cross strip with resistor chains, from [169; p. 200].
c) Delay line readout for a PPAC implemented through coils and capacitors, from [548].
d) Cross delay line (XDL) for an MCP-PMT, from [549].
e) Matrix anode for a metal channel PMT, from [169; p. 190].
f) Matrix anode with resistor network, from [169; p. 196].

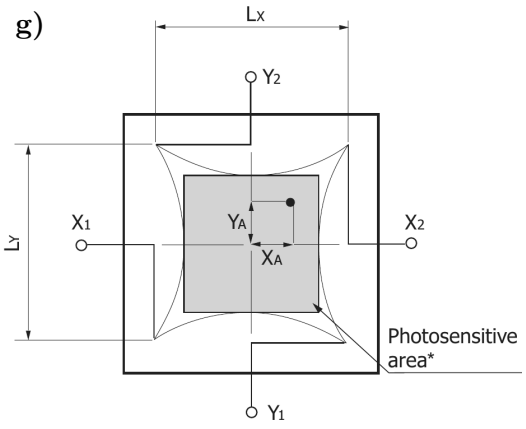


Figure 4.6: Readout types. Continuation from previous page. Discussed in text.
g) Charge division on a resistive surface, for a position sensitive diode (PSD) with 2D readout, from [477; ch. 2, p. 17].

- **Strips/wires**

  - **Strip anode** [80]
    Here, the anode consists of evenly spaced parallel strips or wires. This is the case e.g. for the MWPC, PPAC, MSGC, MGC, S-RETGEM, MHSP, and THCOBRA variants (see fig. 4.5) and the LAPPD [81] [497] (fig. 4.11c).
    This can be used to obtain a 1D position perpendicular to the strips, and even the 2D location (in addition, position on a strip):

    ▷ **Charge division for all strips – 1Dp, 2Ch**
    The electrode strips are connected to resistor chains (as fig. 4.6b but with strips only in one direction) and the position is obtained from the ratio of charges propagating to each end. The centroid algorithm allows to reconstruct the position even if multiple strips are hit, and with good algorithms in this case the resolution can actually be better than with a single strip.

    ▷ **Delay line – 1Dp, 2Ch**
    Here, the anodes are connected to delay lines (fig. 4.6c), and the time difference between the arrival of the signals at each end is used to reconstruct the position. Delay lines for PPACs have been constructed by winding thin Cu wires on bakelite rods (equivalent to a coil inductance) which are intermittently connected to the strips as well as the ground through capacitors.

    ▷ **Individual readout**
    Each wire is read out separately with suitable front-end electronics.

    ◇ **Single readout – 1Di, xCh**
    By reading out each strip, it can be reconstructed which strip was hit. This provides a discrete 1D position – that of the strip.

[80] Reconstruction methods from [515; pp. 220–223] [524; pp. 718, 719].
[81] Large Area Picosecond PhotoDetector

◇ **Single readout + electronic centroid – 1Dp, xCh**
The charge division method can be applied by subsequent signal processing, which allows to obtain a continuous position information.

◇ **Double readout – 2Dp, 2xCh**

a) Charge division: The anode strips are read out from both sides and the position along the wire can be reconstructed from comparison of the collected charges.

b) Delay line: Alternatively, the difference of the arrival times on both sides of the strip can also be used for position reconstruction along the strip. However, this requires ps timing for mm resolutions and places high demands on the used electronics. The LAPPD makes use of this method, since the whole sensor is designed for ps timing anyway, but the achievable position resolution of 3.2 mm is worse than in the direction across strips where charge division is used (0.76 mm) [550] [82]. Other delay line methods instead use a lever (such as coiled wires or serpentine electrodes) to increase the delays to more easily measurable values (see "Delay line" and "Cross delay line anode").

If (electronic) charge division is also performed for the complete layer of strips, a continuous 2D position resolution can be extracted. A drawback of the double readout method is that it requires twice the number of electronic channels compared to single readout.

– **Cross strip anode (XS)**
In this case two orthogonal layers of evenly spaced parallel anode strips or wires [83] are used to reconstruct the 2D position and time (fig. 4.6a) [551] [184; p. 182]. In case of multiple hits, ambiguities can occur, which can be suppressed by use of a third layer at a different angle [515; p. 222].
The XS readout is used in some subtypes of PMT [153; pp. 300–302] and MCP-PMT (e.g. H33D [551]). It also finds use in GDs, where either both layers are anode strips (for example GEMs, fig. 4.5g) or for one layer cathode strips are used (see p. 608; e.g. for µ-PIC and MWPCs, figs. 4.5f and 4.5c).

▷ **Individual readout – 2Di, xCh**
All strips are read out separately with their own preamplifiers, which allows to correct spatial nonlinearities with the charge-sharing technique [153; p. 301]. The obtained position resolution is discrete and corresponds to the strip spacing (unless a subsequent electronic centroid is performed).

▷ **Charge division** [84] **– 2Dp, 4Ch**
In this case, the position is obtained through charge division and subsequent calculation of the centroid. Anodes with the same direction are connected by a linear resistor string leading to two outputs (fig. 4.6b). This linearly divides the signal into four output components which correspond to the center of gravity position of the pulse.

[82] This is one example where an imaging sensor using a centroid outperforms a position sensor – in this case both "sensors" are separate readout structures of the same photosensor.
[83] Then called cross-wire anodes.
[84] Content based on [169; pp. 200–202] [184; pp. 44, 167–185] [551].

When X_1 and X_2 are the output signals at both ends of the x-axis resistor chain and Y_1 and Y_2 those for the y-axis, the position can be reconstructed through (formula based on [169; p. 200])

$$x = \frac{X_1}{X_1 + X_2} \cdot \Delta x \tag{4.5a}$$

$$y = \frac{Y_1}{Y_1 + Y_2} \cdot \Delta y \tag{4.5b}$$

where Δx and Δy are the distances between the outermost wires.
To ensure a high accuracy of the position which is obtained from the centroid, several adjacent strips should be hit and the dimensions of the electron cloud should be optimized for this. The cross-section of both orthogonal strip layers should be identical for uniform collection.
In MCP-PMTs, the strips consist of metal and insulating glass layers on a ceramic substrate, and crosstalk between layers can be reduced by placing grounded strips in between. For these sensors spatial resolutions $< 10\,\mu m$ and a TTS of as low as $\approx 100\,ps$ were achieved.
Besides MCP-PMTs (H33D, LAPPD) and GDs, this method is used in mesh and metal channel PMTs, where the SEs are reflected from the last dynode and collected on two crossing layers of strip/wire anodes above it.

– **Cross delay line anode (XDL)** [85] **– 2Dp, 4Ch**
Here, two orthogonal layers of serpentine conductors (delay lines) are used (fig. 4.6d). The 2D position is obtained from the difference of the arrival times of the signals at the opposing ends of the delay lines. This has the advantage that only four contacts are required.
For MCP-PMTs, spatial resolutions of about 50 µm FWHM [86] and a TTS of ca 100 ps could be reached. However, the gain which is required for readout is 10 times higher than for XS layouts. For this reason XS allows larger count rates and substantially increases the lifetime. This comes at the cost of more internal readout channels and a more sophisticated reconstruction algorithm than for XDL [507].

- **Pixels**
If the anode is segmented into pads/pixels (fig. 4.6e) or the sensor consists of multiple independent sub-sensors (pixels or cells), the position can be obtained from the information which pixels [87] were hit.

  – **Individual readout – xCh**

    ▷ **Plain readout – 2Di**
    In this case, each pixel is read out individually, which requires a front-end chip. This type includes CMOS sensors (fig. 4.10b) and some HPD types (Timepix HPD and Belle II HAPD; fig. 4.14d).

    ▷ **Electronic centroid** [88] **– 2Dp**
    If a centroid algorithm is applied on the previously extracted signals of all channels, this can drastically improve the resolution (e.g. for MaPMTs). When

[85] Content based on [551].
[86] Full width at half maximum
[87] For simplicity, in this discussion the terms pads, pixels and cells are subsumed as “pixels”.
[88] Content based on [196; p. 279].

multiple pixels are hit, accuracies equal to or better than the pixel size can be achieved. This is also due to the fact that gain-non-uniformities can be corrected.
The disadvantages of this method are the high costs due to the large channels multiplicity and that it can limit the count rate.

▷ **Amplified summation network – 2Dp**
The additional noise produced by resistive networks (see next item) can be reduced by amplifying all signals before feeding them into a summation network [196; p. 279].

– **Summation network [89] – 2Dp, 4Ch**
The number of readout channels can be reduced to four (1D: two) by a summation network where the weighting factors are resistors or capacitors (fig. 4.6f). All pixels are connected with e.g. resistive chains, which themselves are attached to resistor strings leading to the outputs terminals (resistive network).
With the Anger method (yet another charge division algorithm) the 2D center of gravity position can be obtained (based on [169; p. 196]):

$$x = \frac{B + C}{A + B + C + D} \cdot \Delta x \tag{4.6a}$$

$$y = \frac{C + D}{A + B + C + D} \cdot \Delta y \tag{4.6b}$$

where Δx and Δy are the distances between the outermost pads and the sum signals A–D are defined in the figure. A dependence of the coordinates on the signal height is removed by dividing the respective signals by the total sum signal. The advantage of the resistive network is that it provides a simple and cheap readout (only 4 channels), large dynamic range, and allows high count rates. However, the resistors introduce noise which deteriorates the spatial resolution and position linearity.
This technique is used for example in multianode PMTs

– **Single readout channel for all pixels – 1Ch**

▷ **Sum readout – 0D**
If the output of all pixels is summed up, the position information is lost. The segmentation of the sensor is performed for other reasons, such as to avoid dead time and increase the dynamic range. This is the case for the SiPM (fig. 4.9e) and dSiPM models which use a digital sum [552].

▷ **Time-distributed readout – 2Di**
The location information can be preserved through serial readout of all pixels, although at the cost of a significantly worse time resolution due to the prolonged extraction process.
This method is employed in the CCD (fig. 4.10a).
The streak camera (fig. 4.12e) is also read out in this manner, but has no position resolution (0D): While this hybrid sensor employs a CCD for detection, this is not used to determine the incidence position, but to improve the time resolution (see p. 659).

[89] Content based on [169; p. 196] [172; p. 7-3] [196; pp. 277–282].

▷ **Digital full readout – 2Di**
A front-end electronics digitizes the signals from all pixels, which allows to use both only a single output channel and preserve the position information. This is the case for some dSiPM variants (fig. 4.9f) [552].

Pixel anodes grouped by gain

Pixel-based sensors can also be distinguished by whether and how the pixels are amplified:

- **Unamplified** [90]
  PMTs can have segmented anodes (multianodes) which either consist of oblong strips (linear array anode) or quadratic pads (matrix anodes, 4 to 256 anodes; fig. 4.6e). This is possible for mesh, metal channel, and foil dynodes, since in this case the PMT effectively consists of multiple parallel amplification channels which act like individual photosensors and preserve the position information. The reachable position resolution is limited by crosstalk between channels, mostly from the spatial broadening of the cascade, and values of 5 mm were reached. There are also prototypes of large box-and-line PMTs (Hamamatsu R7081) for which a segmented anode was implemented (Ma-R7081). [448, 449] This was achieved through a divided cathode, which was connected to the anode segments by four independent dynode chains.
  Unamplified pixels are also used in photodiode arrays (PDAs), CCDs [91], and hybrid sensors based on multianode PMTs, PDAs and CCDs [92].
- **Preamplifier**
  A preamplifier for each channel can be included in the pixel itself (CMOS sensors) or externally in the front-end electronics (many HPD types).
- **Intrinsic amplification**
  This is the case for semiconductor pixel sensors which employ impact ionization [93] or conversion gain (QIS).

Cathode [94]

For some GDs, the cathode [95] is also used for readout. The cathode can be unsegmented, divided in strips (different orientation than the anode strips/wires), or split into pads. Examples comprise the MWPC and µ-PIC (figs. 4.5c and 4.5f).

Dynode

In some sensors even a *dynode* – actually an amplification structure – is used for readout. In PMTs sometimes one of the last dynodes is read out with a second channel to extend the dynamic range. For some vacuum phototriodes (VPTs), a segmented dynode is used to obtain position-sensitivity. [172; p. 1-22] Here, the signal is measured between individual dynode segments and a common anode.

[90] Content based on [169; pp. 189–202] [196; pp. 277–282].
[91] Neglecting the conversion gain, which is spatially separated from the pixels.
[92] DS-QUASAR (see p. 614), some HPDs, IEBCCD (p. 645), and ICCD (p. 648).
[93] EMCCD, APD array, SiPM, dSiPM, SPAD array, and hybrids; see figs. 4.10c, 4.9, 4.12, and 4.15.
[94] Content based on [515; pp. 221–223].
[95] Or one of the cathodes.

Detection channels versus readout channels

The above-mentioned readout types can be grouped into whether they require one or multiple output channels.
Furthermore, sensors can be distinguished based on the multiplicity (one or more) of detection channels. The latter are defined here as parallel structures for photon conversion and/or amplification which are largely independent from each other.
To give a few examples: A CCD contains an array of conversion stages (photodiodes) without gain, while an EMCCD includes many photodiodes but only one amplification stage. A multianode PMT, in turn, has only one photocathode but can be seen as consisting of parallel amplification channels. Finally, an SiPM is composed of many SPADs, each of which is a fully functional micro-photosensor with photocathode and amplification stage.

From these considerations follows that there are four possible combinations of the detection channel and readout channel multiplicities. This simple grouping already allows a coarse assessment of whether a sensor can have a position resolution or not.

For PGDs it is difficult to define "detection channels" and to distinguish them from readout channels.
Usually, an unsegmented photocathode is used, but the amplification (Townsend avalanche) and readout (induction in electrodes) can be intrinsically connected. For instance, for MWPCs and MSGCs the same electrode (wire/strip) is used to accelerate the charges and detect the avalanche. Such double-purpose electrodes are regarded here as both amplification and readout structures and when they are segmented, the substructures are considered parallel amplification channels.
On the other hand, for the RPC the readout electrodes are at ground potential and are not used for acceleration but only for inductive readout. In this case, the photocathode and the gas amplification by the graphite electrodes can be viewed as one continuous detection channel.
The case is simpler for HTGDs. As the name already implies, all such GDs have dedicated parallel amplification structures. For Micromegas, on the other hand, double-purpose electrode strips are used.
Therefore, GDs with segmented acceleration–readout electrodes (MWPC, PPAC, surface-amplification GDs, Micromegas class) or with segmented amplification channels (HTGDs) are considered to have multiple detection channels. Only the RPC and GDs with unstructured anodes (SWC, PPAC) are classified as single detection channel devices.

The sensors can be grouped by multiplicity as follows. For each multiplicity combination, first the known photosensor examples will be discussed. Based on this, then general conclusions can be drawn.

- **Single detection channel**
  - **Single readout channel**
    For semiconductors, photosensors with the single–single configuration are monolithic sensors. This includes the photoresistor, the photodiode family, and devices with gain such as APD, VLPC (visible light photon counter), SPAD, hybrid sensors and the phototransistor.

In vacuum tube sensors, a single–single multiplicity means that there is an unsegmented anode, which can take three forms. A wire anode is used in the phototube, a reflex anode is employed in the VPT, PMT and PMT hybrids, and bulk anodes find use in the micro PMT, CEM and a subtype of the MCP-PMT.
In gas-based sensors also three cases occur (independent of whether the employed photocathode is solid, liquid or gaseous). The sensor can have an unsegmented anode[96] or an unsegmented readout electrode[97]. Alternatively, the device actually *has* a segmented readout but only one detection channel[98]. In this case, the position information is lost and therefore only one readout channel is needed.

None of the abovementioned sensors with the single–single combination shows a position resolution. To enable this, a single detection channel would have to be read out with position sensitivity. This is possible under certain circumstances (see next item) but then requires multiple readout channels. To reduce the number of output channels to one, it would be necessary to additionally digitize the signal in a front-end electronics. So, in principle position sensitivity *could* be possible for single–single sensors, but this would require some effort and would only work for special sensor types. To the author's knowledge, this has not been realized so far.

– **Multiple readout channels**
This is the case for semiconductor sensors which consist of one cell with resistive charge-division readout, namely the PSD, PSAPD and HPSAPD. For GDs, only the RPC falls in this class, since all other GDs with multiple readout channels also have multiple detection channels.

Thus, for the rare cases where single–multi configurations are used, the sensors *are* position sensitive – otherwise, it would make no sense to use multiple channels. Interestingly, all of the known examples use resistive layers to accomplish this feat.

- **Multiple detection channels**

  – **Single readout channel**
  This comprises semiconductor sensors with an array of cells and only one readout channel. Examples include CCD types (CCD and subtypes, EMCCD, EBCCD, IEBCCD, ICCD, and streak camera), the SiPM and its hybrids, and the dSiPM.
  For vacuum tube sensors, some PMT subtypes (PMTs with VB dynodes and micro PMT) have multiple amplification channels but only one bulk anode. MCP-PMTs can also have only one bulk anode [169; ch. 11].
  PGDs and VPGDs can have unsegmented readout electrodes, as well. In addition, a segmented readout can be summed up, which also results in a multi–single configuration.

[96] Gas phototube (anode wire).
[97] SWC, PPAC and RPC.
[98] Hybrid-gas PMT, see p. 694.

Most multi–single devices have no capability to resolve the photon location. In the two cases which do achieve position sensitivity, this is only possible through use of a time-distributed readout scheme (CCD) – with the severe disadvantage of long readout times – or via digitization (dSiPM).

– **Multiple readout channels**
This group contains semiconductor sensors with an array of cells with individual readout channels (PDA, CMOS class, APD arrays, SPAD arrays, and imaging HPDs and HAPDs).
In addition, MaPMTs (mesh, foil, metal channel), the DS-QUASAR, MCP-PMTs and the VPT variant with a segmented dynode are located in this group.
Finally, the vast majority of GDs have multi–multi configurations, since these sensors are explicitly designed for position reconstruction.

All discussed multi–multi photosensors have the capacity to resolve the position, since only in this case it is reasonable to use multiple readout channels.

In conclusion, the single–single configurations so far has not produced position sensitive sensors, while the multi–multi and single–multi combinations always have a position resolution. For the multi–single class most sensors are not position sensitive but special readout techniques can enable this.
So, in order to obtain a photosensor with position sensitivity, in general both multiple detection channels and multiple readout channels are required *unless* some tricks are used (digitization, time-separated readout, resistive surface).

4.1.2 Special sensor types

4.1.2.1 Hybrid sensors

In this work a hybrid photosensor (HPS)[99] is defined as a device which consist of nested or coupled photosensors, where the readout structure of the primary sensor is replaced with the secondary sensor or a scintillator coupled to it.
Usually, an HPS consists of the combination of *two* photosensors, but also more sensors can be stacked. One example with three sensors is shown in fig. 4.16d. This photosensor is based on optical readout of a GD and contains a PGD which is operated in the glow discharge region. The produced scintillation light is amplified by an image intensifier (a light amplifier consisting of photocathode, MCP(s) and a scintillator) and the amplified light then is detected by a CCD.

While at first glance it may seem paradox to install a photosensor in a photosensor, this is an excellent means to circumvent certain shortcomings of each respective sensor type.

[99] In literature the term hybrid photodetector is more common. However, this label is ambiguous, since it has the same abbreviation as the hybrid photodiode (HPD) and because Hamamatsu also uses the term “hybrid photodetector” (*also* abbreviated as HPD) for its line of HAPDs.

Compared to PMTs, individual HPSs can have better properties regarding TTS, early pulse (EP) and late pulse rates, rise and fall times, CE, charge resolution [100], gain stability, position sensitivity, magnetic field tolerance, afterpulse rates, compactness, maximum active area, temperature stability, power consumption, production costs, and electronics costs [538, 545, 553] [153; pp. 297–300] [194; pp. 303–305].
Of course, this also comes with a number of possible disadvantages such as low gain, very high required voltages, chemical incompatibility of components, demanding electronics requirements, and high costs (same sources).
The pro and cons will be discussed separately for each HPS class and type.
Due to their many favorable properties, hybrid photosensors were used as early as 1965 (the first position sensitive sensor 1971) [472], and some of the most promising sensor developments are hybrids.

The primary photosensor is usually useless without the secondary sensor, as it was deprived of its capability for readout. Strictly speaking, it is not even a photosensor anymore, but merely an electro-optical preamplifier for the secondary sensor. The combination of the primary and secondary sensor, however, once more is a photodetector.
Since every diode based sensor is also photosensitive, the border between single-sensor concept and hybrid-sensor is somewhat blurred.
In addition, in many cases secondary photosensors which are based on semiconductors are actually used to detect *electrons* instead of photons, for which the device often is not optimized. For some devices, this mismatch has lead to the production of Si sensors which are specialized for EB. For instance, in the VSiPMT (p. 691) an SiPM is used whose structure was modified to act as an electron detector and current amplifier [554]. This is termed silicon electron multiplier (SiEM) instead, and contains no epoxy resin layer, a thinner SiO_2 passivation layer to allow pe penetration into the Si bulk, and a p-over-n junction to optimize the avalanche trigger efficiency [555] [554]. In the SiGHT (p. 692) , similar measures are considered for the same reasons [556]. One can also assume that Hamamatsu uses APDs which are adapted to EB in their HPD series.

Every sensor which incorporates a light amplifier based on photon-to-electron conversion (e.g. an image intensifier) is automatically an HPS. Such an electro-optical amplifier is formed by the combination of primary photosensor – which converts the incoming photons into photoelectrons – and either a scintillator or the light emission from gas excitations. Correspondingly, a hybrid sensor which contains a scintillating medium and a subsequent photosensor can also be viewed as a light amplifier which is connected to a secondary photosensor.

As already discussed on p. 572, for vacuum tube-based HPSs different focusing types are possible, which strongly influence the resulting sensor properties.

Hybrid photosensor classes

In order to provide an overview of the plethora of hybrid photosensors which have been developed so far, and since the labels in literature often are contradictory and imprecise, a uniform naming convention for HPS classes shall be introduced. The conventions which are used in literature will be retained as far as logical considerations and clarity allows.

[100] Especially due to the crucial first amplification step.

The label of each class consists of the prefix "hybrid" followed by the intermediary amplification steps and concludes with the final photosensor/electron sensor. Since most types are based on EB, this gain method is not specified, whereas the use of scintillation and gas amplification are always included in the name.
In the acronym of the class name, the following abbreviations are used besides the short names of sensors themselves: hybrid (H), scintillator/scintillation (S), MCP/MCP-PMT (M), gas (G), and tynode (T). For clarity, hyphens can be added between gain steps and sensors.
As an example, the three-sensor hybrid shown in fig. 4.16d is a hybrid sensor (H) with gas amplification (G) whose scintillation light (S) is amplified by an MCP-PMT (M) and scintillator (S) and is detected by a photodiode array (PD). This is then abbreviated as HGS-MS-PD. This particular sensor, however, is subsumed under the genre HGS-PD, whose principle of operation is the same except for the extra MS step (image intensifier) [101].

The hybrid sensors which have been developed so far can be grouped in the following classes (cf. fig. 4.1):

- *EB of PD:* **hybrid photodiode (HPD)** [553] [153; p. 297]
  This sensor type is sometimes called "hybrid photodetector" in literature (abbreviated as HPD as well) [557] [194; p. 303] [170; p. 17] and Hamamatsu also uses this appellation and abbreviation for their series of HAPDs (see next entry). To distinguish them, the latter are labeled "Hamamatsu HPDs" (HHPDs) in the present work. In [153; p. 297] hybrid photodiode are also called hybrid photomultiplier tubes and abbreviated as HPMTs, which quickly should be forgotten, because there is a class for which this label would be far more appropriate.

- *EB of APD:* **hybrid APD (HAPD)** [553] [194; p. 304] [515; p. 426]

- *EB of SiPM:* **hybrid SiPM (HSiPM)**
  In [558] this group is called G-HPD (Geiger mode hybrid photon detector).
  If in the future other G-APD sensors such as the dSiPM are included in an HPS, either this class could be extended and renamed to HGAPDs or – which would be clearer – additional specialized classes could be introduced, e.g. HdSiPM.

- *EB of MCP + EB of PD:* **hybrid MCP photodiode (HM-PD)**

- *EB of MCP + EB of scintillator + PD:*
  hybrid MCP scintillator photodiode (HMS-PD)

- *EB of tynode + EB of PD:* **hybrid tynode photodiode (HT-PD)**

- *EB of scintillator + sensor:*

  – *PMT:* **hybrid scintillator PMT (HS-PMT)**
    In [553] this group was labeled hybrid phototubes with luminescent screens (HPLS). [239] calls them crystal-based hybrid photon detectors (X-HPDs; like the photosensor discussed in the same paper and here on p. 657).

[101] In principle, for all HPSs which use scintillation, extra MS steps can be included. However, for many combinations this is unnecessary, since the gain and charge resolution already are sufficient without the extra step.

  - *PD:* **hybrid scintillator photodiode (HS-PD)**
  - *SiPM:* **hybrid scintillator SiPM (HS-SiPM)**
  - *PGD:* **hybrid scintillator PGD (HS-PGD)**

- *GD producing scintillation light + sensor:*
  Here the GD is operated in a range where the avalanche produces photons through radiative relaxations of excited gas particles. This scintillation light is detected by an additional photosensor.
  - *PD:* **hybrid gas scintillation photodiode (HGS-PD)**
  - *PGD:* **hybrid gas scintillation PGD (HGS-PGD)**

The properties of each class will be discussed in detail in their respective sections.

4.1.2.2 Direction-sensitive sensors

For Photosensors with a spherical or hemispherical photocathode, certain parts on the cathode have a directional preference for incident photons for geometrical reasons (fig. 4.7a). However, this is not very pronounced, since for every point on the cathode all directions within a solid angle of 2π can be detected. Furthermore, the internal reflections in the glass can significantly enhance the detection of rays with large incident angles versus the surface normal (see section 2.1.2). For semitransparent cathodes also the double cathode effect (p. 64) must be taken into account. Both effects can complicate the directional reconstruction.
Thus, any position-sensitive photosensor with concave shape can be used as a direction sensor, albeit with a comparatively coarse directional resolution. This was used by NEMO in collaboration with Hamamatsu to develop a prototype of a 10" PMT (Hamamatsu R7081) with discrete direction sensitivity by focusing each quadrant of the cathode onto a separate dynode chain [448]. This device can at the same time also be considered an MaPMT (Ma-R7081) and an imaging sensor.

Direction-sensitive light guide

The direction resolution can be improved by attaching a sectored light concentrator (LC) to the photosensitive window, which maps regions in the angular phase space of incident photons to regions on the cathode (fig. 4.7a). In this manner, any position-sensitive photosensor can be converted into a direction-sensitive one and this can be considered a directionality upgrade. This could be very promising for novel event reconstruction concepts, but has the drawback that the achievable effective magnification of the LC and the detector homogeneity are worse than for LCs specifically designed for such purposes (see 5.3.1).

Such a direction-sensitive light guide was developed (fig. 4.7b) for the QUASAR-370 (a HPS, see p. 654) [559]. It was foreseen to install a multianode PMT in the hybrid sensor which would result in a directional-sensitive QUASAR (DS-QUASAR). No complete sensor prototype was produced.

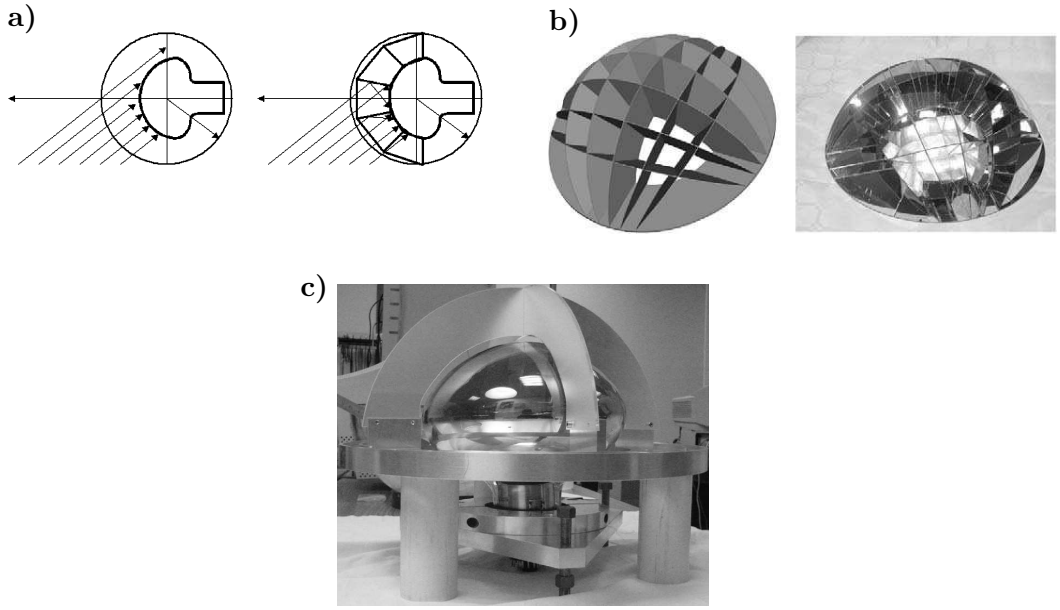


Figure 4.7: Direction-sensitive photosensors: Direction-sensitive light guides can convert a position-sensitive photosensor in a direction-sensitive one or improve the direction resolution of a concave sensor.
a)–b) DS-QUASAR: a) Concept, from [559]. b) Schematic and photo of prototype, also from [559]. See fig. 4.12d for the layout of the photosensor for which the light guide was developed (QUASAR-370).
c) Segmented mirror with a multianode Hamamatsu R7081 for a direction-sensitive optical module for NEMO (see p. 9), from [449].

The employed design basically consists of skewed rectangular cone-concentrators. With regard to light collection, this is a very simple design for which a high fraction of rays is backscattered instead of focused onto the cathode [560]. However, the increase of the effective collection area was not the focus of this development.

Another sectored light guide was produced for the Ma-R7081 mentioned above by the NEMO collaboration (fig. 4.7c) [449]. In this case, mirrors divide the cathode into four quadrants.

4.1.2.3 Wavelength-selective sensors

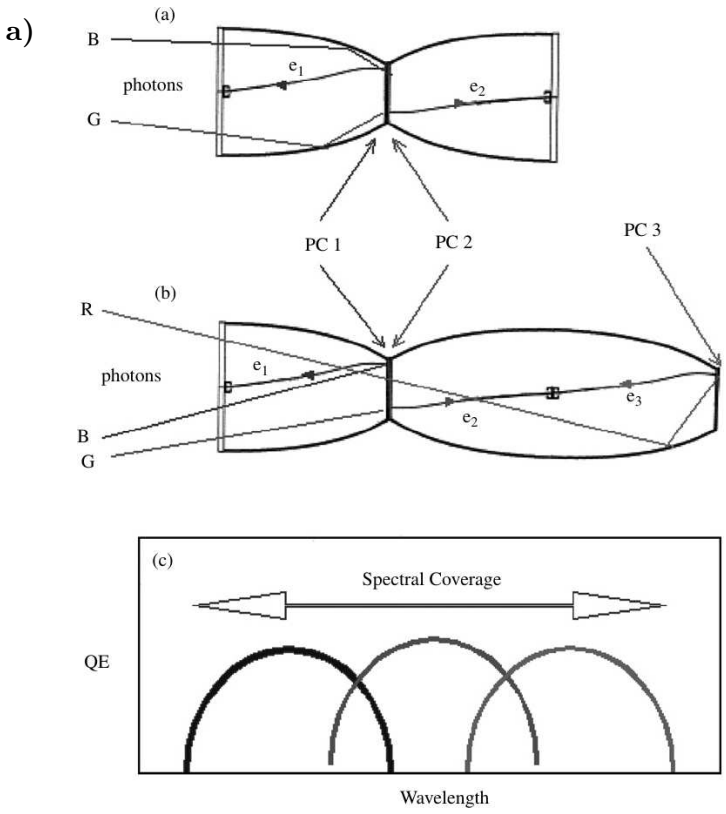
Wavelength sensitivity is a very useful feature for LSNDs and WCDs, as was pointed out by [284, 561]: For a WCD it allows to correct the wavelength dependence of the Cherenkov light that results from dispersion and Rayleigh scattering, which can improve the position reconstruction and time resolution. For LSNDs, it can provide a discrimination between Cherenkov and scintillation light. This allows to reconstruct the event direction and in this way can help to identify solar neutrinos, neutrinoless double beta decays (discrimination from solar neutrinos) and study long baseline neutrino oscillations (separate ν_e from π_0). This is of particular interest for water-based liquid scintillator detectors like THEIA (see 1.3.3). In addition, a discrimination of Cherenkov and scintillation light can aid the α–β discrimination, since Cherenkov light is absent for alpha decays.

There are two different approaches to obtain a wavelength-selective photosensor for LSNDs or WCDs which have been pursued so far: the TransReFerence [561] and the dichroicon [284].

The TransReFerence [102] separates wavelengths through the absorption of different spectral regions in multiple photocathodes. There are two configurations for the detection of two and three wavelength ranges (fig. 4.8a).
In the first layout (top of fig. 4.8a), two ReFerence sensors (see fig. 4.14h and p. 679) with photocathodes for different spectral ranges are optically coupled together. The photoelectrons from the respective cathodes are focused onto PDs or APDs in their ReFerence tubes by a strong electric field. To minimize absorption in the first cathode (Lambert-Beer law) it is sensitive to short wavelengths, while the second cathode (on the back of the first) detects long wavelengths. Thus, the first cathode must be sufficiently thick to convert most short-wavelength photons to electrons, but thin enough to still transmit long wavelength photons. Since short wavelengths are absorbed more strongly in typical photocathode materials (see 2.1.2), even the same material can be used for both cathodes to achieve a limited wavelength sensitivity. The first cathode is used in reflective mode (see p. 65), the second in transmission mode for the first pass and in reflective mode for transmitted photons which are reflected back onto it.
To distinguish three spectral bands, a third ReFerence tube with a cathode which is sensitive to the longest wavelength range is appended as shown in the middle of fig. 4.8a.
The spectral bands are usually broad and will overlap depending on the cathode material and thickness. In order to obtain sharp cuts, one could use reflective interference filters or Bragg reflectors beneath the first and second cathode.

The dichroicon (fig. 4.8b) is a generic concept, which is independent of the sensor type [103]. It allows to separately detect two spectral regions via a wavelength-selective LC which consists of dichroic mirrors. Dichroic short-pass filters are reflective above a threshold wavelength and transmissive with a low absorbance below (vice versa for long-pass filters).
The dichroicon consists of a red-sensitive sensor behind a dichroic long-pass filter and a a blue-sensitive second sensor behind plane dichroic short-pass mirrors which are arranged in the shape of a Winston cone (compound parabolic concentrator (CPC) [560]). Two variants exist: Either a second CPC concentrates light onto the second sensor, or a position sensitive flat sensor like the LAPPD or an array of SiPMs is used. In this case, an acrylic glass light guide directs the transmitted light onto a different region of the sensor / sensor array.
A first prototype with two PMTs was constructed with a simpler, third geometry (not shown). Here, a dichroic Winston cone focused long-wavelength light onto a red-sensitive PMT [104]. The LC was surrounded by a reflective cylinder which guides transmitted short-wavelength photons to a blue-sensitive PMT (Hamamatsu R1408; hemispherical, 8”, VB).

[102] Paragraph based on [561].
[103] Paragraph based on [284].
[104] A Hamamatsu R7600U-200 (18×18 mm² sensitive area, MC, ultra bialkali) or an Hamamatsu R2257 (cylindrical, 2”, LF, extended red multialkali) [189].



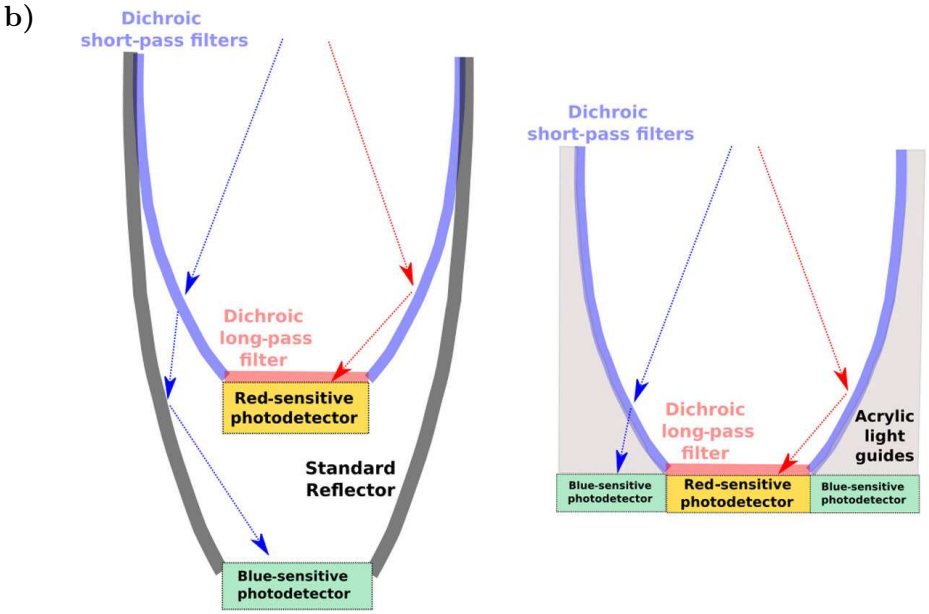


Figure 4.8: Wavelength-sensitive photosensors
a) TransReFerence, from [561]. Photons are detected by ReFerence sensors (see fig. 4.14h). Top: Configuration for two wavelength ranges. Middle: For three wavelength ranges. Bottom: Schematic of the expected quantum efficiency spectra for the three-cathode variant.
b) Dichroicon, from [284]. Wavelength-sensitivity is achieved through dichroic mirrors which transmit or reflect photons depending on wavelength. For the detection, either multiple photosensors are used (left) or a position-sensitive sensor is employed (right), where separate regions are used for different wavelengths.

Both the TransReFerence and the dichroicon use light concentrators in their design, but the underlying principle of operation is different.
The dichroicon relies on the use of LCs to be able to use the same entrance aperture for the detection of two wavelengths regions with similar effective detection areas. It achieves wavelength-sensitivity through dichroic mirrors. In fact, the dichroicon is actually not a photosensor but a dichroic LC and can be considered a wavelength-sensitivity upgrade for a photosensor.
Contrary to this, the TransReFerence does in principle *not* need to use LCs, since the transmissive property of the photocathode is used instead. The principle of operation would also work with cylindrical walls, even if these are not reflective. Here, the wavelength resolution is obtained through the QE spectra of the cathodes and/or the wavelength-dependent absorption (which resembles a dichroic filter). The TransReFerence *is* a photosensor and consists of 2–3 sub-sensors. Since the ReFerence itself is a hybrid sensor, this would make it a hybrid hybrid sensor.

Influence of light concentrators on sensor properties

Apart from the opportunities which are granted by the inclusion of an LC in a photosensor, this also introduces several disadvantages which have to be considered:

- **Photon collection efficiency:** The efficiency of photon collection on the receiver (the photosensor or photocathode) can be significantly lower than 100% depending on design, concentration factor and reflectance (see [560] and 5.3.1). In addition to reflection losses, this is because the used string-method concentrators are ideal LCs (perfect light collection with at most one reflection) only in 2D and for the *specific* light source geometry for which they are designed [105]. If this 2D layout is rotated to obtain a 3D LC, the LC becomes non-ideal and for non-meridional rays (outside the cross-section through the rotational axis) losses occur or multiple reflections can be required before detection. Thus, an additional photon collection efficiency has to be multiplied with the other efficiency factors of the PDE in eq. (2.1). Furthermore, a fill factor of the photosensor (e.g. SiPMs) must be considered.

- **Time-of-flight variance:** The LC introduces a TTS due to a time-of-flight variation of the photon paths for different incident angles and positions on the entry aperture. This is aggravated by a non-ideal 3D LC, which also accepts rays with multiple reflections and even spiraling paths with many reflections [560].
  Simulations performed in the present work for the LENA LCs showed a standard deviation of the time-of-flight of 368 ps [106] for a CPC with 17.9 cm entry aperture [107] and a PMT cathode of 14 cm radius [108] as exit aperture. This TTS can be scaled with the aperture size for a constant critical angle. For stronger concentrations as in the ReFerence and in the dichroicon, the relative TTS should increase significantly. Thus, for a target sensor with low TTS – such as

[105] The CPC, for example, is constructed for parallel incident light like from the sun, but not for a detector in finite distance.
[106] Transit time 1.2–3.4 ns, mean 1.59 ns.
[107] Critical angle 51.4°, geometric concentration factor 1.64, 25.5 cm length. For a CPC, the critical angle is the ray connecting the left end of the entry aperture with the right end of the exit aperture (compare fig. 4.8b).
[108] Spherical segment with 19.5 cm curvature radius.

an LAPPD or SiPMs – the photon path TTS can easily become the dominant source of the total TTS of the photosensor.

- **Limited field of view:** Due to the conservation of étendue (product of cross-section and angular extent of light rays) [109], a spatial concentration (i.e. a smaller exit than entry aperture) automatically limits the accepted angular range [560]. Depending on the use case, this can be a disadvantage or an advantage.

4.1.2.4 Cryogenic sensors

Many sensors which were developed for cryogenic temperatures can not be operated at room temperature, such as calorimetric photosensors based on superconductivity (TES, STJ, SNSPD, KID, superconducting granules; see p. 568). Consequently, these are of no relevance for large neutrino detectors and are not further studied here and not listed in the photosensor tree in fig. 4.1.
On the other hand, some photosensors like the QUPID (p. 680) and SiGHT (p. 692) were developed for operation in cryogenic liquids for dark matter or neutrino-less double beta decay experiments, but in principle *can* also be operated at room temperature. However, this may come at the price of a deterioration of the sensor performance. For example, the SiGHT includes an SiPM whose dark count rate (DCR) is as high as 10^6 Hz at room temperature [556]. These sensors therefore require a closer look to consider their aptitude for LSNDs and are included in the subsequent analysis.

4.1.3 Overview of following sections

The preceding sections have laid down the basic principles required to understand the function and properties of every photosensor relevant here.
In the following sections 4.2–4.4, the individual sensors will be discussed. For each sensor, the contents will be presented in the following order where reasonable: After a summary of the sensor concept and general remarks (connection to other sensors, history, use case), the principle of operation is explained using the previously established concepts. Following this, the mean and best observed properties are listed as far as available in literature, where a similar property order as in ch. 2 is used. Next, sensor variants are addressed, before concluding with a discussion of the pros and cons for a use in LSNDs. The most promising sensors are collected in a short list and their suitability for LSNDs is later discussed in detail in 4.5.1.
The sensors are grouped as shown in the photosensor tree in fig. 4.1, where arrows and lines indicate the progression of nesting. To reduce the number of structure levels in the document, conversion medium and transport medium are skipped where only one option exists. For amplification subtypes, the exact description is summarized in one title or only the most specific type is used to group sensors.

[109] For an ideal LC, the étendue is conserved, for a non-ideal LC it will be lower at the exit aperture.

In the tree and the following text only sensors are listed which have been realized or proposed so far. Based on combinations of the above-mentioned principles and materials, there are many other conceivable combinations, some of which will be discussed in 4.5.3.

All known sensor developments which might be relevant for LSNDs are listed in fig. 4.1 for reference. To stress similarities and maintain a clear overview, sensors were subsumed under their main sensor class where applicable (photodiode, CCD, CMOS, PMT, MCP-PMT, HPD, SPGD, GPGD, LPGD). Some subtypes are separately displayed where seen as justified by marked development steps (e.g. CCD → pnCCD, Skipper CCD; CMOS → sCMOS).

Overlap of sensor classes

For some developments it is difficult to place them. The boundaries between sensor types can be blurred, or projects overlap and a subset of people is involved in the development of both sensors, and some designs even changed sensor *classes* during development.

For example, the LAPPD (figs. 4.11b and 4.11c) and H33D (fig. 4.6d), although labeled with a proper noun, in the end are just MCP-PMTs (fig. 4.11a) with a special readout layout (strip line + delay line, XS, XDL) and large areas (LAPPD). In addition, H33D and LAPPD were developed by largely overlapping groups of people, and in some papers the borders blur between a development branch of the H33D and a predecessor of the LAPPD [551, 562].
Similarly, the NNVT GDB-6201 and GDB-6203 are simply MCP-PMTs with fountain focusing instead of proximity focusing.
This poses the question, whether a different focusing or readout type is sufficient to assign a sensor to a new class. For PMTs this is *not* the case, and mesh and MC PMTs are generally considered subtypes of PMTs, although their detection properties are vastly altered (position sensitivity, magnetic field tolerance, TTS, size limit, ...). The fundamental detection principle – as illustrated in the photosensor tree in fig. 4.1 – is ultimately still the same for all PMT variants: Conversion in a photocathode is followed by EB on dynodes for amplification. Thus the LAPPD, H33D and GDB-6201/6203 here are also considered subtypes of MCP-PMTs.

The term HPD is both an HPS class and used as a proper name for some sensors of this class. In the latter case, the definition is more stringent and usually refers to electron-bombarded PDs where the pixels/pads contain no preamplifier and have individual readout channels. In most cases, each channel has an external preamplifier in the front-end electronics, which is bump-bonded to each pixel and contained in the same vacuum tube. If instead the preamplifier is *part* of the pixel, the device is considered an EBCMOS, which also belong to the HPD class. In this work, the label given by the inventors of each device was used, although some devices which are called HPD are actually EBCMOS types (Timepix HPD).

The X-HPD, in turn, is a continuation of the C2GT designs (hybrids using either PDs or APDs) with adapted technology (scintillator + PMT). The X-HPD was originally foreseen as an HPD, then was realized as an HS-PMT, and in the end it was considered turning it into an HSiPM [558].

4.2 Internal photoeffect sensors

The first sensor family to be addressed uses the IPE (see 4.1.1.1).
As already observed on p. 570, the only secondary IPE detection mechanisms which are relevant for LSNDs are photoconductivity and in principle the photovoltaic effect. Sensors based on thermal and photochemical interactions, the photon-drag effect, or the photorefractive effect are not capable of UV/visible photon counting at room temperature and thus are not included here. These sensors were already covered in 4.1.1.1. All subsequent sensors are based on photoconductivity except where explicitly stated otherwise.
The studied IPE-based photosensors are shown in fig. 4.1 and their schematics are summarized in figs. 4.9 and 4.10.

Unamplified sensors

Photodiode (PD) [110]

Photodiodes (PDs) employ semiconductor diodes in which incident photons create electron–hole (e–h) pairs via the photoconductive effect (figs. 4.9a and 4.9c). The created charge carriers are drawn by a reverse bias to opposing electrodes where they induce a signal.
Usually, silicon PIN [111] or p–n diodes are used. However, many other types exist as well, including Schottky barrier (metal–semiconductor junction) [511; pp. 680–682] [563], MSM (metal–semiconductor–metal) [511; pp. 712–716] [564], and MIS (metal–insulator–semiconductor) [564] diodes. A reverse-biased LED (light emitting diode) can in principle be used as a photodiode as well (although not optimized for light detection). Solar cells also are a type of PDs with large areas, but they operate without external bias and thus use the photovoltaic effect instead [511; p. 667]. This requires high intensities and is not suited for LLL detection.

The QE of PDs can reach peak values above 90%, decreases for long wavelengths due to the growing absorption length, and falls towards low wavelengths, since the absorption length becomes similar to the thickness of the inactive surface layer. The sensitive spectrum can range from 190 to 1100 nm.
For PIN diodes, a reverse bias extends the depletion zone in which created charge carriers can effectively be collected, with a typical thickness of about 100 µm. This results in lower noise due to a smaller capacitance and increases the QE at long wavelengths, where the absorption length is in the same order as the thickness of the depletion region.
PDs are compact, robust, require only low supply voltages (< 50 V), are insensitive to magnetic fields, have a low temperature sensitivity and are affordable. Their rise time lies between 10 ps (p–n, Schottky) and 0.1–10 ns (PIN) [511; p. 666]
However, they have no gain and can only detect high light levels of more than a few hundred photons even when using slow low-noise preamplifiers. Their TTS is

[110] Content based on [194; pp. 306–308] [515; pp. 421–424, 436] [170; p. 12].
[111] p- and n-doped regions around an undoped intrinsic region.

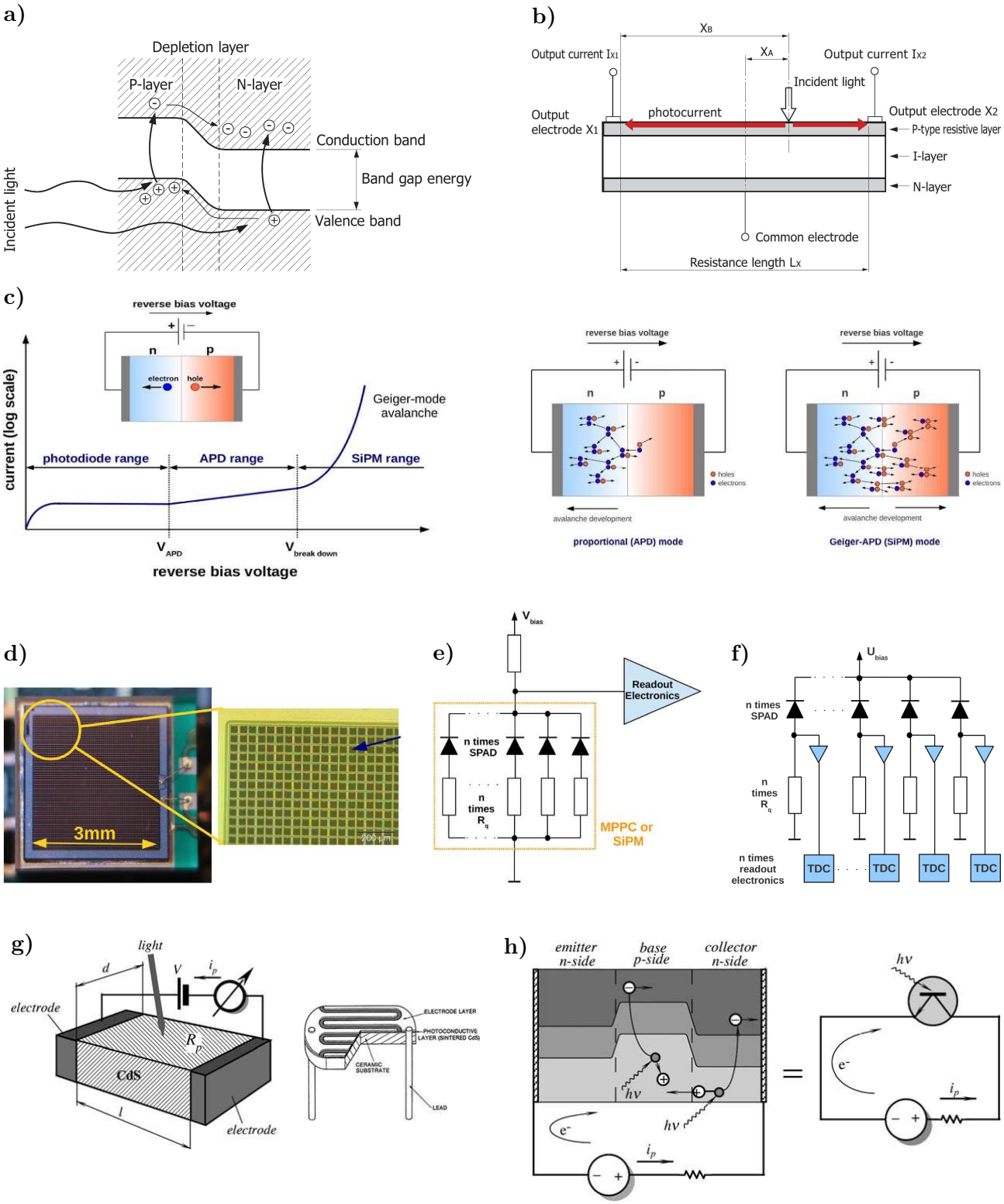


Figure 4.9: Semiconductor photosensors based on photoconductivity and photovoltaic effect. Discussed in text.
a) p–n photodiode (PD), from [477; ch. 2, p. 3].
b) Position sensitive diode (PSD), from [477; ch. 2, p. 17].
c) Operating modes of photodiodes for reverse bias voltage. In order of increasing voltage: Photodiode mode (PDs), linear mode (avalanche photodiodes, APDs), Geiger mode (silicon photomultipliers, SiPMs). From [293].
d) Photo of an SiPM, a Hamamatsu S10931-3050P. From [293].
e) Analog SiPM (SiPM). From [293].
f) Digital SiPM (dSiPM). From [552].
g) Photoresistor. From [480; p. 473].
h) Phototransistor. From [480; p. 471].

comparatively large (ca. 2 ns) for their small size (up to 20 mm diameter), the self capacitance of large diodes reduces the bandwidth, and they show high noise and very large dark currents.

Since spe detection is not possible, PDs cannot be employed in LSNDs.

Photodiode array (PDA) [112]

A photodiode array (PDA) is simply an array of PDs, either a 1D linear array or a 2D array.
PDAs are also available with amplifiers for each channel in a readout chip. This should be distinguished from a CMOS sensor (see below), where the amplifier is part of the pixel.

Like for PDs, spe detection is not possible, which excludes this sensor.

Position sensitive diode (PSD) [113]

A position sensitive diode [114] (PSD) is a PD which is fabricated from high resistance silicon and uses resistive charge division (also called lateral effect) to obtain a continuous position sensitivity in 1D or 2D (figs. 4.9b and 4.6g).

Devices have rise times of 1–2 µs and a spectral response between ca. 320–1100 nm. The sensitive area is $1\times(1\text{–}12)$ mm for 1D PSDs and in 2D sensors takes the shape of a square with 4–27 mm side length. The spatial resolution can be notably below 1 mm if low bandwidths are acceptable.

Since only high photon fluxes are detectable and the response is slow, this sensor in ineligible for LSNDs.

Impact ionization

Linear mode

Electron-multiplying CCD (EMCCD) [115]

An electron-multiplying CCD (EMCCD) is a CCD (described on p. 632) which includes an electron multiplication register between the last shift register and the output amplifier (fig. 4.10c). This is basically a shift register in which far larger voltage differences (35–40 V) are used. The resulting higher electron velocities can lead to impact ionization with a small probability of about 0.5–1.5% (linear mode).

[112] Content based on [477].
[113] Content based on [480; pp. 310–314] [565].
[114] Called position sensitive detector by Hamamatsu.
[115] Content based on [508] [478; pp. 271–272].

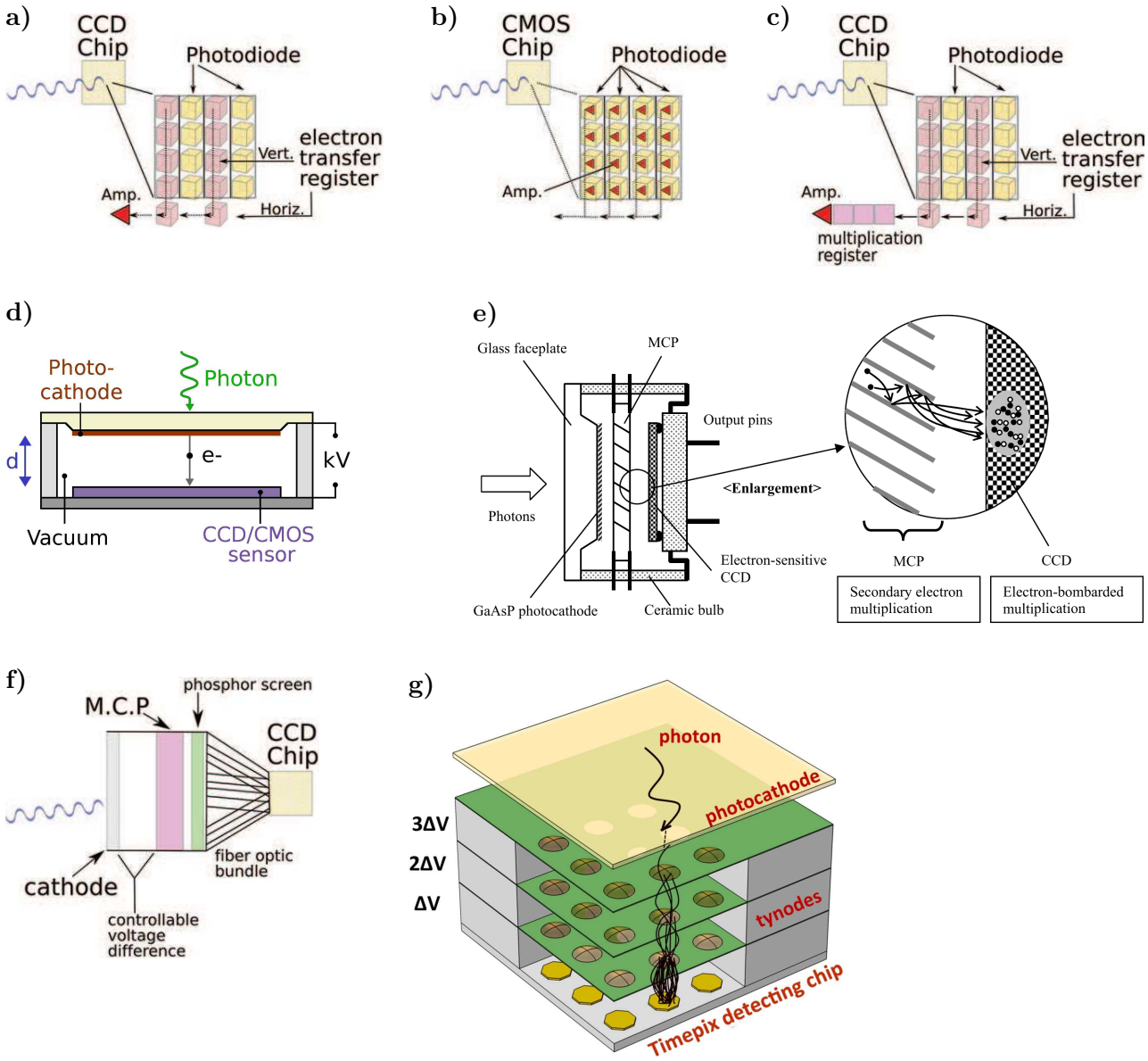


Figure 4.10: Photosensors based on photodiode arrays. Discussed in text.
a) Charge coupled device (CCD), from [478; p. 269].
b) CMOS image sensor (CIS), here active pixel sensor (APS) variant. From [478; p. 269].
c) Electron-multiplying CCD (EMCCD), from [478; p. 269].
d) Electron bombarded CCD (EBCCD) and electron bombarded CMOS (EBCMOS), from [566].
e) Intensified EBCCD (IEBCCD), from [567].
f) Intensified CCD (ICCD), from [478; p. 269].
g) Timed Photon Counter (TiPC), from [568].

While this produces only a minute gain per stage, a great number of stages (400–590) is used, which results in a total gain of several 100 or 1000. This actually makes the EMCCD sensitive to single photons.
However, the dark current becomes the dominant noise source (the EMCCD is *also* sensitive to single thermally excited electrons) and the sensor has to be cooled to −80 to −100 °C for operation. In addition, the high number of low gain processes leads to a large ENF (see eq. (2.65)), the gain is unstable and deteriorates over time, and the devices are expensive.

An electron-multiplying CMOS (EMCMOS) concept has also been demonstrated [566].

Due to the large DCR at room temperature and the slow readout, EMCCD and EMCMOS are not included in the short list.

Avalanche photodiode (APD) [116]

Compared to PDs, avalanche photodiodes (APDs) are operated at a far higher reverse bias (several 100 V) and contain an additional p–n junction with high doping, which creates a strong field of typically ≈ 1 kV/mm near the junction. Photons are absorbed in the thicker lower-doped depleted region. When the created electrons drift to the anode, they pass the high field region, where they start an avalanche through impact ionization (compare fig. 4.9c, additional junction not shown). The multiplication is in the linear mode and the gain usually is between 10 and several 100, which allows to detect light intensities down to 10–20 photons. The gain can reach 10^4 and with such a high gain at 77 K less than 8 photons/pulse could be detected [569].

Depending on the material and from which side the APD is illuminated, the spectral range can extend from 200 to 1700 nm. A QE of 77% at 400 nm and a PDE of up to ca. 70% can be achieved. APD diameters normally lie between 0.04 and 40 mm, but a device with 45 cm^2 has also been fabricated [569]. The TTS (σ_T) can be as low as 20 ps, the response time is around 100 ps [511; p. 666] (limited by self-capacitance), and APDs have good magnetic field tolerance.
However, in addition to the shot noise from leakage currents, statistical fluctuations occur in the generation of the avalanche. These cause a quite high ENF of 2–10, and the S/N exceeds that of PDs only for low light intensities. The gain depends strongly on voltage and temperature, which requires to use stable power supplies and often cooling. Furthermore, the dark current is much higher than in PMTs.

While the QE and TTS of APDs are excellent and their insensitivity to magnetic fields is favorable, the inability to detect single photons excludes these sensors for LSNDs.

APD array

APDs can also be pixelized to form an imaging sensor [569, 570] [477; ch. 3, p. 7]. For example, in the proximity-focused HAPD of Belle II, four 6×6 APD arrays are used as the bombarded stage to obtain position sensitivity [498].

Like APDs, this sensor is not considered due to its lack of spe sensitivity.

[116] Content based on [515; pp. 423–424] [194; p. 308] [170; pp. 12–13].

Position sensitive APD (PSAPD) [117]

The position sensitive APD (PSAPD) uses the same principle as the PSD for an APD. A resistive layer with four electrodes in the corners is placed on the back surface and the interaction position can be reconstructed with a charge division algorithm. A resulting pincushion distortion can be corrected numerically.
Radiation Monitoring Devices produced PSAPDs with diameters from 8×8 to 28×28 mm^2, where the large diameter specimens had to be cooled to −40 to −20 °C to reduce the leakage currents.

Since large photon fluxes are required for detection, PSAPDs cannot be used in LSNDs.

Visible light photon counter (VLPC) [118]

The visible light photon counter (VLPC) [119], developed in 1987, is an APD variant where a very high donor (As) concentration in Si creates an impurity band 50 meV below the CB. The band gap of Si is used for photon absorption, while the small impurity band gap is used to create an avalanche of electrons and ionized donor atoms, since the high density of impurity atoms allows impurity band conduction (charge hopping between impurities). Due to the small impurity band gap, a low field of only 6–7 V is sufficient to reach a gain of up to $5 \cdot 10^4$, which enables spe detection with a very low charge resolution.
However, because of the small band gap, the VLPC is also sensitive to IR photons and must be cooled to 6–8 K. Even then, the DCR amounts to 5–10 kHz for a sensor with 1 mm diameter.
The QE is about 90%, the sensor is fast and insensitive to magnetic fields, but its size is limited to a few mm [573].

In spite of their excellent QE, VLPCs are excluded for LSNDs due to the cryogenic operating temperatures, the high DCRs and their small size.

Geiger mode

Single photon avalanche diode (SPAD) [120]

The single photon avalanche diode (SPAD), first developed in 1963 [574] [121], is a photodiode which uses avalanche multiplication in the breakdown regime as gain mechanism and is therefore also called Geiger-mode APD.

A SPAD is operated about 10–20% above the breakdown voltage and the fields are far higher (order of 10 kV/mm) than in an APD, which enables the holes to also gain sufficient kinetic energy for impact ionization. Therefore, a single photon can trigger a self-sustaining avalanche (cf. fig. 4.9c), which has to be quenched by lowering the bias voltage below the breakdown voltage. Before the next photon can be detected, the voltage has to be restored, which introduces a dead time of

[117] Content based on [569].
[118] Content based on [570–572] [194; p. 308].
[119] Initially this was called solid state photo multiplier.
[120] Content based on [478; pp. 13, 42–45] [293] [515; pp. 424–426].
[121] According to [170; p. 14].

typically several 10 ns. This layout results in high gains of about 10^5–10^7 with nearly constant output signals and allows spe detection above gains of ca. 10^5. However, since the output pulse is independent of the number of photons, the SPAD is a binary device [304] – it can only detect whether light is incident or not – and is unsuited for photon counting [122].
SPADs are manufactured from Si using a custom process or a CMOS process, or from III/V materials for near-infrared (NIR) detection. The quenching and sometimes also the readout circuit is included in the device. The avalanche is detected by a sensing resistor, a transistor or a trans-impedance amplifier, and a subsequent discriminator can be used to obtain a digital output.

The active area generally is circular to obtain uniform fields, and diameters range from 10 to 500 µm.
The peak PDE is typically 60% and lies in the blue spectrum. Here, the PDE is the product of QE, fill factor (FF) and the probability that photo-generated charge carriers reach the high-field region and trigger a detectable avalanche (avalanche trigger probability).
SPADs have an extremely low TTS of a few 10 ps, where the tiny remaining variations are due to different absorption positions and the statistical avalanche buildup time. The transit time distribution (TTD) shows a Gaussian with an exponential tail of ca. 2.5 ns lifetime. The latter originates from the carrier diffusion time for photons which are absorbed in the neutral region. The tail is significantly more pronounced for longer wavelengths due to the then higher absorption in the neutral region.
Quenching is possible in different ways. In the most common case, a relatively large resistor is connected in series with the SPAD (passive quenching). This causes a voltage drop for high currents and recharges the diode with a time constant determined by the resistance and the total capacitance. During the exponential recharge, incident photons *can* trigger pulses that then have a reduced charge, which makes it difficult to define the dead time. Alternatively, the bias voltage can be regulated by a transistor (active quenching), which results in a faster recharge and a well-defined dead time. Mixed active/passive quenching is also used.
The pulse length is between 5 and 20 ns.
The DCR is typically 10–1000 Hz and is caused primarily by thermal carrier generation from states in the band gap. The dark current drops with falling temperature by a factor of 2 ca. every 10 °C.
Afterpulses occur due to trapping and release of charge carriers in deep and shallow traps in the high-field region. This produces exponential or multi-exponential delay distributions, and the delays rise for lower temperatures. To reduce the AP rate, the recharge time can be increased so that most APs occur before the SPAD is fully recharged. In addition, about $3 \cdot 10^{-5}$ photons are created per charge carrier in the avalanche and can be absorbed in the neutral region. The generated charge carriers can diffuse to the high-field region and cause additional avalanches. The rate of this AP type (diffused APs) can be reduced by using a low-lifetime substrate.
SPADs are also insensitive to magnetic fields.

The PDE and TTS of these sensors is excellent. The inability to detect more than one photon is offset by the low multi-hit probability due to the small size, but

[122] It has, however, been possible to determine the number of photons from the slope of the rising flank for up to four coincident photons [575].

this would require to use front-end electronics inside a detector to reduce the number of channels. Regardless of this, the DCR is far too high for LSNDs without cooling (which is not possible there), which excludes SPADs.

SPAD array (SPADA) [123]

By arranging multiple SPADs in an array, larger sensor sizes can be realized. In addition, an imaging sensor or multi-photon detection can be realized in this way. In a SPAD array (SPADA), every SPAD is read out *individually*, which allows binary imaging. As opposed to this, in an SiPM (see below) the signals of all SPADs are *summed up* to obtain multi-photon detection with a high dynamic range but no position sensitivity.
SPADAs can be manufactured in 1D or 2D arrays using either a custom process or a CMOS process. The latter allows to include quenching, control and readout electronics in the pixel. Compared to custom processes, however, a CMOS process results in higher DCRs and has other disadvantages, since the implants and fields are not optimized. Time-resolved photon counting can be implemented through the inclusion of a time-to-digital converter (TDC) and a counter in each pixel, but this significantly reduces the FF. SPADAs have also been used to resolve a signal temporally instead of spatially, similar to a streak camera (see p. 659).

Since SPADAs suffer from the same high DCR as SPADs, they are not included in the shortlist.

Silicon photomultiplier (SiPM) [124]

A silicon photomultiplier (SiPM) also consists of an array of SPADs (fig. 4.9d), but here the output of all pixels is connected in parallel in one readout channel (fig. 4.9e). There is a great number of different names for this sensor owed to the large number of manufacturers [125], but the term SiPM has found widespread use [126] [170; p. 14] [515; p. 429] [293; p. 1] [244; p. 47].

An SiPM contains between 100 and 10^5 SPADs (in this context called microcells, cells or pixels), usually with quadratic shape with side lengths of 10–100 µm [127]. This results in total active areas of typically 1×1 to $10\times10\,\mathrm{mm}^2$.
Due to the large number of cells, the probability of a double hit on one pixel is small and the summed output signal is proportional to the number of triggered cells. This permits photon counting with a high dynamic range, which is limited only by the number of pixels and the uniformity of illumination. A linear range of 10^7 photons/s was measured for 15 µm pixels. For high intensities, a strongly nonlinear response is observed due to multiple hits, but the photon number can be restored by applying a correction factor.

[123] Content based on [478; pp. 42, 46–48] [575].
[124] Content based on [478; pp. 13, 48–55] [293] [515; pp. 429–33, 436] [170; pp. 14–17] [194; pp. 308–309].
[125] Hamamatsu, for example, calls them multi-pixel photon counter (MPPC) [477; ch. 3].
[126] Although one might argue that an SiPM and a PMT have little in common, save for their capacity to resolve single photons.
[127] For very small cells with pitches between 5 and 12.5 µm, recently also circular shapes have been used to obtain higher fill factors [478; p. 54].

Each cell consists of a SPAD and a quench resistor (polysilicon strip with $\approx 400\,\mathrm{k\Omega}$), which also decouples the pixels and determines the fall time of the pulses. This simple design trades the in-pixel signal processing capabilities of the SPADA for a higher fill factor.
The peak QE lies between 70 and 80% and the sensitive range extends between about 400 and 1000 nm. The FF from dead space ranges from 20–80% and is lower for smaller cells. The avalanche probability amounts to about 90%. The resulting peak PDE can reach up to 60%.
The bias voltage is typically around 50 V and lies about 10–15% above the breakdown voltage. This results in a gain of about 10^5–10^6. Since the avalanche is in saturation, the amplitude is the same for all pixels. This leads to an excellent charge resolution and an ENF close to 1 (down to 1.03), which allows photon counting up to over 10 pe (see fig. 2.25).
Other than in a single SPAD, the TTS is mainly limited by electronic noise, which causes fluctuations of the threshold crossing time (see eq. (2.188)), and varying lengths of the readout lines between the cells and the output. Compared to SPADs, this leads to higher TTS values of typically 100 ps (down to 70 ps).
Due to the very fast avalanche process, the rise time is only ca. 0.5ns, while the fall time is typically 30 ns to over 100 ns. The recovery time lies below 100 ns.
SiPMs show extremely high DCRs of up to several $\mathrm{MHz/mm^2}$ at room temperature. This is mainly caused by a small percentage of noisy SPADs and can only be improved through better production processes and higher material purity. The DCR shows a strong temperature dependence (decrease by a factor of 2 every ca. 7.5–8 K, DCR levels out below 120–160 K) and at liquid nitrogen temperatures can be as low as several $10^{-3}\,\mathrm{Hz/mm^2}$ for optimized SiPMs [85, 170, 304]. In the last years, the DCR could been reduced to levels of about $30\,\mathrm{kHz/mm^2}$ at room temperature and low overvoltages [304]. However, higher overvoltages have to be used to obtain the optimal PDE, which raises the DCR.
Compared to a SPAD, additional AP effects are possible. Apart from the AP types which occur in the same cell – trapping APs (delays of several 100 ns) and diffused APs (photon absorbed in the neutral region of the same cell) – in addition, photons which are created in the avalanche can also cause optical crosstalk in neighboring cells. If a photon is absorbed in the depleted region of another cell, this results in a near instantaneous second pulse and doubles the output pulse height (direct crosstalk). If the neutral region of a neighbor cell is hit, the carrier drift time leads to a second pulse with a delay of a few ns to several 10 ns (delayed crosstalk). Furthermore, external crosstalk can occur if light from an avalanche is reflected outside the SiPM onto another cell. This light-feedback process can limit the operational range at high gains. The AP rate can be lowered through improved materials, and crosstalk can be reduced by optically isolating cells with trenches, which are etched out and filled with a dielectric. In the last years, AP rates of down to 1–2% and crosstalk of as low as 1% could be achieved for small overvoltages.
SiPMs are insensitive to magnetic fields and can be operated in up to 7 T. In addition, other than PMTs, they are resistant to damage from extreme light levels.

For a large LSND, the required DCR per area is in the range of $0.02\,\mathrm{Hz/mm^2}$ (see table 1.1) and photosensor cooling is likely not an option, as it would introduce convection and distribute radioactive impurities[128]. Even when cooled to −50 °C,

[128] Unless the whole detector is cooled, as in JUNO-TAO (see p. 20).

the DCR of the currently best SiPMs is still a factor of 4000 too high; to comply with LENA requirements, cooling to 128 K would be required. This eliminates SiPMs as possible photosensors for large LSNDs.

Digital SiPM (dSiPM) [129]

A digital SiPM (dSiPM) differs from an (analog) SiPM in that the signals of each cell are summed up digitally (fig. 4.9f).
The output is already digitized in the cell, which reduces electronic noise. To measure the photon arrival time, either one global TDC is used or each cell contains a separate TDC (fully digital SiPM). Furthermore, the cells typically are actively quenched with a transistor and recharged by a second transistor, which results in fast quenching. All components can be integrated on a single chip.
SPADs with high DCR can be turned off, which can significantly reduce the DCR at the cost of a small reduction of the PDE. However, this is offset by the higher DCR which results from the CMOS processes which have to be employed for dSiPMs.
In addition, position sensitivity could be implemented and a correction of the internal transit times could be included to lower the TTS [576].
Compared to SiPMs, however, the extra in-pixel circuits reduce the FF, photon absorption in the additional CMOS layers lowers the QE, and dSiPMs have a higher power consumption.
A TTS of 140 ps was reported for dSiPMs.

In sum, dSiPMs show much potential, but cannot be employed in LSNDs due to the still extremely high DCR at room temperature.

Transistor

Phototransistor [130]

A phototransistor is a bipolar npn transistor where the incident photons create the base–collector current (fig. 4.9h). This results in a significantly higher emitter–collector current and thus causes a current gain. The base–collector junction acts as the photosensitive region and is larger than in a regular bipolar transistor, but otherwise the same manufacturing methods can be used and small voltages are sufficient for operation.
The collector is positively biased versus the emitter, while the base is floating and acts as a potential barrier for electrons from the emitter and as a well for holes. The holes created by IPE are collected in the base, which reduces the electron barrier and strongly increases the electron flow from emitter to collector. Therefore, a small hole current causes a large electron current if the electron transit time through the base is small compared to the minority carrier lifetime. The photoelectrons either flow to the collector or the emitter depending on the impact position; in the latter case, this can minimally reduce the current gain.

[129] Content based on [515; pp. 433–434] [478; p. 48] [293] [304].
[130] Content based on [489; pp. 536–537] [511; pp. 694–697].

The gain ranges from 50 to a few 100 for homojunction phototransistors (junction of materials with equal band gaps but different doping). It can reach 10^4 for heterojunction devices (materials with different band gaps). Since the photocurrent alters the base potential, the gain is not linear, however. In addition, the dark current is amplified by the same gain factor.
The large base–collector area increases the capacitance and limits the response time. For homojunctions this lies in the range of 1–10 µs, for heterojunctions it can be below 1 ns. However, the response time is inversely proportional to the gain. Usually, devices with two leads are used, but a base lead can also be included, which allows to apply a voltage and decrease the response time at the cost of a reduced gain.

Since a high gain – potentially allowing spe detection – and a fast response time are not possible at the same time, phototransistors are not suited for LSNDs.

Photo-Darlington [131]

A photo-Darlington consists of a phototransistor whose emitter is connected to the base of a normal bipolar transistor.
The total gain is approximately the product of the current gains of both transistors. However, the response time is significantly larger than for phototransistors and typically amounts to about 50 µs.

Their slow response excludes these sensors for LSNDs.

CMOS image sensor (CIS) [132]

A CMOS [133] image sensor (CIS) is based on the same technology which is used to manufacture integrated circuits. Like a CCD (fig. 4.10a), a CIS (fig. 4.10b) consists of an array of photodiodes (often p–n), but here additional functions are integrated in each pixel, such as direct read out, amplification, digitization and even digital signal processing. The more structures are included in a pixel, the more insensitive area results per pixel, which lowers the fill factor.

In a passive pixel sensor (PPS), the simplest and earliest form, each pixel only contains a PD and a MOS [134] transistor (select transistor), which lets the collected charge flow to a bus wire.
The active pixel sensor (APS) is the most common type and contains a charge amplifier (a transistor whose gate is controlled by the photocurrent) in each pixel in addition to the structures of a PPS. The collected charge is converted to a voltage in the pixel and sent to the bus wire with a transistor. This provides a far better S/N but lower linearity and uniformity than in a PPS, where the same amplifier is used for all pixels. Additional transistors can be included to serve as shutters or for noise reduction.

[131] Content based on [489; p. 537] [511; p. 697].
[132] Content based on [577] [511; pp. 711–712] [489; pp. 543–545] [478; p. 272].
[133] Complementary metal–oxide–semiconductor
[134] Metal–oxide–semiconductor

In a digital pixel sensor (DPS), moreover an analog-to-digital converter (ADC) and digital signal processing (e.g. gain control) is included in each pixel. Due to the digitization, no signal loss occurs in subsequent stages, but at least five transistors are required per pixel.

Compared to a CCD (see p. 632), a CIS has a considerably higher readout speed (up to several thousand frames per second, unit fps), a larger S/N, and requires lower voltages and significantly less power (up to 100 times less). Furthermore, it is cheaper than a CCD, since inexpensive CMOS technology is used and less external electronics is required. In addition, a CIS is smaller, has no charge transfer losses, and is easier to integrate with additional components (e.g. dark current compensation and signal processing) on a CMOS chip.
However, the noise is higher, the dynamic range is smaller, and the fill factor is lower than in CCDs. The FF can be increased by minimizing the size of insensitive structures, using microlenses above each pixel (which can cause optical crosstalk) or by back side illumination.
As for CCDs, a CIS is not capable of producing a signal or trigger when a photon is detected. Rather, it becomes apparent only after readout (clocked or triggered externally) whether photons were registered in the preceding integration period (frame). Therefore, the TTS is equivalent to the frame length.

Since the sensor is based on Si, it is sensitive in the visible and NIR range with a QE of up to over 90% (cf. fig. 2.8). The pixel pitch is typically 2–30 µm.

Despite the far higher readout speed than in CCDs, generally the TTS of a CIS is still far too high for the use in LSNDs (with one possible exception which is discussed in the IEBCMOS section on p. 646). Furthermore, the ENC of the in-pixel preamplifier is too high for spe detection. This excludes this sensor type for LSNDs.

Scientific CMOS (sCMOS) [135]

A scientific CMOS sensor (sCMOS) is a hybrid between CCD and CIS.
Here, each column is read out with a shift register (see below) and amplified and digitized individually with two channels (high and low gain). This results in low readout noise (1.5–25 e^- RMS) and a large fill factor, like in CCDs, and high readout speed (> 100 fps) as in a CIS.
The QE lies between 55 and 70%. The first sCMOS had a pixel size of 6.5 µm, 5.5 megapixels and a dynamic range of $1.6 \cdot 10^4$ [578].

The large TTS and incapacity for spe detection exclude the sCMOS.

Conversion gain

Charge coupled device (CCD) [136]

A charge coupled device (CCD) is based on a matrix of MOS capacitors (photogates) or photodiodes, which is read out sequentially (fig. 4.10a).

[135] Content based on [478; p. 272] [578].
[136] Content based on [153; pp. 497–501] [577] [489; pp. 543–545] [478; pp. 270–272] [508].

Each pixel contains a potential well that collects the electrons which are produced by IPE near the surface. After an exposure period (frame), the charges of all pixels are usually transferred to a light insensitive array of cells (transfer gate). The charges are then repeatedly moved from one row to the next by applying a voltage pattern to the electrodes which shifts the potential wells (shift register). The last row is moved onto another shift register, which serially transports the charge of each pixel onto a charge amplifier. This produces a voltage, which is digitized by an ADC. The CCD thus does not produce pulses in response to incident photons but integrates the signals over a fixed exposure time. Thus, the time resolution is equivalent to the exposure window.
The shift registers typically have three cells per pixel to avoid charge mixing during the transfer. The transfer efficiency between wells is very close to unity. By using a transfer gate, the pixels can be exposed to light during readout, which is read out in the next frame. One can also directly use the pixels as a shift register, but this requires to use a shutter to avoid illumination during charge shifting and introduces a dead time. To speed up readout, the charges can be extracted separately from each quadrant of the CCD by using four readout nodes instead of one.
The surface structures reduce the fill factor, which can be increased by back side illumination through a substrate with reduced thickness.
The pixels can be easily manufactured with standard integrated circuit techniques. Only one amplifier, ADC and output node are required, so the linearity and uniformity is better than for CMOS sensors (p. 631) and less pixel defects occur.

Since the device is based on silicon, it is sensitive in the visible and partially the IR range and the QE in the visible spectrum can be around 80–90% (fig. 2.8). The pixel pitch is between 2 and 30 µm and the whole structure is a square with a few cm side length and several 100 µm thickness.
The pixels itself have no gain, but the readout amplifier can have very low capacitances of 0.03–0.1 pF, which leads to a low readout noise (down to 2–3 e^- RMS) and allows the detection of down to several 10 photons with cooling [170; p. 17]. For such a low readout noise, however, large amplifier shaping times of ca. 10 µs per pixel are required, which limits the readout rate and thus the time resolution. The frame rates typically lie between 3 and 16 fps [578].
The thermal excitation of electrons in the pixels causes a dark current, which contributes to the noise and can be reduced through cooling.
The dynamic range depends on the number of electrons which can be stored in a pixel (full well capacity) and thus is connected to the pixel size.

Since spe detection is not possible and the TTS is extremely high, CCDs are not suited for LSNDs.

pnCCD

In a pnCCD, p–n diodes are used in place of MOS photogates.
In this case, the pixels are typically larger (75–150 µm edge length), are depleted over the complete thickness [137] (several 100 µm), and have a higher radiation hardness [153; p. 498] [579].

These devices are excluded for LSNDs for the same reasons as CCDs.

[137] This is beneficial for particle detection.

Skipper CCD [138]

A skipper CCD employs a different readout technique and measures the charge non-destructively with the floating gate of a transistor (skipper amplifier). Therefore, this process can be repeated thousands of times and the readout noise can be substantially reduced through averaging, at the cost of a significantly longer readout time. Since only the readout structure is altered, this technique can be implemented for a conventional CCD without notably affecting the design.

In [580] a fully depleted skipper CCD was reported which achieves a readout noise of as low as $0.068\,e^-$ RMS at 140 K. This allows single photon detection and unambiguous photon counting with a linear range of over 1500. Cooling is required to reduce the dark current, which can be the limiting noise component. For a single sampling the readout noise was $3.55\,e^-$ RMS with a readout time of 10 µs. To achieve a readout noise below $0.1\,e^-$ RMS, ca. 1200 samples were required, which raises the readout time to 12 ms per pixel and 3 h for the whole sensor of 4126×866 pixels with four readout nodes. The authors aimed at a use in dark matter, coherent neutrino-nucleus scattering and astronomical experiments.

Skipper CCDs are an extreme example of the tradeoff between low detection threshold and fast readout which is encountered in sensors with charge amplifiers (e.g. CCD, QIS, IEBCCD, EBCCD, HPDs, HAPDs). Here, this is further aggravated by the serial readout and the repeated sampling. The spe capability comes at the price of a huge TTS, which firmly rules out skipper CCDs for LSNDs.

Quanta image sensor (QIS) [139]

A quanta image sensor (QIS) is basically a CIS which uses an extremely high conversion gain to achieve photon counting, has very small pixels, and employs an adapted readout method.
The target is to obtain sub-diffraction limit pixels (smaller than the wavelength) with a standard CMOS process to allow further pixel miniaturization. Furthermore, with regular CMOS processes it is possible to achieve comparatively high QE, low dark current, fast readout speed, low power dissipation, and low production costs. Since the full-well capacity and thus the dynamic range of a pixel shrinks with smaller sizes, this has to be compensated by a larger number of smaller pixels with photon-counting capability. The light flux then can be reconstructed statistically and in this way the total dynamic range can even be increased significantly.
In the QIS, the pixels are called jots (Greek for "smallest thing") and consist of a pinned PD which collects the charge that is created by IPE during an integration period. During readout the charge is transferred to a floating diffusion with a very small capacitance. This results in a very high conversion gain between the collected charge and the produced voltage (see eq. (4.1)), which is sufficient to measure single photons. The full-well capacity of a jot can range from about 1 to 200 and the number of collected electrons is digitized in the pixel output as either single-bit (photons detected yes/no) or multi-bit (number of photons). In a single-bit QIS,

[138] Content based on [508] [580].
[139] Content based on [510, 581] [478; p. 48].

the output of all pixels is a binary bit plane. To reconstruct the image, spatial and temporal binning is performed, i.e. groups of jots are binned to form pixels and a series of subsequent bit planes is summed up. This requires high-speed readout to minimize the probability of collecting more than 1 pe per jot.
The read noise is dominated by the noise of the in-pixel source follower, where the major contribution is 1/f noise, which probably stems from carrier trapping and re-emission at the Si–SiO_2 interface. In order to enable photon counting without cooling the sensor, the readout noise has to be smaller than $0.5\,e^-$ RMS. At this value, the npe distribution is nearly continuous and the counting error rate [140] is over 10% for low occupancies. Below $0.15\,e^-$ RMS, the npe distributions do not overlap and photon counting with $< 0.1\%$ error is possible. This can be achieved by either reducing the output voltage noise or by increasing the conversion gain. In the QIS the latter strategy is used by minimizing the capacitance of the floating diffusion.

In [510] a QIS was reported, which had 10^6 jots at 1.1 µm pitch and was based on a commercial CMOS process. This achieved a mean read noise of $0.21\,e^-$ RMS at room temperature and thus was capable of photon counting. The conversion gain was as high as $368\,\mu V/e^-$, the readout rate was 1040 frames per second, the QE was 70–80% between 450 and 650 nm, and the mean dark current was 0.16 Hz/jot.

While the QE and spe charge resolution of this device are excellent, its TTS is not sufficient for LSNDs.

Conservation gain

Photoresistor [141]

A photoresistor (also called photoconductor) shows a significantly reduced resistance when illuminated (fig. 4.9g).
In this sensor, electrodes with ohmic contacts are placed on both ends of a semiconductor, for which most commonly CdS and CdSe are used. The device requires a power source, since no photocurrent is generated. Donor and acceptor levels reduce the mobility of thermally created electrons and holes without illumination, which leads to a high resistance and a small dark current. Incident photons excite electrons into the CB and the created electron-hole pairs increase the conductivity. An additional acceptor level catches the holes, which reduces their mobility and the recombination probability and thus increases the carrier lifetime. The applied voltage draws the charge carriers to the electrodes and creates a current. When the electrons reach the anode, the holes still remain inside the sensor, so to conserve charge neutrality, for each extracted electron one is inserted at the cathode. The current continues to flow and decays with the carrier lifetime while the electrons go through the detector many times. This creates a gain of 1–10^6 (typically 10^3).
The gain is maximized for short, wide sensors, which can be achieved by a serpentine shape of the gap between electrodes (right in fig. 4.9g). The current pulse duration can be adjusted by artificially reducing the lifetime with dopants or radiation damage,

[140] The probability to reconstruct the wrong number of detected photons.
[141] Content based on [489; pp. 538–540] [194; p. 392] [511; pp. 666–671] [582; ch. 10].

which can provide fast timing at the cost of a lower gain. Response times can also be shortened by a small electrode spacing and a high field, and range between 10^{-3} and 10^{-10} s.
Photoresistors are outperformed by PDs in many use cases, but are frequently used for IR detection; CdS, however, is sensitive only below 515 nm.

Since the gain of photoresistors is very low for acceptable rise, fall and recovery times, these sensors are unsuited for LSNDs.

4.3 External photoeffect sensors

4.3.1 Vacuum-based

Unamplified

Phototube [142]

A phototube [143], invented as early as 1893, is an electron tube which contains a photocathode and anode (also [173; p. 129]).
Usually, a vacuum tube is used, which results in a photosensor without gain whose output current is proportional to the light intensity (called vacuum phototube or simply phototube here).
However, there are also devices which use a gas-filled phototube to obtain a gain through gas multiplication. Due to the different detection mechanism, this is treated as a different sensor type (see p. 693) in this work.
In addition, the term phototube is used as a class name for all photosensors which are based on the components of a vacuum phototube and contain additional structures, such as dynodes or secondary sensors. This comprises all sensors branching off from the phototube in fig. 4.1 and is why the PMT is called a photo *multiplier* tube: It is a phototube which multiplies the photoelectrons before readout.

A variety of photocathode and window materials can be used and the QE can reach up to about 25%. The applied voltages are small (about 15–50 V), since they only have to facilitate pe collection on the anode, and the dynamic range is large (also [586]).

This sensor is not capable of spe detection and is thus irrelevant for LSNDs.

Discrete reflection dynodes

Vacuum phototriode (VPT) [144]

A vacuum phototriode (VPT) is a PMT with only a single dynode. The photoelectrons from a cathode on an UV-glass window pass through the anode mesh in front of the dynode and create SEs on the dynode, which are collected on the anode.

[142] Content based on [583] [584].
[143] Also called photoelectric cell, photocell, photoelectric tube or vacuum photodiode [584] [169; p. 10] [585; p. 174].
[144] Content based on [515; pp. 419–420] [327] [587].

Since the transmission probability of the anode mesh is typically 50%, a large fraction of photoelectrons is not amplified. Furthermore, a significant number of SEs is not collected on the anode but hits the dynode a second time, which results in a small SEY of about 0.5–0.6, since the impact energy is the SE emission energy (a few eV). Both effects reduce the gain. For voltages in the order of 1 kV, a total gain of about 10 results and a large ENF of about 2–3 is observed.
The main advantages of VPTs are that they can be used in high magnetic fields of about 2 T, where PMTs are unsuited, and that they are radiation hard, which allows to use them in calorimeters in place of APDs.
The peak QE is typically 20–25% with a sensitive range between 300–550 nm, the TTS σ_T is about 1 ns and the sensitive area is below 50 cm^2.

VPTs can also be configured as image sensors. In [588] a prototype was studied by Hamamatsu which had a transmissive fine mesh dynode and a 4-segment anode behind it.
In [172; p. 1-22] VPT types with segmented dynodes are mentioned. Here, the signal is read out between the dynode segments and the common anode. A strong axial magnetic field significantly reduced the crosstalk between channels.

The very low gain makes VPTs incapable of spe detection and thus unsuited for LSNDs.

Photomultiplier (PMT)

The properties of photomultiplier tubes (PMTs, see fig. 2.1) were already discussed at length in ch. 2 and were summarized for large modern PMTs with B+L and LF dynodes in ch. 3. The properties which can be achieved with other dynode types were elaborated on p. 577.

PMTs are the sensor of choice for LSNDs, WCDs and many other applications. They head the short list, and all other photosensors are compared to their properties.

Micro PMT [145]

Micro PMTs are miniaturized PMTs (fig. 4.3e) with total dimensions of only a few mm. This is possible by fabricating the electron multiplier (dynodes and anode) with micromachining techniques (micro-electro-mechanical systems, MEMS) on a Si wafer. The Si substrate is then sandwiched between two glass plates, on one of which the cathode is placed. This manufacturing technique allows mass-production and the resulting structure is highly shock-resistant, which allows the implementation in portable devices.
The principle of operation is exactly as in a PMT: First, a photocathode converts photons into photoelectrons. These are accelerated sideways into the electron multiplier where subsequent EB on the dynodes produces a gain through SE production. Finally, the cascade is read out by an anode.

With a bialkali cathode, a spectral response from 300–650 nm with a peak QE of typically 26% at 350 nm is obtained. For multialkali the sensitivity extends from 300 to 850 nm. At 900 V a gain of $2 \cdot 10^6$ results. Due to the small cathode area, the dark current is only 0.3 nA. The T_{FWHM} [146] is 1.3 ns and in the TTD a long trailing flank

[145] Content based on [169; pp. 203–210] [267].
[146] TTS using the FWHM of the main peak in the TTD.

up to a lateness of ca. 9 ns is observed. The rise time amounts to 1.2 ns and the fall time to 1.7 ns. Micro PMTs are nearly insensitive to the geomagnetic field (GMF) and in the most sensitive direction a reduction of the output to 50% is found at 1.4 mT.

These properties are very favorably, but the low FF (here, the fraction of sensor surface occupied by the photocathode) and the small size (would require an extreme amount of channels) prohibit a use in LSNDs.

Continuous reflection dynodes

Channel electron multiplier (CEM)

This photosensor was already described on p. 581.
The dynamic range of only 1 pe (binary sensor) and the long dead time exclude CEMs for LSNDs.

Micro-channel plate PMT (MCP-PMT) [147]

In a micro-channel plate PMT (MCP-PMT), the photoelectrons are accelerated by a strong electric field from the photocathode onto a stack of one to three MCPs (see p. 582) in which they are amplified through secondary emission (fig. 4.11a). The generated cascade is read out with an anode, which can be segmented to obtain position sensitivity. Generally, proximity focusing with distances of a few mm is used between cathode and first MCP and between last MCP and readout. The first visible light MCP-PMT prototypes were developed in the 1970s, and commercial devices became widely available at the end of the 20$^{\text{th}}$ century.

Ionization of residual gas produces IBF to the photocathode, which originates mostly from near the capillary exits, where the electron density is highest. This can create afterpulses and damage the cathode, which reduces the sensor lifetime. It also limits the gain in an MCP with straight channels to 10^3–10^4, which is not sufficient for spe detection.
To reduce IBF and achieve higher gains, several MCPs with inclined straight channels (typically 8–10°) can be stacked with alternating capillary directions. The "chevron" configuration uses two MCPs (fig. 4.11a), the "Z-plate" three. Alternatively, a single MCP with curved capillaries can be used ("C-plate"). The Z-plate (gain about 10^8) achieves a significantly better IBF suppression than chevron stacks (gain ca. 10^7), since ions are trapped at two MCP interfaces rather than one, and can reduce cathode aging by a factor of ten. C-plate MCPs are mechanically most stable, have the lowest charge resolution, and provide the best IBF reduction. However, they have lower gains (order of 10^6), proved to be extremely difficult and expensive to fabricate, and are currently not commercially available. Therefore, Z-plates are most commonly used to obtain high gains.
A further method to reduce IBF is to place a thin aluminum oxide film of 5–10 nm thickness in front of the first MCP stage. This prevents ions from reaching the cathode (ion barrier) but reduces the CE by ca. 40%. To avoid this loss, the ion barrier can be placed between two MCP stages.

[147] Content based on [196; pp. 282–292] [515; pp. 420, 436] [169; ch. 11] [589; pp. 398, 399] [194; pp. 302–303] [267; pp. 48–50]. [589; ch. 22] provides an excellent review.

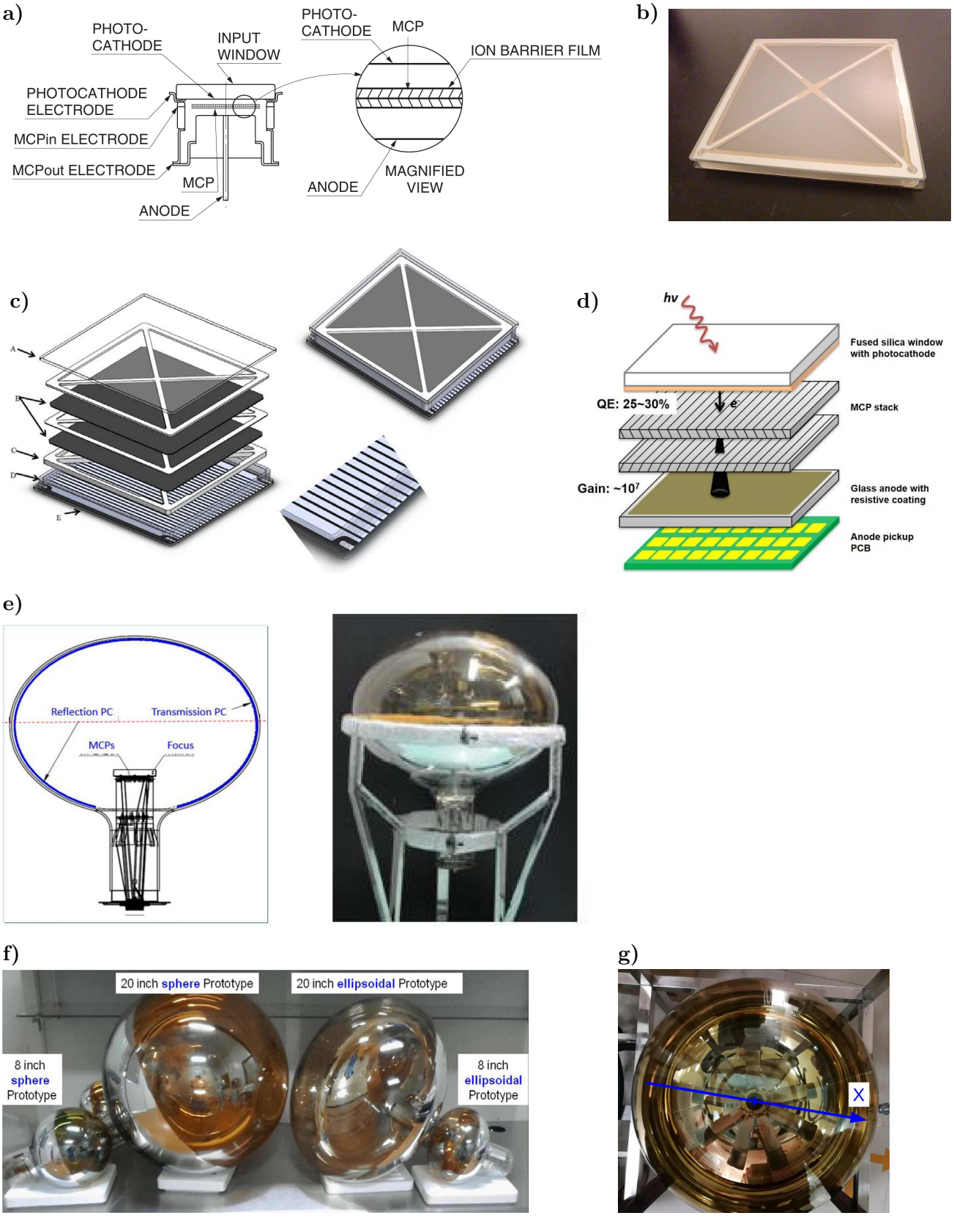


Figure 4.11: Micro-channel plate PMT (MCP-PMT) variants. Discussed in text.
a) MCP-PMT, from [169; p. 213].
b)–d) Large Area Picosecond PhotoDetector (LAPPD): b) Photo, from [550]. c) Layout (GEN I), showing the window (A), micro-channel plates (B), spacers (C), anode with strip lines (D) and side walls (E). Assembled LAPPD in top right. Close-up view of the strip line exit in bottom right. From [497]. d) Argonne LAPPD with capacitively coupled anode pads, from [590].
e)–g) Fountain focused MCP-PMTs: e) Layout of the NNVT GDB-6201 for JUNO (see p. 18), from [591]. f) NNVT MCP-PMT prototypes for JUNO. The elliptical 20" MCP-PMT is the GDB-6201 [346]. From [592]. g) NNVT GDB-6203 for Hyper Kamiokande. This sensor is based on the GDB-6201 with an additional asterisk-shaped focusing electrode to improve the transit time spread (TTS). From [593].

In addition, IBF can be lowered through a better vacuum and by cleaning the MCP surfaces with electron scrubbing.
Recently, also atomic layer deposition (ALD) has been used in MCP fabrication. This allows to sequentially deposit atomic mono-layers through alternating exposure with different gases, since the gases form a film on the material surface, where they chemically react with each other. This improves the SEY and gain and prevents contaminants in the MCP glass from evaporating, which drastically improves aging.

The spectral response of MCP-PMTs usually ranges from 300 to 550 nm with a peak QE of up to 25%. Diameters lie between ca. 18 and 150 mm.
The gain typically is between 10^4 and 10^7, and in saturation mode the SER shows a distinct spe peak with only few underamplified pulses.
Due to the high voltage of several kV, the small distance of cathode and anode to the MCP stack, and the short length of the MCPs, the transit time is very small (down to 0.5–1 ns). For the same reasons, the field lines are almost parallel between cathode and MCP, and the pe emission angle and velocity are nearly negligible. These properties lead to an excellent TTS of as low as ca. 20 ps. The factors which still limit the TTS are lateral pe momentum, backscattering on the MCP input surface, rise time (jitter from noise), and time walk due to charge sharing between anode pads [545]. If NEA cathodes (GaAs, GaAsP) are used (see p. 65), the TTS is notably larger due to the photoemission lag (see p. 156). In the TTD of a Hamamatsu R3809U-50, several LP types can be observed [169; p. 217]. One type with a lateness of up to ca. 450 ps and several percent probability likely is from backscattering on the MCP surface (Lm) [148]. LPs with a short lateness < 100 ps could originate from L1 or L11. Furthermore, also secondary LPs are found (e.g. LmLm). Regarding the pulse shape, the pulse width amounts to a few 100 ps. The Hamamatsu R3809U-50 has a rise time of 150 ps, a fall time of 360 ps and a pulse width of 300 ps for 550 ps transit time and 25 ps T_{FWHM}.
MCPs preserve the impact position information, where the accuracy is determined by the channel pitch. This allows to achieve very small position resolutions depending on the employed readout method. With XS anodes the position can be reconstructed from the charge centroid of several strips. In this case, the strip pitch and the stack–anode distance need to be optimized so that several strips are hit by the cascade. For a pitch of 0.7 mm and 2.5 mm stack–anode distance, a spatial resolution of about 17 µm FWHM at $1.4 \cdot 10^6$ gain was reached.
MCP-PMTs are insensitive to magnetic fields and can be operated at up to 0.5–2 T for axial fields, whereas for fields perpendicular to the MCP axis a sharp drop in the output current is observed above 70–100 mT.
The channel recovery time is relatively long (between 10 and 20 ms) and the lifetime is low compared to other photosensors.
MCP-PMTs are mostly used as a gain stage in image intensifiers (here the anode is replaced by a phosphor), e.g. in night vision devices, and can be gated by including a mesh with switchable potential between cathode and MCP stack.

There are various subtypes of MCP-PMTs which are based on different readout methods or alternative focusing types. For example, the H33D detector [507, 549] here is considered an MCP-PMT with XDL or XS anodes, which are read out by a field programmable gate array (FPGA). This sensor is not further addressed, since

[148] Please refer to 2.3.5 for the LP nomenclature.

its size, QE and readout rate are insufficient. Instead, two MCP-PMT variants which are of direct relevance for LSNDs and WCDs will be covered subsequently.

Regarding LSNDs, MCP-PMTs have excellent TTS, high linearity, very good magnetic field immunity, and can be made sufficiently large. However, their PDE (QE 25%, FF of the MCP 50–60%) is below the exclusion limit of 20%. Also, aging and ionic afterpulses (iAPs) can be problematic, so ALD coated MCPs should be used. Furthermore, lead-glass based MCPs are very expensive. MCP-PMTs are therefore not included in the short list. This exclusion does, however, not extend to the fountain-focused and LAPPD subtypes, whose layout is quite different. These variants require a separate analysis.

Large Area Picosecond PhotoDetector (LAPPD)[149]

The Large Area Picosecond PhotoDetector (LAPPD) is an 8×8" square MCP-PMT with stripline readout, which is based on a novel MCP substrate with ALD coating and a significantly lower cost per area (4.11b and fig. 4.11c). This results in a large photosensor with several 10 ps TTS and mm position resolution. The LAPPD was developed for nuclear and high energy physics and Cherenkov detectors.

First generation LAPPDs (GEN I) are housed in an all-glass hermetic package with a fused silica/borosilicate glass window, a borosilicate bottom plate, and borosilicate side walls (size 230×220 mm², height 22 mm). This contains an Na_2KSb high-temperature bialkali cathode (see p. 55) with 195×195 mm² active area, two stacked MCPs in chevron configuration, and silver microstrip anodes on the bottom plate. Window, MCPs and anodes are separated by X-shaped spacers to provide uniform distances and mechanical stability for the flat vacuum tube. The spacers reduce the effective area to 91%. The striplines are operated as double readout delay lines, which allows to reconstruct a continuous 2D position from the time difference along the strip and from a charge centroid across adjacent strips [596]. Using striplines is more economic than pads, since the number of readout channels scales linearly with sensor size instead of quadratically.

Currently, second generation devices (GEN II) are being developed. Here the striplines are replaced by a thin metal plane in order to capacitively couple external anodes on a PCB through the bottom plate (compare fig. 4.11d). This allows to use customized anode shapes (pads, lines) and is expected to significantly reduce costs.

The MCPs in the LAPPD are not based on lead-glass, since this would lead to reduced lifetimes through IBF, has limited size and thickness due to fragility, and is very expensive ([590] lists a price of $12k for a 53×53mm² Photonis MCP-PMT). Instead, a glass capillary array developed by Incom is used. [597] This is created by first drawing hollow glass tubes to a diameter of 20 µm. The tubes are then fused to large blocks, from which MCPs with 203 mm side length and 1.2 mm thickness are cut. In this method, it is not necessary to remove the contents of the capillaries through etching, and thus MCPs with large areas, 20 µm channels and open-area ratios above 65% can be produced at low costs – a cost reduction by at least a factor of 10 is expected. Open-area ratios of 60–74% were reached [547, 598, 599] and up

[149] Content based on [497, 550, 594, 595].

to 85% are considered possible [600]. These MCPs are then coated through ALD with separate layers of first a resistive material and then a secondary emitter. This provides high gain and – unlike in lead-glass MCPs – the substrate, resistive and emissive materials can be chosen independently.

The LAPPD was commercialized by Incom in 2014 and the first fully functional prototypes were manufactured in 2017. While optimizations are ongoing (e.g. for the QE), the sensor is now reaching the point where properties can be consistently reached and has entered the stage of pilot production. About 60 devices have been fabricated so far and the fabrication capacities allow to produce several 10 sensors per year.

At the moment, GEN I LAPPDs achieve a QE of $25 \pm 2\%$ at 365 nm and ca. 21% at 420 nm (with fused silica window 30% at 365 nm and about 26% at 420 nm) with a typical uniformity of $\pm 8\%$. The active area amounts to 350 cm^2, which makes them the largest proximity focused MCP-PMTs produced to date.
The gain ranges from 10^6 to 10^7 at 900–1000 V per MCP and 200–400 V between cathode and MCP stack. The P/V is typically between 1.5 and 2. Up to 4 coincident photons per strip can be distinguished algorithmically [601] and a dynamic range of 500 kHz/cm^2 at 10% gain drop (resulting from MCP recharge) is reported.
The σ_T lies between 50 and 70 ps (resolution possibly limited by electronics) and is expected to improve in the future with smaller 10 µm pores. The TTD shows several LP effects, likely Lm (0.8 ns lateness, $\approx 1.8\%$ probability), Λm, LmLm (up to 1.7 ns), and LPs with short lateness (L11 or L1). The total LP probability is about 4%. In addition, some EPs with an earliness of up to 400 ps are present. These might originate from photoelectrons whose incidence angles match the channel tilt and which thus hit the channel at a later position or even skip one MCP. This would result in an earlier and smaller pulse. Since it is the defined goal of the LAPPD to reach the smallest possible TTS, it should prove beneficial to apply TST to this sensor to calculate occurring time shift effects (TSEs). This would allow to adapt the design to minimize these effects.
Regarding the pulse shape, a rise time of 0.85 ns and a pulse width of 1.1 ns was found.
The spatial resolution along strips is 2.4 mm, across strips 0.76 mm (both RMS).
The DCR amounts to a few 100 Hz/cm^2 at $5 \cdot 10^6$ gain and to several 100 Hz/cm^2 at 10^7 gain when using a threshold of $8 \cdot 10^5$ gain (0.16 and 0.08 pe). For high cathode and MCP voltages, a sharp upturn of the DCR is observed, which is likely due to feedback of some sort (ions, photons).

The 8×8" LAPPD has an excellent TTS and the EP and LP probabilities are within the allowed limits (the LPs also have a short lateness). However, the P/V (1.5–2) so far is barely sufficient or below the required value of 2. The DCR (several 100 Hz/cm^2) is far too high, which is due to the multialkali cathode, but might be low enough if a K_2CsSb cathode were used. The PDE of 14–18% at 420 nm (QE 21/26% and FF ca. $91\% \cdot 74\%$) is still too low. Due to the fill factors from MCP and spacer, high quantum efficiency (HQE) cathodes with at least ca. 30% QE would be required to fulfill the target PDE of 20%. However, the shape of the sensor allows a high optical coverage (OC) close to 100% in a LSND, which can compensate a lower PDE. Finally, at the moment only several ten sensors can be fabricated per year.

Thus, currently the properties of LAPPDs are not sufficient for LSNDs. Nevertheless, the LAPPD is included in the selection, since this sensor is very promising and is still being developed.

Argonne LAPPD[150]

The Argonne National Laboratory is developing a variant of the LAPPD with the target to obtain a low-cost pixelated photosensor with extremely small TTS for particle identification in high magnetic fields and high radiation.
The Argonne LAPPD (fig. 4.11d) has a size of 6×6 cm² with a fused silica window for the detection of Cherenkov light down to 180 nm. To obtain better TTS and magnetic field tolerance, newly developed smaller 10 µm pores are used and the spacing between bialkali [598] cathode, MCPs and anode is reduced. As for the GEN II LAPPD, a capacitively coupled external readout with a resistive anode (in place of the striplines) and pixels on a PCB is used. This allows to adapt the pixel size and pitch.

The QE is 14.3% at 370 nm (12.5% at 420 nm).
A gain of $3.1 \cdot 10^6$ for 10 µm pores and $1.4 \cdot 10^7$ for 20 µm was reported.
For 10 µm pores, a TTS of 35.8 ps (σ_T) and an RMS over the whole TTD of 83.1 ps were measured, compared to 20 µm pores with $\sigma_T = 68.4$ ps and a full TTD RMS of 132.1 ps. For 20 µm pores, a significant fraction of pulses are LPs with delays up to 500 ps and the TTD can be described by two Gaussians; for 10 µm the LP probability is greatly reduced.
The rise time was 0.54 ns for 20 µm and 0.44 ns for 10 µm, the fall time was 1.2 ns for 20 µm, and the pulse width was 0.8 ns for 20 µm.
For highly energetic particles, a position resolution (σ) of 0.7 mm could be reached for 2×2 mm² pads with direct signal pickup. With capacitive pickup and 4×4 mm² pads a value of 0.8 mm was found.
With 10 µm pores and a reduced internal spacing, the gain dropped to 50% only at about 0.64 T and remained above 10% of the zero-field value at 1.5 T. This could be further improved by using 6 µm pores, which are currently being developed by Incom.

The Argonne LAPPD has excellent TTS, a low LP probability and is tolerant to high magnetic fields. On the other hand, its current QE is too low and the sensor is quite small (6×6 cm²) which would require many channels or local front-end electronics. Since the higher possible OC balances the QE deficit, it is included in the shortlist.

Fountain-focused MCP-PMT

Between 2009 and 2015 a new MCP-PMT subtype which uses *fountain* focusing instead of proximity focusing was developed by the JUNO experiment and NNVT with the goal to obtain a large photosensor with higher PDE than a PMT of the same size.[151] This led to the commercialization of the 20" diameter (508 mm) GDB-6201 by NNVT of which 15000 sensors were ordered as one of three photosensor types for JUNO.

[150] Content based on [590, 598].
[151] Paragraph based on [250–252, 602].

In this sensor, the entrance window is coated with a transmissive photocathode, while the back of the glass bulb is covered with a reflective cathode instead of an Al mirror as in PMTs (fig. 4.11e). The photoelectrons are focused onto a stack of two MCPs (ALD). This layout takes advantage of the higher QE of reflective cathodes (see p. 65), expected at 40–50%, and the double cathode effect (p. 64) to target a PDE of about 30% at low costs, even without HQE transmissive cathodes [250].
To achieve this goal, a high CE is essential. In initial prototypes (left of fig. 4.11f), two back-to-back MCP stacks were used in a spherical bulb. Here, photoelectrons from the top and bottom (or left and right) hemispheres were collected on different MCP stacks with nearly 4π collection angle. However, this resulted in a CE of only ca. 50% and was abandoned in favor of an ellipsoidal bulb with only one MCP stack (fig. 4.11e) which increased the CE to 65% [251].
Apart from the electron optics, the CE is limited by the open-area ratio of the MCPs, which is difficult to raise above 70%, since photoelectrons which hit the MCP surface are lost and produce no pulse (Tm). Therefore, the open-area ratio acts like a fill factor. By coating the top electrode (Ni-Cr) with a layer of the secondary emitter material Al_2O_3, incident photoelectrons can produce SEs that are focused into neighboring channels and start a cascade, which eliminates the losses. This raised the CE from about 60% to 98%. However, this came at the cost of a significant deterioration of the TTS from 12 to 20 ns FWHM, since this amplification process (Rm) produces LPs and occurs for over 30% of the pulses. Photonis has also used coated MCP surfaces for proximity-focused MCP-PMTs to improve the CE from about 65% to close to 100% and observed a slight degradation of timing [516].
Furthermore, in the GDB-6201 a gap with a reverse voltage was added between the two MCPs to increase the gain and improve the charge resolution, since this rejects peripheral low energy electrons. Introduction of the coated MCP surface and the reverse-bias gap allowed to increase the QE, PDE and P/V but also raised the TTS, DCR, rise time and fall time. In addition, during manufacturing the peak QE was improved from 30 to 35% by applying some of the methods described on p. 64.
In marked contrast to proximity focused MCP-PMTs, the fountain-focused layout is *not* position sensitive anymore and the resulting TTS is very *large*. The latter is, on the one hand, due to the pronounced CTS, which results from the 4π pe collection and a relatively low voltage of 400 V between the cathodes and the MCP stack [252, 591]. In addition, Rm and backscattering from the MCP surface (Lm and Λm) produce a large amount of LPs, which also raises the TTS.
The final version of the GDB-6201 (values from [603, 604]) showed a mean QE of 31.3% at 400 nm (PDE 30.9%), a gain of 10^7, an R_σ [152] of 30.9%, and a P/V of 7.1 at a voltage of on average 1737 V. The T_{FWHM} was 20.0 ns and the TTD showed a Gaussian with a strong, broad right flank from LPs which overlap with the main peak. [252] When the MCP surface was coated, this flank became even more pronounced and longer and produced lateness values of up to about 22 ns after the peak. The rise time amounts to 1.44 ns, the fall time to 23.5 ns, the DCR is 37.8 kHz, and the dynamic range (10% deviation) is 1400 pe. An AP number–charge rate (see 2.5.2.1) of 1.0% was measured, which was distributed on five main clusters at 215, 841, 3070, 4430 and 16600 ns with smaller peaks in between.

In [517, 605] possible TTS improvements of the GDB-6201 were discussed based on simulations. By increasing the diameter of the focusing electrodes, the extraction

[152] Charge resolution based on the standard deviation in the SER, see eq. (2.64).

fields near the cathodes can be increased, which reduces the transit time variations from emission energy and angle to < 3 ns for the whole cathode. A CTS of below 1 ns was found up to 70° from the surface normal, but then a steep drop occurs down to −35 ns transit time difference vs. the center at 130°. For this reason, the authors suggested to coat the bottom hemisphere with a mirror instead of a reflective cathode. In addition, an alternative cambered shape of the Ni-Cr top electrode which results in a bowl shaped channel entrance was studied. When this electrode is coated by a high-SEY material as in the GDB-6201, the SEs are focused into the adjacent channels which resulted in a CE of 98.8% with less than 2% LPs from pe backscattering. Through these measures the simulated TTS was lowered to under 3 ns.

Based on the GDB-6201, the Hyper-Kamiokande collaboration and NNVT developed a variant with improved timing (GDB-6203) by adding an asterisk-shaped [153] focusing electrode (see fig. 4.11g). [154] This significantly lowered the CTS and thus reduced the T_{FWHM} from 11.5 to 4.3 ns, but also reduced the CE from 98% to 85%. A QE of 30% at 400 nm was measured [607], the charge resolution was 40% (R_σ) at 1900 V and the gain was found to be constant within 5% for magnetic fields of up to ± 100 mT.

The GDB-6201 shows very good QE, CE and PDE values, its charge resolution is acceptable, the P/V is very good, the linearity is very high, and a low AP rate is found. The DCR of 18.6 Hz/cm² (cross-section) is higher than the allowed 2.16 Hz/cm², but might be sufficiently low after long times in dark. The LP probability is large but the LPs have a comparatively low lateness. However, the TTS is far too high for LSNDs that focus on a low time resolution, which excludes this sensor for LENA. Nevertheless, the TTS and LP rate could be significantly improved by an adapted layout and cathode shape and then might meet the LENA requirements. Thus, the GDB-6201 is included in the short list for reference.
The GDB-6203 has a good PDE of about 25.5%, an acceptable charge resolution and high magnetic field tolerance. Its TTS of 4.3 ns is far lower than the GDB-6203 but still higher than the LENA requirement of 3 ns FWHM. Since this value is not far from the target and the properties are otherwise good, it is included in the selection to allow a comparison with other sensors.

Hybrid micro-channel plate photodiodes (HM-PDs)

A hybrid MCP photodiode (HM-PD) is a phototube which contains a stack of MCPs whose SEs bombard a photodiode matrix, either a CCD (IEBCCD) or a CIS (IEBCMOS). The combination of both gain steps is sufficient to create an imaging photosensor which is capable of photon counting [566]. The slow readout speed of the secondary sensor can limit the utility of HM-PDs for LSNDs, however.

Intensified electron bombarded CCD (IEBCCD) [155]

In an intensified electron bombarded CCD (IEBCCD), a vacuum tube contains a photocathode whose photoelectrons are proximity-focused onto an MCP (fig. 4.10e).

[153] In order to fit through the stem with 90 mm diameter.
[154] Content regarding GDB-6203 based on [251, 279, 346, 593, 606].
[155] Content based on [567, 608].

The produced SEs are accelerated (also proximity focusing) onto a back-thinned CCD with EB gain. The photosensor thus is like an EBCCD (p. 675) with an additional MCP between cathode and CCD to provide an extra gain step. This allows to achieve spe detection with a lower bombardment voltage than in EBCCDs or ICCDs (p. 648), and also a smaller number of stacked MCPs than in ICCDs is sufficient. Through this, the amount of IBF – visible in form of intense white spots in the images – is significantly reduced. This is aided by using an MCP with tilted channels (see p. 582).
Compared to a CCD, the MCP limits the achievable spatial resolution, but less than in an ICCD, where an additional electron–photon conversion takes place in a phosphor screen. The PDE is lower than for an EBCCD due to the FF of the MCP (open area ratio) [566].
Achievable readout rates are limited by the CCD frame rate, although a binned or sub-array (only a fraction of pixels) readout can be used to obtain higher frame rates. The exposure time, on the other hand, can be in the ns or ps range, since the MCP can be gated.
The IEBCCD is particularly suited for single molecule fluorescence imaging (as of 2007) due to its low multiplication noise and the lack of iAPs.

To the author's knowledge, only a single such sensor has been developed by Hamamatsu [567]. This device consisted of a ceramic tube with a glass window and a GaAsP cathode with a peak QE of up to 54% at 500 nm (ca. 44.5% at 420 nm). A single MCP with 6 µm channel diameter, 240 µm thickness, an open area ratio of 65% (resulting in a maximum PDE of 35%) and a gain of 12–75 at 600–700 V bias voltage was installed. The CCD had a diameter of 1", 640×480 pixels (20 µm side length, full well capacity 120 ke$^-$), a readout noise of 100 e$^-$ RMS, 30 frames/s and 12 MHz clock rate. At 4–6 kV bombardment voltage, an EB gain of 420–890 resulted. The total gain ranged from 5000 to $6.7 \cdot 10^4$. On average the charge is distributed over 3 pixels, and a minimum gain of 3000 is required for spe detection with an S/N of 10. To reach this S/N, for an EBCCD a voltage of 15 kV and for an ICCD 2 MCPs with at least 1.5 kV and a gain of 10^5 would be required. Since in the IEBCCD the gain of both stages can be lower, this strongly reduces the ionization rate of the residual gas and the iAP rate is over 10 times lower than in an ICCD. A complete camera with 2×2 pixel binning, sub-array readout, 75×95 mm² size, 160 mm length, a HV supply and a computer interface was developed. The detection of single photons was demonstrated with a spatial resolution of about 70 µm. 20 sensors were manufactured, and it was planned to improve the spatial resolution and mean QE and commence production.

The low readout-rate excludes the IEBCCD sensor for LSNDs.

Intensified electron bombarded CMOS (IEBCMOS)[156]

An intensified electron bombarded CMOS (IEBCMOS) is like an IEBCCD with a CMOS sensor instead of a CCD: A photoelectron from a photocathode is proximity-focused into a channel of an MCP stack (usually chevron configuration), where it is amplified. The electron cascade which leaves the channel is accelerated across a small gap onto a CMOS chip. Usually, the bare metal input pads of the CMOS chip are used to detect the incident electron cloud, so no EB gain is created on the chip.[157]

[156] Content based on [609–612].
[157] Unlike e.g. in the Timepix HPD (p. 674) where Si diodes are bonded to the chip pixels for this purpose.

This allows to use lower voltages, which increases the lifetime through a reduced IBF. Depending on the distance between MCPs and chip, the electron cascade either hits a single or multiple pixels. The latter allows to achieve sub-pixel position resolutions through a centroid. However, this comes at the cost of a lower maximum counting rate, since for proper event reconstruction clusters from different photons may not overlap.

In [611] an IEBCMOS prototype was tested for different MCP types with channel diameters down to 4 µm (5.5 µm pitch). Here, the intrinsic (low) QE of the MCPs was used as a UV photocathode. The MCPs were placed 1.1 mm above a custom pixelated analog ASIC [158] with 15×15 mm² area, 300×352 pixels (50 µm size) on a hexagonal grid, and $< 100\,e^-$ ENC. The electron cascade was collected on metal pads on top of the ASIC and the position was obtained through a centroid.
For a chevron MCP stack with 13° bias angle, spe detection with good uniformity, high S/N and < 6 µm FWHM spatial resolution was reported. Concurrent photons could be detected if separated by at least about 0.4 mm. The T_{FWHM} amounted to ca. 130 ps and the pixels were read out either every 20 µs (all pixels) or with 10 MHz readout rate (area around pixels exceeding the threshold; externally triggered). A low DCR was reported.

Other IEBCMOS sensors are based on the different variants of the Medipix/Timepix CMOS chips of the Medipix collaboration, which is led by CERN [609]. Apart from the detection of UV/visible light (Cherenkov light, fluorescence lifetime imaging) these sensor have also been used for soft X-rays, alpha particles, ions, and neutrons (with different particle–electron conversion stages).
The layout consists of a window with a photocathode, whose electrons are proximity-focused onto a chevron MCP stack (to lower IBF). The produced SEs are accelerated across a small gap onto the bare metal pads of an array of 2×2 fast pixelated ASIC readout chips, which are wirebonded to vacuum feedthroughs in a ceramic board. Due to a relatively low ENC of the pixel amplifiers of typically $75\,e^-$, a low MCP gain of about 10^4 is sufficient for spe detection without noise, which leads to a faster MCP channel recharge and a longer lifetime. Typical detector voltages are about 3 kV. A position resolution of 5 and 7 µm (limited by pore pitch) has been achieved using a charge centroid. The TTS is limited by the Timepix chip, which has a temporal resolution of ≈ 10 ns for the Timepix version and ≈ 1.56 ns for the Timepix3; for Timepix4, which is currently being developed, ≈ 200 ps is expected. The sensors can detect concurrent events and GHz counting rates for 28×28 mm² size and have low dark current. However, the chip introduces a quite high heat dissipation into the vacuum tube, which requires cooling. Also, there is a dead area between chips.
The newest member of the Medipix/Timepix IEBCMOS family is a sensor concept which is based on a Timepix4 ASIC and has an area of 40×35 mm², a height of 12 mm (without PCB and heat sink) and a bialkali cathode (7 cm² active area) [610]. For the MCPs, ALD coating and a low gain of a few 10^4 are foreseen, which improves the lifetime by a factor of 100 compared to lead glass and 19 as opposed to a gain above 10^6. In addition, to improve the sensor vacuum, assembly in 10^{-10} mbar and getter strips in the tube are foreseen. The Timepix4 chip (under development as of 2018) will have 512×448 pixels (55 µm side length), where each channel contains an amplifier, discriminator and TDC. The latter has a time-stamp with 195 ps precision

[158] Application-specific integrated circuit

(56 ps RMS) and determines the time-over-threshold, which is used for time-walk correction and the charge centroid position reconstruction, with ca. 1 ns resolution. A sparse readout (only pixels containing events) allows to detect up to 1.2 Ghits/s (limited by how many pixels are hit per photon). A planned ENC of 50–70 e^- per pixel would result in a detection threshold of 500–700 e^-, which would allow single photon counting. A drawback is the large power dissipation (a few Watt) inside the tube. This requires active cooling to avoid unstable operation and damaging the vacuum seals, e.g. with a heat sink and a connected cooling system at the back. The MCP–chip gap should be large enough to use a centroid to reduce the spatial resolution to the pore spacing of 5–10 µm. The TTS is reduced to 20–30 ps (limited by the TTS of the MCP) in a similar fashion by using multiple sampling of the same charge cloud in different pixels.

Contrary to expectations, CMOS sensors *can* have excellent time resolution when using cutting edge readout electronics. However, their readout speed is still slow, which would require a different layout of the trigger electronics in an LSND. Furthermore, since active sensor cooling or significant heat production would introduce convection in such a detector, the high power consumption might be problematic. For many properties no numerical values are known (e.g. QE, charge resolution, DCR, iAPs) and the sensors are all small due to using proximity focusing. However, likely none of these problems cannot be solved and thus the IEBCMOS is included in the short list.

Hybrid micro-channel plate scintillator photodiodes (HMS-PDs)

As in an HM-PD, the phototube of a hybrid MCP scintillator photodiode (HMS-PD) contains a stack of MCPs for primary gain. However, here the SEs are accelerated onto a scintillator to produce light, which is then detected by an imaging photodiode array. Again, either a CCD (ICCD) or a CIS (ICMOS) is used and this layout results in an imaging photon counting detector with slow readout, as well [566].

Intensified CCD (ICCD)[159]

An intensified CCD (ICCD) is an image intensifier which is optically coupled to a CCD camera with readout electronics (fig. 4.10f). The intensifier contains a photocathode and an MCP stack and uses a phosphor as anode, which converts the electron cascade into visible light that is detected by the CCD. ICCDs are commercially available [566] and have been used e.g. in many space missions since the late 1970s for UV and visible light detection from single photons to high fluxes [589; ch. 25].

Usually, a transmissive photocathode is applied on a window. However, e.g. for the detection of vacuum ultraviolet (VUV) light below 110 nm – where no window material is transparent – the window can be omitted and instead a reflective cathode is placed on the top MCP surface. This, however, then requires operation in vacuum. One to three MCPs are used (for photon counting usually at least two) with < 1 kV potential difference per MCP.

[159] Content based on [589; ch. 25] [613].

The SEs are proximity focused onto the anode with a high electric field (up to 6 kV/mm) to increase the detected photon yield (dpy) of the phosphor and to reduce the lateral spread. For spe detection, on the other hand, a much larger spatial spread and charge centroiding is used. The MCP output surface and the phosphor screen are coated with a thin metal layer (e.g. 50 nm Al) to apply a uniform field, reflect photons back to the CCD, and provide a light shield.
The phosphor screen has an emission spectrum (typically from near-UV to green) which matches the spectral response of the CCD, a high light yield (ly; 100–1000), and a short decay time (100 ns to 10 ms drop to 10% height, adapted to the CCD readout time).
Three methods can be used to optically couple the phosphor to the CCD. In the simplest (but seldom used) case the CCD is merely placed close to the phosphor. Alternatively, the output plane can be imaged onto the sensor plane with a lens. In this case, the phosphor is located on a transparent window, the image can be (de)magnified to fit the sensor size, and the camera can be removed or changed. Third, the phosphor can be deposited onto a fiber optic plate, which consists of many fibers (typically 3–25 µm size) and acts as the output window. For this purpose, the CCD window needs to be removed and the fiber optics is glued with optical cement onto the CCD, which provides a far better coupling efficiency than a lens system. A fiber plate can be either used for 1:1 imaging onto the CCD or tapered for (de)magnification (usually up to 4:1 with up to 100 mm diameter [614] for the broad end). For all coupling methods, additional attenuation can be used to adapt the output to the full-well capacity of the secondary sensor.

The ICCD properties depend in large part on the CCD. The QE is lower than for a CCD and is determined by the used photocathode, as is the DCR. The total gain can reach 10^6 [614] with an ENF of about 1.4 (higher than in EBCCDs [507]), which allows photon counting and the discrimination of the DCR of the CCD. A high gain, however, limits the dynamic range through the saturation of the pixels and the image intensifier. For photon counting, even a single photon might be enough to fill the full-well-capacity of a pixel. If charge centroiding is used, the event rate per pixel must also remain below 100 Hz to avoid overlaps.
The frame rate can be up to about 1 kHz and is limited by the CCD and possibly the phosphor (the common P43 phosphor has a decay time of several 100 µs). However, gating of the cathode or MCPs (through the HV) allows very short exposure times of below 1 ns.
Compared to a plain CCD or an EBCCD, the position resolution is limited by the MCPs (also [507]), and an ICCD can have 2–3% optical distortion from the MCPs, electric fields and the coupling optics.
The sensors show iAPs in form of intense white spots in the images [567] [508; p. 931].
ICCDs are very expensive (about $30k–40k as of 2018 [478; p. 272]) due to the MCPs.

The slow readout speed and high costs[160] makes this sensor unattractive for LSNDs.

Intensified CMOS (ICMOS)[161]

The layout and properties of an intensified CMOS (ICMOS) are exactly the same as those of ICCDs (fig. 4.10f), except that a CMOS image sensor (or an sCMOS)

[160] Unless ALD-GCA-MCPs as in the LAPPD are used, see p. 641.
[161] Content based on [589; ch. 25] [613].

is used in place of a CCD. This provides the advantage of a much faster readout. ICMOS cameras have been used in space experiments since 2006 and have begun to replace ICCDs, since they are more radiation hard and allow random readout.

The iQID [162] is an example of an ICMOS. For this sensor, various layouts have been used, which reached total gains of 10^4–10^6, achieved spe detection (for 2 or more MCPs), and used CMOS (ICMOS) or CCD (ICCD) image sensors. [614] One configuration uses a fiber optic *input* window of 125 mm to couple a scintillator detector to a bialkali cathode. The photoelectrons are then proximity-focused to two 40 mm diameter MCPs in a chevron configuration, which is followed by a P43 phosphor and a CMOS with up to 1280×1025 pixels or 400 Hz frame rate (lower spatial resolution).

One very promising concept was put forward by [615], which proposed an ICMOS with an GaAsP cathode, three MCPs (gain 10^6), a P47 phosphor (100 ns decay time), a megapixel CMOS and FADC [163] readout with 1–2 GS/s [164]. This device should be capable of spe detection with 100 ps TTS, $>40\%$ QE between 400 and 600 nm, 20 µm position resolution and counting rates above 1 MHz. The idea was to obtain the time information by reading out the *recharge* current of the lowest MCP with coarse position resolution. By combining this with the good position resolution but coarse time information of the CMOS, one can obtain a signal with both low position and time resolution. Regarding the position-sensitive readout of the MCP, Hamamatsu investigated various possibilities and successfully completed a feasibility study. A prototype was constructed, which, however, had a too low gain to achieve spe detection with the MCP signal and a too slow scintillator (P43, 1 ms decay time) to obtain CMOS frame rates above 1 kHz. It was planned to finish the development of the sensor in 2006, but no follow-up publications could be found.

Regular ICMOS sensors are too slow for LSNDs, but the concept study of [615] shows that in principle ICMOS sensors could become excellent sensors if the proposed concept can be realized. Alternatively, the inclusion of a Timepix4 CMOS (see p. 647) would likely also result in very good properties, depending on the used phosphor type. Nevertheless, the high costs of lead-glass MCPs would still be an exclusion criterion. This might be addressed by using cheaper ALD-GCA-MCPs as in the LAPPD – but at this point one is basically constructing an LAPPD with an additional phosphor screen and a CMOS. Although the phosphor will significantly deteriorate timing (unless MCP readout is added), this concept might still merit further consideration due to the improved position resolution.

Transmission dynodes

Timed Photon Counter (TiPC) [165]

The Timed Photon Counter (TiPC or "Tipsy") is based on transmission dynodes and was in large part already explained on p. 582.
In this hybrid sensor (fig. 4.10g), the photocathode is placed above a stack of curved tynodes (inter-tynode distance ≈100 µm) on top of a CMOS chip (Timepix).

[162] Ionizing-radiation Quantum Imaging Detector
[163] Flash analog-to-digital converter
[164] Gigasamples/s
[165] Content based on [253, 616].

Focusing electrodes (not shown) concentrate the photoelectrons on the first tynodes to improve the CE. For a TSEY of 4 and with 8 successive tynodes, a gain of $6.5 \cdot 10^4$ could be reached, which would be sufficient for spe detection with a Timepix3 CMOS chip even without an in-pixel amplifier. Tynodes with an TSEY of up to 5.5 have been fabricated. A time resolution in the order of only 1 ps is predicted, while the spatial resolution depends on the pixel pitch and size. Compared to the SEY of dynodes in PMTs, one drawback of the low TSEY is that the charge resolution will be very large. Assuming an optimistic TSEY of 5.5, spe detection can be achieved with 7 tynodes ($g = 1.5 \cdot 10^5$). According to eq. (2.65), then R_σ will be 47.1% from statistics alone and before considering the influence of inelastic backscattering. From eq. (2.68) follows an R_F [166] of at least 111%, which is worse than for most PMT series (cf. table 3.7). The npe peaks will, however, likely still be discernible and the P/V should be in the region of 2.

So far, only a prototype has been produced where an electron gun substituted the photocathode, and a single layer of Al_2O_3/TiN/Al_2O_3 tynodes and a Timepix1 sensor (512×512 pixels with 55 µm pitch and 25 ns temporal resolution) was used [616]. The charge which is collected on each pixel could only be measured indirectly and non-linearly through a time-over-threshold mode. However, a TSEY of 0.1–0.9 for 1.2–1.8 keV impact energy could be deduced, which is in agreement with scanning electron microscopy measurements. In the spatial distribution of the electrons on the CMOS sensor, a separated banana-shaped cluster was attributed to SE emission from the sample holder (Rm).
Next steps will include the upgrade of the CMOS sensor to a Timepix3 chip [167], which can measure the absolute charge and has a time resolution of 1.56 ns. Furthermore, it is planned to test high-TSEY samples in the TiPC prototype, stack multiple tynodes, reduce the tynode thickness to allow to use lower operating voltages, and to produce and include a photocathode to obtain a fully functional TiPC.

At its current state, the TiPC can not be included in the shortlist, but future developments should be closely monitored.

4.3.1.1 Electron bombarded scintillator

Hybrid scintillator photomultipliers (HS-PMTs)

Hybrid scintillator PMTs (HS-PMTs) consist of a vacuum phototube where the photoelectrons are fountain-focused onto a fast scintillator crystal (see figs. 4.12a–4.12d). The photons which are generated through electron-bombardment are detected by a small PMT. Usually, a very high EB voltage of 20–40 kV is applied, which results in 20–50 detected photons in the PMT and has a favorable effect on sensor properties. For better uniformity, the photocathode has a spherical or hemispherical shape, which allows to construct large sensors with 8–15" diameter and raises the effective mean QE through the double cathode effect (see p. 64). This HPS class therefore shows high CE, good charge resolution, good TTS (depending on scintillator), low amounts of EPs and LPs, low AP rates, and possibly higher magnetic field tolerances than PMTs at comparatively low costs [553, 617].

[166] Charge resolution based on the FWHM of the SER, see eqs. (2.66) and (2.67).
[167] The Timepix4 chip, which is currently being developed, has even better properties, see p. 647.

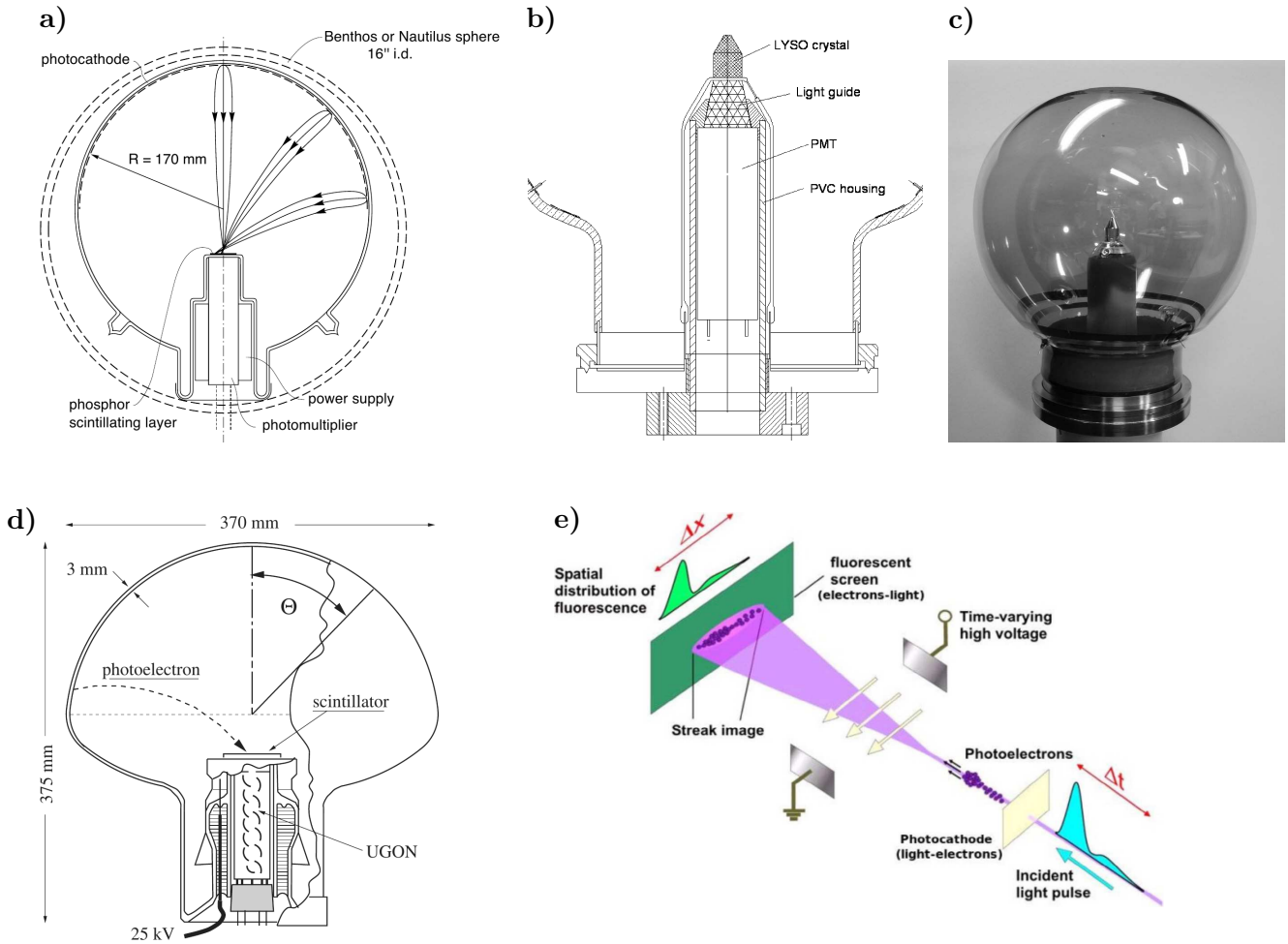


Figure 4.12: Hybrid photosensors based on electron bombardment of a scintillator. Continued on next page, discussion see text.
a) SMART (Photonis XP2600), from [172; p. 1-23].
b)–c) X-HPD: b) Layout, from [618]. c) Photo, from [618].
d) QUASAR-370, with small UGON PMT, from [553].
e) Streak camera, from [619].
f)–g) ArcaLux: f) Side view, g) top view, both from [519]. The hemispherical photocathodes (2) and the reflectors (3) are realized as cavities in a flat panel (1). Incident photons (ph) are converted to photoelectrons (e), which are accelerated onto a scintillator (center). This is coupled to a fiber (6) which leads to an SiPM (outside figure). The voltage is applied to both the cathode and the scintillator by a resistive plate (4, green) and the vacuum is maintained by getters (5).

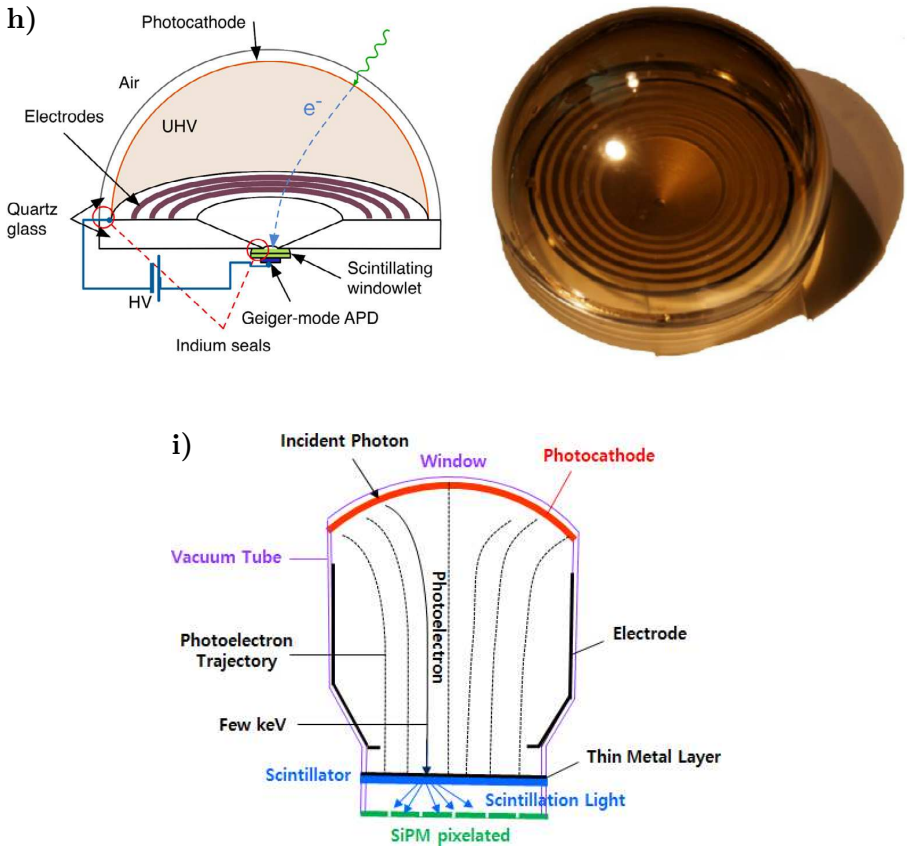


Figure 4.12: Hybrid photosensors based on electron bombardment of a scintillator. Continued from last page, discussion see text.
h) ABALONE, from [283].
i) Silicon Photomultiplier Tube (SiPMT), from [620].

SMART (Photonis XP2600) [168]

The Photonis XP2600 ("SMART" [169]) was developed by Philips (which later became Photonis) from 1981–1986 [352] for the DUMAND program, which aimed to build a deep-sea WCD for high-energy neutrinos (see p. 9). About 30 sensors were produced, some of which were operated in the Baikal neutrino telescope (also p. 9).

The XP2600 (fig. 4.12a) consists of a spherical glass bulb of 34.5 cm diameter with a hemispherical K_2CsSb bialkali photocathode, which covers a surface of 3π sr of the bulb. Photoelectrons are accelerated by a voltage of 25 kV (up to 40 kV) onto a thin, disk-like monocrystalline YSO phosphor (Y_2SiO_5:Ce, also called P47, 40–60 ns decay time, ly $\approx 10^4$ photons/MeV). The phosphor disk is placed ca. 6 cm below the center of the sphere on a glass viewing window and is coated with a thin Al foil to apply the voltage [170]. Since the Al foil is typically 100 nm thick, the pe loses only a few kV of impact energy, and because photoelectrons penetrate only a few µm into the scintillator, a thin phosphor is sufficient. About 300 photons are

[168] Content based on [621; pp. 58–61, 71–73] [239, 553, 622, 623] [172; p. 1-22].
[169] The moniker "smart" photomultiplier was coined to denote the capability to distinguish 1 pe from 2 pe events and in this manner discriminate the majority of background sources (environment, DCR).
[170] This also prevents light feedback.

produced, of which ca. 20–50 (typically 30) are detected by a small fast PMT (e.g. Philips XP2982), which is placed in a recess at the back of the phototube. This high first stage gain leads to a good charge resolution and a low TTS. The anode section consists of a glass recess with a viewing window for the PMT and is connected to the bulb with welded metal seal rings. The power supply for the EB voltage is placed in this recess, as well.

The QE is about 24.5% at 380 nm, with a maximum increase from pole to equator of about 44.5% at 405 nm due to the double cathode effect. A CE of ca. 100% is reported, which is uniform over nearly the complete cathode area.
The charge resolution R_F is ≈ 50% (45–70%), which allows to distinguish up to 3 pe, and the gain is uniform over the cathode area.
The TTS T_{FWHM} is ca. 5 ns [621] which mainly originates from the CTS (4–5 ns between 0 and 90° from zenith), photon production statistics in the scintillator and the TTS of the PMT (ca. 0.5 ns). For spot illumination at 0° and 90°, virtually no EPs are found, but a trailing flank of LPs with up to above 19 ns lateness is present. The pulse shape is dominated by the long decay time of the scintillator, and individual photons are visible as distinct pulses.
DCR pulses stem mainly from thermionic electrons from the cathode with 1 pe charge and the dark current rises linearly with EB voltage up to ca. 30kV, where a steep rise begins. APs are practically absent.
The sensitivity to magnetic fields is lower than for PMTs and no shielding is required for the GMF; however, the PMT is furnished with a mu-metal cylinder.

A good PDE [171], a good charge resolution, and a low EP probability are found and the XP2600 can tolerate the GMF. For a use in LSNDs, however, the TTS and possibly also the LP probability are too large. In addition, this sensor is no longer fabricated.

QUASAR-370 [172]

The QUASAR-370 is an HS-PMT which was developed based on the XP2600 by the INR (Moscow) and the KATOD company in the 1980s for the Baikal neutrino telescope (also [239, 348]). The first designs (QUASAR-300 and QUASAR-350 with 30 and 35 cm diameter) still had a spherical shape like the XP2600. The final version QUASAR-370 (fig. 4.12d), however, uses a mushroom-shaped borosilicate glass bulb of 37 cm diameter, which provides a significantly lower CTS and TTS. The development continued until 2012 in form of a large number of modifications of the scintillator and the secondary photosensor. Over 200 QUASAR-370 sensors were fabricated and used in the Baikal neutrino telescope and the TUNKA experiment for over 15 a, but the sensor is not produced anymore.

The QUASAR-370 contains a hemispherical K_2CsSb bialkali cathode with a viewing angle of over 2π and ca. 2000 cm^2 sensitive area. The photoelectrons are accelerated by ca. 25 kV onto a thin, flat layer of a fast inorganic scintillator with high light yield, which is placed near the center of the glass bulb. In the standard version of the QUASAR-370, the scintillator is a disk of YSO (ly ca. 10^4 photons/MeV,

[171] QE ca. 24.5%–44.5% depending on photon path, CE ca. 100%, multiplication efficiency (ME) from backscattering unknown but likely significantly lower than 100%.
[172] Content based on [502, 553, 624, 625].

decay time 30–50 ns) with 20 mm diameter. The scintillator is either a pulverized phosphor (6 µm thick – optimized for 25 keV) or a monocrystal (≈ 100 µm thick for mechanical stability). The EB results in a dpy of ca. 20–30 pe (up to 50) for 25 kV. A 100 nm thick Al foil covers the scintillator to suppress light feedback and to reflect light and thus increase the dpy. The scintillation light is detected by an "UGON" PMT with 3 cm diameter, which was specially developed by the INR and MELZ. This PMT has a 25 mm wide spherical K_2CsSb bialkali cathode (ca. 18% QE at 410 nm) on a plano-concave borosilicate glass window, 12–13 dynodes, a gain of 10^7, 2.2 ns T_{FWHM}, a P/V of typically 1.3, and a DCR ≤ 1 kHz for 0.25 pe threshold at 20 °C (also [626]). This results in a total gain of the hybrid sensor of about $2.5 \cdot 10^8$.

The QE of the QUASAR-370 is about 20% at 420 nm with a peak between 400 and 420 nm. The double cathode effect leads to an improvement by ≈ 22% from pole to equator (QE ≈ 24%) at 380 nm [239]. For good tubes, the PDE uniformity is better than ± 10% up to 90° zenith angle, but this strongly depends on the precision of the mechanical assembly.
A voltage of 25 kV is considered safe for stable operation and results in an R_F of 70–80% and a P/V of about 2.5.
The high EB voltage and the hemispherical shape of the cathode lead to a very small CTS with a maximum transit time difference of about 0.8 ns. The pulse shape consists of a superposition of single PMT pulses, which are distributed in time according to the exponential decay of the scintillator. For this reason, one should use two discriminators to minimize the TTS: one with 0.25 pe threshold for timing and one with 0.25 · dpy pe threshold for background discrimination. This effectively suppresses LPs and small EPs from the PMT and results in a mean T_{FWHM} of 2 ns. In the TTD, the flanks of the main peak follow exponential functions with a significantly longer right flank. No additional EP or LP peaks are observed between −10 and 22 ns from the main peak.
The left flank of the main peak is determined by the scintillator rise time, the TTS of the PMT and the electronics jitter. The absence of EPs was attributed to the optical decoupling of the small PMT, which renders photonic EPs such as $\gamma 1$ (see p. 282) impossible. Nevertheless, the present author would like to remark that the used two-discriminator method will still be sensitive to electronic EPs in the PMT, since these can shift the threshold time forward. For E1 (p. 283) the earliness is similar to the rise time of regular pulses in the PMT, and PMT pulses typically show a plateau around the pulse peak. Therefore, at the time when the threshold is normally passed, E1 pulses are close to their peak and introduce a near-maximal forward shift. The effect will be visible even for a single E1 pulse and will become more pronounced for multiple EPs. Assuming an E1 probability of merely 1%, the likelihood of one E1 per photon is already 22%, since about 25 pulses per photon are created in the PMT. For two E1 pulses it is 2.4% and for three 0.18%. BE1 EPs will also be present with far higher charges and should have an even stronger effect. Thus, it is concluded that PMT EPs *will* cause EPs in the QUASAR-370, which broaden the rising flank of the main peak in the TTD.
The falling flank of the main peak (LPs up to ≈ 7 ns lateness) is determined by the scintillator response and can be well described by the function $\exp(-g_s\, t/\tau)$ with g_s the dpy, t the time, and τ the scintillator decay time. From this follows that the TTS is improved by a higher dpy and a shorter decay time. Photoelectrons which are inelastically backscattered from the scintillator and refocused onto it (BLs,

where s stands for scintillator) do not enter the TTD, since they produce a primary pulse (PP) with regular time. However, they are visible as early afterpulses (EAPs) in form of pileup in the pulse shape with delays of ca. 20 ns (BLs) and 40 ns (BLBLs). These delays correspond to two and four times the transit time between cathode and scintillator. At 25 keV the probability for elastic backscattering is $< 0.1\%$, which explains why no clear Ls peak is observed in the TTD.
The mean DCR amounts to 30 kHz at 0.25 pe threshold and 20 °C.
The iAP rate is very low and only an upper limit of ca. 0.1% can be given. No peaks are observed in the delay distribution (DD) up to 32 µs delay. This is ascribed to the separated vacuums of PMT and phototube which suppresses iAPs. Furthermore, bremsstrahlung photons which are produced by the photoelectrons here have high energies of typically 10 keV, for which the photocathode sensitivity is low. This reduces the amount of photonic afterpulses.
The QUASAR-370 is immune to the GMF.

The scintillator is of crucial importance for the operational characteristics of the sensor. When the QUASAR-370 was designed, YSO ($\rightarrow$ ly $\approx$ 10k/MeV, $\tau = 30$–50 ns, $R_F = 80\%$, $T_{\mathrm{FWHM}} = 2$ ns) was the most effective and fast inorganic scintillator. [624, 627] However, since then a number of materials with better properties have emerged and have been tested in the QUASAR-370. A good scintillator should have a high ly for 10–30 keV electrons (the dpy is proportional to the ly and a high value raises the gain and reduces the charge resolution and TTS). In addition, a low decay time is required (improves TTS). A low effective atomic number Z_{eff} is important to suppress backscattering, the emission spectrum should match the spectral response of the PMT, and a good proportionality of the scintillator is necessary. The material should also be compatible with vacuum and the photocathode manufacturing process (high temperature and aggressive chemical atmosphere; for compatible materials no transfer technique has to be used, which reduces costs). The compatibility requirements limit the selection to inorganic nonhygroscopic scintillators.
The most promising scintillators which were tested in the QUASAR-370 so far are:

- **YAP** ($YAlO_3$:Ce): monocrystal, ly $\approx$ 16k/MeV, $\tau = 25$–30 ns, $R_F = 40$–60%, $T_{\mathrm{FWHM}} = 1.2$–1.5 ns
- **SBO** ($ScBO_3$:Ce): phosphor, ly $\approx$ 16k/MeV, $\tau = 30$ ns, $R_F = 40$–60%, $T_{\mathrm{FWHM}} = 1.2$–1.5 ns
- **LSO** (Lu_2SiO_5:Ce): monocrystal, ly $\approx$ 30k/MeV, $\tau = 40$ ns, $R_F = 35\%$, $T_{\mathrm{FWHM}} = 1$ ns

With LSO (QUASAR-370LSO), up to 7 pe can be distinguished in the SER and for higher voltages the TTS drops even further (0.8 ns at 40 kV). However, the material has a large nonproportionality and a high Z_{eff} of 66, which increases the LP probability.
The sensor properties could be improved substantially still, if ZnO:Ga (ly $\approx$ 16–40k/MeV, $\tau = 0.4$ ns) were used [553]. For this material, a $T_{\mathrm{FWHM}} < 1$ ns and a pulse width < 1 ns are expected [352]. So far, only small scale prototypes with low quality could be produced, however [348, 628].

In addition to the scintillator studies, a two-channel device (QUASAR-370-II) was fabricated. This exploited that photoelectrons from the central cathode region

are focused onto the center of the scintillator, whereas those from peripheral regions are focused onto the edge [502, 625]. In this variant, a modified UGON PMT with mesh dynodes and two anodes (central and peripheral regions) was used, which had a σ_T of 340 ps, a P/V of 1.5, and 1% crosstalk between channels. In this sensor, background events from the environment or DCR as well as EPs, LPs and APs can be suppressed through a coincidence of both channels.
The DS-QUASAR (already mentioned on p. 614) is a direction-sensitive variant, which is achieved through a sectored LC and an MaPMT [559]. Simulations showed that slight changes of the bulb shape and the scintillator position would improve the position separation of photoelectrons on the scintillator and drastically lower the σ_T from 1 ns (in agreement with measurements) to 0.2 ns. Construction of a prototype was commenced (glass bulb and LC), but apparently not concluded.
The QUASAR-370G is a low gain variant for high currents from night sky background, which features a lower cathode resistance, an adapted scintillator and an UGON-6PMT with only six dynodes [626].
A variant with low background glass exists as well (QUASAR-370L) [625].
Furthermore, two prototypes with PIN PDs instead of a scintillator [173] were produced and studied (Quasar-370D), but no properties were reported in the sources mentioning them [352, 624, 625].
In [519] and the references therein, HS-SiPM prototypes were reported, for which a QUASAR-370 was retrofitted with a single 1x1 mm² SiPM instead of a PMT. This allowed to resolve the npe peaks in the SER and test the position sensitivity of the QUASAR.

In summary, the QUASAR-370 has an acceptable charge resolution (R_F 70–80%), a good P/V (2.5), a good T_{FWHM} (2 ns), the EP and LP probabilities are likely within the limits, the AP rate is very low ($< 0.1\%$) and it is immune to the GMF. The LSO variant shows even better properties (R_F 35%, T_{FWHM} 1 ns) but also will have a higher LP probability. The PDE requirement of 20% likely is narrowly missed (QE 20–24% depending on angle, CE likely close to 100%, ME probably significantly lower than 100%). In addition, since the DCR of 27.9 Hz/cm² (target 2.2 Hz/cm²) is quite high, it is questionable whether it would show a sufficient decrease after long times in the dark. It would be preferable to improve the QE and DCR to the level of a modern standard quantum efficiency (SQE) PMT, which should in principle be possible. Unfortunately, the QUASAR-370 is not produced anymore, however. Since the DCR and PDE would be the only properties excluding this sensor (as is the case for many PMT series), it is added to the short list for comparison.

X-HPD [174]

The X-HPD photosensor (figs. 4.12b and 4.12c) was developed by CERN, Photonis and IN2P3 (by the latter for KM3NeT) between 2007 and 2010 [175]. It is based on the spherical 15" HPD that was conceived for the C2GT project and which has a cube-shaped anode that is covered by 5 silicon PDs or APDs (see p. 673 and fig. 4.13e and 4.13f). To simplify the design at the cost of a slightly lower performance, this layout was adapted to an 8" spherical HS-PMT with a non-planar scintillator, with

[173] These sensors thus belonged to the HPD class.
[174] Content based on [281, 618, 629].
[175] Apparently the development was prematurely ended by the cessation of all PMT activities by Photonis in 2009.

the goal to achieve better properties than in SMART and QUASAR-370. It was planned to develop X-HPDs with up to 15" diameter [630].

The X-HPD consists of a nearly spherical glass envelope with a round opening, which is sealed by a base plate with a recess for the PMT. A light guide couples the PMT to the scintillator crystal through a window.
The glass bulb is covered by a semitransparent bialkali cathode with 120° viewing angle, which corresponds to a solid angle of 3π and allows to use the double cathode effect. The reachable viewing angle is limited by the geometry and field distortions of the anode.
The photoelectrons are accelerated by a voltage of 20–30 kV onto the scintillator. The field is radial ($\propto 1/r^2$) apart from distortions near the anode which are corrected by a ring electrode at ca. −1 kV that is placed on the glass bulb below the cathode. The spherical symmetry results in a uniform CE of almost 100% over the full viewing angle. In addition, it produces a very small TTS, since the pe paths have only sub-ns flight time differences (CTS), and provides immunity to the GMF.
A non-flat scintillator crystal acts as the anode and is mounted in the center of the sphere. It is coated with a reflective Al film to apply the potential and prevent light feedback. The shape (cylinder, hemisphere, cone) affects the dpy through the photon path lengths and the number of reflections in the scintillator. Due to focusing inaccuracies and the spot size (see p. 179), the crystal must have a minimum size for uniform detection. Monte Carlo simulations and measurements showed that the dpy is highest for the zenith angle (0°) and significantly drops for larger angles. The EB leads to a dpy of 20–60 pe, depending on HV, crystal and optical coupling, which allows moderate photon counting capabilities.
The scintillation light is detected by a small, low cost sensor such as a PMT. The secondary sensor is mounted outside the vacuum tube due to its incompatibility with the bake-out process. This comes at the cost of light coupling losses at the interfaces between crystal, window, light guide and PMT. In particular, substantial losses are incurred by a small vacuum gap at the crystal–window interface.
Due to the uniform fields, in principle it should be possible to construct an image sensor by using segmented scintillators and secondary photosensors.
Because of the simple layout with few components, the X-HPD is expected to have an affordable cost.

Measurements with a first fully operational X-HPD prototype were reported in [281, 618]. This sensor had a diameter of 208 mm, a cathode surface of 1020 cm², and a conical LYSO ($Lu_{2(1-x)}Y_{2x}SiO_5$) crystal anode (12 mm diameter, 18 mm height, tapered with 60° near the 3 mm wide flat top, assumed ly of 30 photons/keV). The crystal was coated with 100 nm Al and was placed on top of a glass cylinder with a thin, flat window. The light guide consisted of a truncated cone of Al-coated plexiglass and was coupled to a 25 mm diameter PMT (Photonis XP3102, gain $6 \cdot 10^6$, 1000 V, P/V 2.7, PDE ca. 22.5%).
The double cathode effect increased the QE by about 60% between ca. 40° and 115° at 420 nm; no absolute QE values were measured. In a previous prototype, an increase by a factor of 2 and a QE of 50% at 350 nm was observed for side illumination. The larger increase was attributed to a thinner cathode.
At +20 kV, the SER showed an spe peak with a P/V of about 1.6 and a strong exponential component of undersized pulses. The latter resulted from about 50% backscattered photoelectrons (due to the high $Z_{\text{eff}} \approx 63$ of LYSO [558]) which deposit

only a part of their energy in the crystal. Simulations showed that the charge resolution and TTS are limited by the large variation of photon escape paths in the scintillator but can be lowered through an optimized crystal shape and improved optical coupling. A thin phosphor layer on a hemispherical window would significantly reduce such fluctuations and could also improve the gain uniformity over the cathode, but would result in a lower mean dpy.
The dpy was between 16 and 59 for −12 to −26 kV, and between 29 and 45 for +18 to +24 kV. At −20 kV the value was 40 and for positive voltages it was ca. 15% lower. This was attributed to an optical effect of the insulator between the window and light guide.
The DCR amounts to 100 kHz (100 Hz/cm²) at 0.1 pe threshold and ≈ 25 °C for positive voltage. For negative HV it was higher by a factor of 20, which indicates micro discharges between corrector electrode and base flange.
At 20 kV a TTS of 2.5 ns FWHM was measured and the TTD showed an LP tail.

In [558] an alternate HSiPM version of the X-HPD was suggested (fig. 4.15a). This variant took up the original C2GT HPD layout of a central anode cube, which is covered by 5 diode-based sensors, but used SiPMs instead of PDs/APDs. This would in principle allow to use it as an image sensor or a direction-sensitive sensor. Compared to the HPD (C2GT) or HS-PMT design, a lower EB voltage of about 10 kV instead of 20–25 kV would suffice. This would reduce the problems that result from the photocathode evaporation process, during which all components are exposed to alkali vapors. In addition, the PDE should be larger, since the gain is at least a factor of 100 higher than with PDs and the lower $Z \approx 14$ of Si would lead to a significantly lower fraction of backscattered photoelectrons. Furthermore, the low TTS of SiPMs would lead to a better time resolution than achievable with p^+n diodes after amplification and shaping or with a scintillator plus PMT, and the TTS would be limited only by the CTS. However, the enormous DCR of (uncooled) SiPMs could be a serious performance limit for many applications.

The X-HPD has a very good QE of 25–50% (depending on angle), a CE of almost 100%, a good T_{FWHM} of 2.5 ns, and is immune to the GMF. However, it suffers from ca. 50% backscattering due to the high Z_{eff} of LYSO, which affects the PDE, SER and TTD. The P/V therefore is only 1.6, which lies below the acceptance threshold, and it can safely be assumed that the LP probability will be very high. Furthermore, the DCR of 294 Hz/cm² (cross-section) is far too high (which in part is due to the large cathode surface per cross-section). The X-HPD is thus excluded for LSNDs.
The HSiPM version is not included in the shortlist either, since this was only a concept study and all HSiPMs which produced prototypes are excluded due to their enormous DCR (see p. 689).

Hybrid scintillator photodiodes (HS-PDs)

Streak camera [176]

Despite of incorporating a CCD, the streak camera (or streak tube) is not an image sensor (see fig. 4.12e). Rather, the device trades a position resolution for superior temporal resolution – state-of-the-art devices can reach values of as low as 200 fs (σ_T).

[176] Content based on [210, 631, 632].

The principle of operation is based on the rapid deflection of an incident light beam, which transforms a temporal into a spatial distribution.
Simple types of streak cameras use optical deflection, e.g. by a rotating mirror, and achievable time resolutions are limited by the moving mechanical parts.
Optoelectronic types are most common, in which case incident photons are converted into photoelectrons in a photocathode. These electrons are accelerated by a HV, deflected by a fast linear field, and hit a phosphor. Optionally, the electrons are amplified in an MCP first (in this case the sensor is an HMS-PD instead of an HS-PD). The phosphor converts the electrons back to photons, which then are detected by a low-noise CCD with a high spatial resolution.
Thus, the deflection (typical streak speeds 2–5 ps/mm) is proportional to the photon arrival time. The CCD pixels should be small enough to not limit the resolution. Usually, the pixel size is ca. 10 µm, which would be sufficient for 20 fs at 2 ps/mm. Also, the diameter of the light beam in direction of the deflection should be similar to the pixel size, which can be achieved through a slit. Apart from the layout of the streak camera, the achievable resolution is limited by space charge effects in the electron beam (mutual repulsion of the electrons – weak signals thus have a lower TTS) and dispersion in the optics.
Since the CCD is a 2D image sensor, the spatial distribution along the slit can be measured, and in this way 1D or 2D (through event repetition and adjustable optics) images *can* be obtained.
Drawbacks of this photosensor are the low dynamic range and high costs [322]. For a defined deflection offset, one also needs to coordinate the deflection mechanism and the recorded event.
Streak tubes are often used to measure the duration or peak power of an optical pulse.

Due to the small sensitive surface and high costs, streak cameras are not suited for LSNDs – this sensor has a different field of application.

Hybrid scintillator silicon photomultipliers (HS-SiPMs)

ReFerence (HS-SiPM variant)

See p. 679.

ArcaLux[177]

The ArcaLux was a conceptual study for a photosensor which is composed of an array of HS-SiPMs. The layout is depicted in figs. 4.12f and 4.12g and consists of a square matrix of hemispherical cavities (diameters of a few cm) in a flat panel, whose surfaces are covered with photocathodes. In each hemisphere, the photoelectrons are accelerated by a voltage of at least 15 kV onto a scintillator film. The latter is covered by a thin Al film, which serves as a light shield and allows to apply the potential. A high ly scintillator with a dpy of at least 30 was foreseen so that the signal exceeds the background which results from the DCR of the SiPMs. The resistive bottom plate of each cavity acts as a continuous voltage divider between anode and cathode and is coated with a dielectric mirror to reflect transmitted photons back onto the cathode. The photons which are produced in the scintillator are transported to

[177] Content based on [519].

an SiPM (gain ca. 10^6) outside the vacuum enclosure, either with a light guide or through a window. Cavities with cross-shaped cross-sections are located between the hemispheres to reflect additional photons onto the cathodes and thus improve the fill factor. To maintain the vacuum, getters are placed inside these concentrator cavities. The flat window allows a good optical coupling to liquid scintillators, water or ice. In addition, the layout has a high pressure resistance, which is improved by using a quadratic instead of a hexagonal arrangement. Furthermore, the sensor is designed for full angular acceptance (which can be a drawback in detectors such as LENA) and the hemispherical cathodes provide a low CTS.
A very high CE, a low TTS, and insensitivity to the GMF were expected. Due to the pixelated design, the ArcaLux would be an image sensor. In a very rough cost estimate, a final price of $< \$1000/m^2$ sensor area was assumed. No prototypes were produced to the author's knowledge.

In an alternative layout, APDs are used instead of the scintillators and SiPMs, which places this variant in the HAPD class.

Since both the HS-SiPM and HAPD versions only reached the stage of a concept study and no prototypes were fabricated, their potential cannot be assessed and the sensors are not included in the short list.

ABALONE [178]

The ABALONE [179] photosensor (fig. 4.12h) is a medium-sized HS-SiPM which was invented at the University of California, Davis (UC Davis). This sensor is being developed for APP (in particular for a potential extension of IceCube), cryogenic dark matter (DM) search and low temperature double beta decay experiments. The goal is to obtain a photosensor which is robust and affordable, can be mass-produced, and has very good performance.
The sensor has reached the stage of functional prototypes for the IceCube variant with fused silica glass, whose current layout and properties will be summarized here.

The ABALONE consists of a dome and base plate with ca. 10 cm diameter, which are made of fused-silica glass (for UV sensitivity and high radiopurity) or regular glass. A small scintillator window covers a hole in the center of the base plate. Since glass does not release trapped gases like metals, no bake-out is needed and plasma cleaning is sufficient. Thin film deposition of ultrapure metals is used to apply the vacuum seals between dome, base plate and scintillator, the electrodes on the base plate, and the film covering the scintillator. The vacuum seals are also used to apply the ground potential and the HV, which allows to omit feedthroughs. A thin-film getter covers 90% of the base plate.
In the prototype, a Cs_3Sb cathode (see p. 55) is used, since this material combines to an optimal stoichiometric configuration during deposition, but bialkali or multialkali cathodes can be used as well.
Photoelectrons are accelerated with 25 kV (highest practical value) onto the scintillator, which results in a CE of nearly 100%. The focusing field is shaped by passive, floating electrodes in the shape of concentric rings on the base plate. These assume the appropriate potential shortly after powering on (similar to an RPC, see p. 590).

[178] Content based on [282, 283, 633].
[179] Named after the large iridescent shell of the eponymous sea snail.

A conductor which is at cathode potential is placed below the base plate to improve the focusing field and to increase the CE of the photoelectrons originating from 5–10 mm above the base plate.
The scintillator window either is entirely made of scintillator or consists of a glass plate covered by a scintillator film of $< 2\,\mu m$ thickness. A thin metal alloy film coats the scintillator to prevent light feedback, apply the potential, and reflect photons back to the cathode. Photoelectrons with 20 keV lose only about 0.3 keV in the metal film and penetrate $< 2\,\mu m$ into the scintillator. In the prototype a $6\times6\times1.5\,mm^3$ bulk LYSO scintillator was used, which for 25 kV produces ≈ 650 photons, of which 100 are detected by the SiPM ($6\times6\,mm^2$ SensL J-type MicroFJ, 35 µm cells, gain $6.1 \cdot 10^6$ at the recommended bias). This results in a total gain of about $6 \cdot 10^8$. The dpy raises the gain and allows to discriminate photons from SiPM dark noise but lowers the dynamic range [634]. LYSO has a decay time of 42 ns and acceptable performance and was chosen for mechanical and chemical robustness, but faster scintillators with higher ly are available.
The P/V amounted to about 2.9. The charge resolution was better than for PMTs but significantly higher than for SiPMs, since inelastically backscattered photoelectrons fill up the valleys between peaks in the SER.
The pulse height was ca. 6 mV and the pulse width ≈ 150 ns.
When using a constant fraction discriminator with 5% threshold, this resulted in a T_{FWHM} of 740 ps, including an LED jitter of about 500 ps. After quadratic subtraction of the latter, a TTS of about 550 ps remains. Simulations showed a transit time of 2.3 ns between cathode and scintillator for central illumination and 1.5 ns from the brink. No TTD was shown in available publications, but inelastic pe backscattering from LYSO ($Z_{eff} = 66$) with 30–40% probability is reported. The backscattered electrons are either refocused onto the scintillator (BLs, undersized PP and EAP) or lost (BLsTm, undersized PP). Elastic scattering (Ls) should also occur, but will be suppressed in the TTD for the same reason as discussed for the QUASAR-370.
Due to the relatively noisy Cs_3Sb cathode material, a dark current of $0.2\,fA/cm^2$ was found at 25 kV.
The AP rate was only $0.5 \pm 0.2\%$ at 0.25 pe threshold. The dominant contribution in the iAP DD (cf. 2.5.2.3) is a peak at 185 ns from He^+/He^{++} with additional contributions from $C^+/N^+/O^+$ (205–245 ns). APs occurred up to ca. 400 ns (measured delay range ca. 170–880 ns). Large AP charges of up to at least 8 pe were observed.
An early prototype was tested through 4 a of continuous operation. During this time, the AP rate dropped significantly to $0.08 \pm 0.04\%$ at 1 pe threshold ($7 \pm 5 \cdot 10^{-3}$ % for 4 pe). This decrease was attributed to ion implantation: The ion energy is sufficiently high to pass the thin (≈ 15 nm) cathode and penetrate ca. 50–200 nm into the glass, where all ion types except He remain, which results in an efficient vacuum pump mechanism. This is not possible in PMTs and MCP-PMTS due to the lower ion energies, for which the ions are chemisorbed in the cathode (poisoning it) or return to the vacuum. Since He cannot be implanted or chemisorbed, it is the dominant iAP type in the ABALONE. Furthermore, the ionization probability is significantly reduced compared to PMTs due to the large pe energies.
The sensor is resistant to shock and vibrations, pressure and low temperatures, and is not damaged by accidental exposure to strong light.
The authors claim that a cost of less than $\$5k/m^2$ sensitive area could be achieved with a fully developed production process due to the simple layout, the use of prefabricated components, and a continuous production line.

Presently, prototypes have been produced in a downscaled production line and PhotonLab has acquired a facility with the intention to perform R&D and gradually scale up the production.

When used in cryogenic DM and double beta decay experiments, it is foreseen to combine multiple ABALONEs to panels.
The SiPM can also be replaced by an APD, PIN diode or a small CCD. The latter was in fact used during prototyping, which resulted in an HS-PMT hybrid sensor.

The ABALONE has an extremely high gain ($6 \cdot 10^8$), a good P/V (2.9) and sufficient charge resolution, a very good T_{FWHM} (550 ps) and a low AP rate (0.5%), which further drops during operation. Drawbacks are the limited dynamic range, the strong backscattering on LYSO (which leads to a low ME, a high LP probability and a large EAP rate), and likely a low QE and high DCR due to the use of a Cs_3Sb cathode in the prototype. At the moment, some crucial properties are not known, although it is probable that the current prototypes do not fulfill all criteria for LSNDs. However, the backscattering might be reducible by employing a different scintillator, the use of a bialkali cathode should improve QE and DCR, and the sensor is still being developed. Therefore, the ABALONE is included in the shortlist and future progress should be closely watched.
An APD variant should also prove interesting.

Silicon Photomultiplier Tube (SiPMT) [180]

The Silicon Photomultiplier Tube (SiPMT) is another HS-SiPM which is currently being developed. The goal is to obtain a large low-cost photon counting sensor with a manageable DCR and a moderate operating voltage for use in neutrino detectors (e.g. KNO [181]). Presently, the sensor is in an early stage of development (working proof-of-principle demonstrator), so only target properties can be quoted.

In the SiPMT (fig. 4.12i), photoelectrons from a curved photocathode are accelerated by a field of a few kV and are fountain-focused onto a scintillator by an electrode on the side wall. An SiPM is attached to the scintillator and counts the produced photons [182]. Again, the scintillator is covered by an Al layer to prevent photon feedback.
A dpy of ≥ 10 and an SiPM gain of 10^6 is foreseen, which would result in a total gain of 10^7.

A demonstrator with a polycrystalline Ag cathode (UV-sensitive only) of 12 mm diameter, a $10\times10\times3\,\mathrm{mm}^3$ CsI scintillator (without Al layer), and a $6\times6\,\mathrm{mm}^2$ SiPM with 35 µm pixels was constructed. An Al ring placed in front of the scintillator served as anode. This sensor was successfully operated with 3–4 kV EB voltage and 266 nm laser light.

[180] Content based on [620].
[181] Korean Neutrino Observatory; a proposed second WCD in the J-PARC neutrino beam in addition to Hyper-K to improve the sensitivity for the CP-violating phase and mass ordering [41, 635].
[182] In the schematic, the scintillation light appears to be collected by an *array* of SiPMs, while the text in [620] only refers to a *single* SiPM and the description of the field simulations in the same source also indicates that a single SiPM is intended.

In addition, field simulations were performed to find the cathode shape which minimizes the CTS. The lowest CTS of 0.8 ns was found for an ellipsoidal cathode. Currently, work to construct a prototype is underway.

The present state of the development is not sufficient to allow to evaluate the aptitude of this sensor for LSNDs yet.

Hybrid scintillator photosensitive gaseous detectors (HS-PGDs)

Hybrid gaseous photomultiplier (HGPM)[183]

The hybrid gaseous photomultiplier (HGPM) is based on a vacuum chamber that contains a photocathode and a scintillator (fig. 4.16f). The latter is bombarded using proximity focusing, and the scintillation light is detected by an MWPC-based PGD (cf. fig. 4.5c). This layout provides position sensitivity, is simple to manufacture, and does not require a large demagnification in the electron focusing as in HPDs (due to the small size of the diodes).

The vacuum chamber is a 4.5 mm thick disk with 30 mm diameter and contains a bialkali or Cs_3Sb photocathode. An EB voltage of 12 kV was used, and as scintillator a thin layer (200 µm) of metalized (50 nm Al) $KMgF_3$ on top of a CaF_2 window was employed. The emitted VUV photons (140–200 nm) pass through the window and hit an MWPC-PGD (readout chamber). This secondary sensor contains a CaF_2 window and a CsI cathode and is filled with CH_4 or an $Ar + CH_4$ mixture at 1 atm.

A peak QE of about 12.3% at 360 nm (10.6% at 420 nm) was measured, but the PDE was only ca. 1.4% at 410 nm (1.3% at 420 nm). The position resolution amounted to about 2 mm FWHM. No other properties were reported.

In a preliminary study, PMTs were used in place of the MWPC ("PM" in fig. 4.16f), which resulted in an HS-PMT sensor.

Even though only few properties are known, the very low PDE and small size already suffice to exclude this sensor for LSNDs.

4.3.1.2 Electron bombarded semiconductor

Hybrid photodiodes (HPDs)

Hybrid photodiodes (HPDs) are based on the EB of photodiodes.
The sensor types of this class can be distinguished based on their readout and gain layout. An HPD in which each PD has an individual readout channel is an HPD sensu stricto if the photodiode(s) contain no inbuilt amplifiers, and is termed EBCMOS instead if they do. On the other hand, if unamplified diodes are read out sequentially as in a CCD, the sensor is called an EBCCD. In this sense, the PD variants of the ReFerence (p. 679) and TransReFerence (p. 680) are HPDs sensu stricto, but since their layout differs significantly from other such sensors (integrated LC and reflective cathode), they are not considered members of this group here.

[183] Content based on [636, 637] [196; pp. 295–296].

In most HPDs, multiple diodes are used in form of a PDA. Strictly speaking, this would make the sensors hybrid PDAs. Likewise, the EBCMOS and EBCCD could be considered hybrid CMOS and hybrid CCD sensors. However, to point out the similarities between sensors – the use of EB PDs without in-pixel impact ionization – only one class is used here, namely HPDs.

HPDs sensu stricto

HPDs are a very popular sensor group. The earliest examples date back to 1965 [472] and HPDs have been used e.g. in the CMS hadron calorimeter and the LHCb RICH [184] detector. Correspondingly, many overviews of their general properties have been presented in literature [185], whose contents shall be summarized in the following.

HPDs (see fig. 4.13) combine the photon counting abilities and large detection areas of PMTs with the charge resolution and position sensitivity of segmented diode sensors. In these hybrids, a vacuum phototube contains a photocathode and one or more electron-bombarded silicon diode(s). The latter are read out with a separate chip, which is connected to the diodes with bonding or flip-chip assembly. This results in an image sensor with very good spatial resolution and flexible segmentation layout. The readout electronics can be included in the tube to improve the S/N and to reduce the number of feedthroughs. For this, however, the electronics must be compatible with the bake-out that is carried out before cathode deposition.

HPD sizes range between 1 and 15” diameter, where the fraction of the cathode area to the total sensor area is 50–80%.
The EB voltage has negative polarity – which can be problematic in a liquid – and typically lies within 10–20 kV (up to 30 kV). This provides a CE close to 100%.
Fountain, cross or proximity focusing are used (an early type was magnetically focused). The demagnification of fountain and cross focusing allows to use photocathodes that are larger than the diode area, which reduces the noise through smaller diode capacitances and leakage currents. Furthermore, with these focusing types higher gains of 5000–6500 are possible, but for magnetic fields above a few mT the spatial resolution deteriorates and the image distortion increases. For proximity focused HPDs, in turn, the small gap between cathode and diodes (down to a few mm) leads to very compact sensors (length below 3 cm) and improves the magnetic field tolerance up to above 1 T (especially for axial fields). However, this layout limits the applicable voltage, and thus only gains of ca. 3000 can be achieved.
The photoelectron energy is reduced by entry losses of typically a few 100 eV to 1 keV in the ca. 150 nm thick rear-side contact layer of the diodes (insensitive layer). The e–h pairs are produced close to the diode surface, and the gain shows a linear dependence on the EB voltage apart from the losses. This provides a better gain stability vs. voltage fluctuations than for PMTs (compare eq. (2.102)).
Since only one amplification step – the EB – is used with a very high gain, this leads to small total gains but very low charge resolutions. The latter are dominated by the noise of the readout electronics and are larger than for SiPMs. Thus, for photon counting low noise amplifiers with high gain are required. The readout noise is minimized by a small input capacitance and a long peaking time of the preamplifier. Therefore, the pixels should be as small as possible (this also lowers the dark current)

[184] Ring-imaging Cherenkov
[185] [196; pp. 292–298] [515; pp. 424, 428–429, 435, 436] [194; pp. 303–305] [170; pp. 17–18] [545] [153; pp. 297–300] [327, 472, 553, 558, 617].

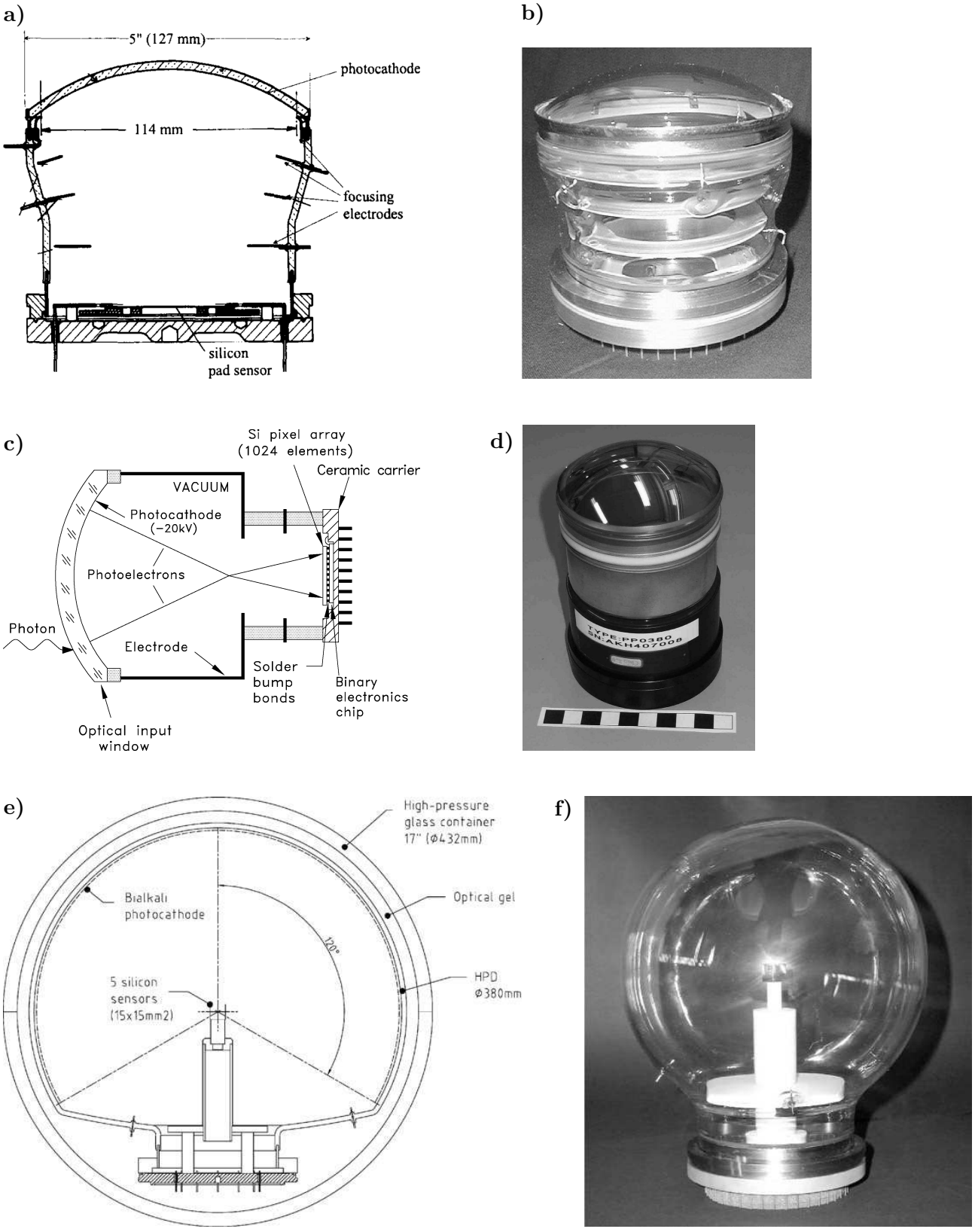


Figure 4.13: Hybrid photodiodes (HPDs). Discussed in text.
a)–b) Pad HPD (first generation): a) Schematic, from [472]. b) Photo, from [503].
c)–d) Pixel HPD: c) Schematic, from [638]. d) Photo, from [639].
e)–f) C2GT HPD: e) Schematic, from [499; p. 1127]. f) Photo of the half-size prototype, from [499; p. 1127].

and usually have side lengths in the order of 100 µm. ENCs down to 500 e^- can be reached for a detector input capacitance of a few pF and 100 ns peaking time. This provides well-separated npe peaks up to over 10 pe in the SER and low thresholds losses for single photons. However, this comes at the cost of a slow detector response, which strongly deteriorates the TTS.
In addition, (in)elastic pe backscattering at the silicon surface occurs with 17–18% probability (the backscattering coefficient k) for 10–30 keV energy. The energy spectrum of the backscattered electrons is relatively flat and on average only half the impact energy is converted into e–h pairs before scattering. The scattered electron is frequently not refocused onto the diodes (BL1-Tm). In this case, only an undersized PP is produced, which notably fills up the regions between the npe peaks in the SER [186], raises the charge resolution, and introduces a finite threshold efficiency (TE, see 2.1.3). The resulting ENF is similar to PMTs. The TE can be calculated by assuming a flat energy distribution of the backscattered electrons: [558]

$$\mathrm{TE} = 1 - k \cdot \frac{q_t}{\sigma_n \,\mathrm{S/N}} = 1 - k \cdot \frac{q_t}{q_1} \tag{4.7}$$

where q_1 is the spe charge, q_t the threshold, and σ_n the noise RMS of the pedestal. For e.g. $k = 18\%$, $q_t = 4\sigma_n$, and an S/N of 10 the TE is 92.8%. In addition to backscattering, the signal height can be reduced by charge sharing at the boundaries between pixels.
The small penetration depth of the pe in the diode leads to short rise and fall times before the amplifiers, typically a few ns down to < 1 ns. The achievable TTS strongly depends on the readout electronics. For large photon fluxes [187] and small peaking times, a TTS of as low as 50 ps was achieved, but for photon counting such numbers are not reachable. HPDs have less EPs, LPs and APs than PMTs.
These hybrids are comparatively insensitive to magnetic fields, and for proximity-focused sensors an axial field can even *improve* the performance, because it reduces both backscattering effects and the influence of field inhomogeneities at larger radii (see p. 670). Compared to PMTs, the temperature fluctuations and the power consumption (no divider current) are far smaller. Due to the latter, the power supplies can be simpler and cheaper than for PMTs despite the higher voltages.
However, the chemical incompatibility of the very different components of an HPD can be quite challenging. The aggressive alkali metals which are used to create the photocathode can attack the Si structures. This results in leakage currents from surface contaminants, which drastically increases the noise. For this reason, expensive transfer techniques are required during production, which makes HPDs more expensive than PMTs.

Due to their low gain, it is questionable whether HPDs can achieve both photon counting and a ns TTS at the same time, which basically eliminates this sensor class for LSNDs.

INTEVAC intensified photodiode (IPD) [188]

INTEVAC offered a commercial HPD called intensified photodiode (IPD) in the 1990s.

[186] For multi-pe events, the probability that all n photons create the full charge is only $(1-k)^n$. This leads to a broad quasi-continuous distribution which blurs the peaks at high n.
[187] In this case, however, statistical averaging improves the TTS for *any* photosensor, see eq. (2.131).
[188] Content based on [472, 640].

This HPD had an 8 mm diameter GaAs or GaAsP cathode which was grown with molecular-beam epitaxy (MBE) and showed a high QE of 45% at 500 nm. The photoelectrons were fountain-focused with 2–10 kV onto a single-pixel GaAs PIN diode with 1 mm diameter and an EB gain of ca. 1500. However, the HQE GaAsP cathodes made the sensor extremely expensive [641].

The Max Planck Institut für Physik (MPP) studied the applicability of this sensor for imaging atmospheric Cherenkov telescopes (IACTs) [640]. It was concluded that the gain was too low for spe detection with ns timing, since even the best available amplifiers had an ENC of a few 1000 e^- in the 1–10 ns time domain. To achieve a higher gain, INTEVAC produced a new sensor where the diode was replaced by an GaAs Schottky avalanche diode (HAPD class).
For 10 kV EB voltage and ca. 30 V diode bias, an avalanche gain of 10–15 and a total gain of $2.5 \cdot 10^4$ was measured. When using a filter time constant of 50 ns, up to about 6 pe were resolvable in the SER.

The small size, likely insufficient timing, and high costs eliminate both sensor variants from the selection for LSNDs.

Imaging silicon pixel array tube (ISPA-tube)[189]

The imaging silicon pixel array tube (ISPA-tube) was developed in 1994 as a result of the CERN RD7 and RD19 programs.
Here, for the first time the readout electronics was included in the vacuum tube, which allowed to use a finer segmentation of 1024 pixels (64×16 pixels, 75×500 µm size). The pixels were bump-bonded to a binary readout chip with a preamplifier and discriminator for each channel. A voltage of 20 kV was applied and the electrons were *magnetically* proximity-focused with a 500 G solenoidal field. The total readout times were between 6 and 12 µs.

The small size, the necessary magnetic field and possibly the long readout times exclude this sensor for LSNDs.

Pad HPD[190]

The Pad HPD (figs. 4.13a and 4.13b) was a photosensor prototype which was developed for the LHCb RICH detector in the late 1990s and early 2000s.

The sensor consists of a cylindrical vacuum tube with a 127 mm diameter window which is made of UV borosilicate glass window and has the shape of a spherical segment. Due to different thermal expansion coefficients, the tube and window are joined by a Kovar skirt.
The K_2CsSb cathode has an active diameter of 114 mm (81% of total sensor area), a peak QE of 15–20% at 350 nm, and $< \pm 10\%$ PDE uniformity. The cathode is applied by *external* evaporation before the tube is connected to the stainless steel base plate with cold indium sealing (transfer technique).
The photoelectrons are accelerated by typically −15 to −20 kV onto the Si sensor. Four stainless-steel ring electrodes provide fountain focusing with a demagnification

[189] Content based on [472, 506] [196; p. 295].
[190] Content based on [503, 642].

of 2.7 over the full active diameter. The first electrode is a so-called bleeder electrode and corrects non-linearities and cross-focusing effects near the cathode brink, while the second electrode controls the demagnification.
The secondary sensor is a 300 µm thick aluminized circular disk with 52 mm diameter (50 mm active) and 2048 pads with $1\times1\,\mathrm{mm}^2$ ($\mathrm{p}^+\mathrm{n}$ diodes).
At −20 kV, a gain of 5450 was measured, which showed a linear voltage dependence with ca. 2 keV losses in the inactive surface. Furthermore, a very good charge resolution (R_σ of Gaussian fit 6.5%, but notable backscattering between the peaks), a S/N of 19 and a TE of 90% for single photons were measured. A position resolution of 300 µm (σ) was found.
The signals are read out capacitively by low noise analog electronics which is enclosed in the vacuum tube. The electronics consists of 16 multiplexed IDEAS VA2 chips, which are wire bonded to 40 feedthroughs and have a low ENC of 285 e^- but a slow shaping time of 1.3 µs. Each channel contains a charge amplifier, a shaper and a sample-and-hold unit to determine the peak voltage. With analog electronics, lower ENCs than for digital readouts could be obtained.

Sensors with this layout (first generation) were mostly limited to <15kV due to imperfections of the ring electrode (sharp edges, rough surface, positioning) and alkali deposition on the side walls, which led to micro discharges and light generation. Therefore, a second-generation design with only three simplified electrodes was developed. Here, the bleeder electrode was a Ni wire with 2 mm diameter and the other two electrodes consisted of polished Ni rings with rounded edges. Furthermore, the cathode evaporation process was improved by placing screens around the sources. This produced QE peak values close to 26% at 350 nm (ca. 21% at 400 nm), resulted in an active area of about 90%, and allowed stable operation at −20 kV with low noise.
To achieve the required rate of 40 MHz, it was foreseen to use the SCTA128 chip with 25 ns peaking time but a higher expected ENC of 600–700 e^- in the final Pad HPD version.
Alternative designs included the VA-TAGP chip which is capable of self-triggering.

Judging from the known properties, this sensor appears to fulfill most requirements. However, even assuming the fast readout variant, the timing is very slow and the electronics is not designed for self-triggered detection with a TTS smaller than the readout rate. Therefore, the Pad HPD is not included in the short list.

Pixel HPD[191]

The pixel HPD (figs. 4.13c and 4.13d) was developed by CERN and Photonis-DEP in parallel to the Pad HPD for the LHCb RICH detector and was selected as the final sensor for this experiment (also [645]).

Here, a 75 mm diameter multialkali cathode is evaporated onto a 7 mm thick quartz window. A peak QE of 30% at 270 nm was achieved, which improved to over 35% during production. In early prototypes the QE was typically >20% QE between 250 and 450nm (typically 19% at 400 nm).
With cross-focusing fields at −20 kV EB voltage, a demagnification factor of ≈5 and an image distortion of <10% (which can be corrected offline) was reported. A gain of

[191] Content based on [545, 638, 643, 644].

about 5000 and an ENC of typically ca. 150–160 e^- was reached. Charge sharing and backscattering can lower the pixel charge below the threshold and reduced the TE to 84–89% (usually 85%).
The secondary sensor is 300 µm thick, has 16×16 mm^2 sensitive area and consists of a matrix of 8196 (256×32) pixels (p–n diodes) with 62.5×500 µm size, which are operated at a bias voltage of 80 V. Over 99% of the pixels were working. The small pixel size reduces the load capacitance and thus results in low noise and high speed. To reduce the charge sharing for operation in LHCb, 8 sub-pixels are coupled together, which results in 32×32 channels (1024 logical pixels) with 0.5×0.5 mm^2 size.
A binary CMOS readout chip (LHCBPIX1) with 25 ns peaking time [646] and 40 MHz clock is enclosed in the tube. The chip is bump-bonded to the diodes to minimize the detector capacitance and obtain a high S/N and a fast response. Each channel contains a charge preamplifier, a shaper and a discriminator and the readout must be triggered externally.
The spatial resolution is about 2 mm [647]. However, the position information can be incorrect if photons are reflected on the back of the window (1.4% probability) or from the Si sensor (2–6% depending on wavelength) onto the cathode.
Due to the multialkali cathode, the DCR was very high and ranged between 0.03 and 3 kHz/mm^2, in most cases < 1 kHz/mm^2.
iAPs were observed with ca. 250 ns delay and low rates $< 0.5\%$, which were attributed to the good vacuum of typically $1.3 \cdot 10^{-7}$ Pa.
The sensor is susceptible to magnetic field and requires a mu-metal shield for operation in 1–3 mT.

Regarding LSNDs, the pixel HPD has a too slow timing, a far too high DCR, is not-self-triggering, and the small size also could be problematic (many channels or in-detector front-end electronics). It is thus not included in the short list.

CMS HCAL HPD [192]

Another HPD was developed for the CMS hadron calorimeter (HCAL) to read out bundles of 2–18 wavelength-shifting fibers with an emission wavelength of ca. 520 nm in a field of 4 T. The development was performed in collaboration with DEP and Canberra between 1998 and 2003 by modifying an existing commercial 7-channel HPD of DEP. About 600 sensors were fabricated and passed acceptance tests, of which ca. 500 were installed [647]. The high magnetic fields narrowed the sensor choice to APDs and HPDs. The latter showed a better S/N, were less expensive and had smaller dependences on temperature and voltage fluctuations, despite having a lower QE than APDs. Recently, HPDs have been started to be replaced by SiPMs in the detector [652].

This CMS HCAL HPD [193] is proximity focused with an HV of 10–12 kV (losses 800–1200 V) and a hexagonal Si PDA (PIN diodes) with 19 pads (27 mm diameter, 5.4 mm pad size, 40 µm gap between pads, no dead area, 80 V bias). The pads are bump-bonded to feedthroughs to external readout electronics.
A second version with 73 channels was eliminated for cost reasons.

[192] Content based on [648–651].
[193] Schematic shown in in [648], photo in preprint version of [472].

A multialkali cathode with 11–18% QE at 520 nm (broad peak) was used, which had a required uniformity of 8%. The sensitive area was increased to 27 mm diameter to accommodate the fiber bundles, but this resulted in smaller distances between the diodes and walls, which led to HV breakdown for some devices.
A gain of typically 3300 (2500 after 10 a of operation) and a linear range (5% deviation) of $7 \cdot 10^4$ pe were required. In the detector, a voltage of 8 kV was applied, which led to a gain of 1500 [647].
The gap between cathode and diodes was reduced to 3.4 mm to reduce position inaccuracies from alignment errors in the magnetic field. Smaller distances than this involved the danger of discharges.
Crosstalk was found to occur due to three reasons:
First, without a magnetic field, pe backscattering occurred with 8% probability (double scattering was also observed) and a mean backscattered energy of 7 kV for 10 kV impact energy. However, an axial field of 4 T strongly reduced the maximum pe travel distance to ca. 100 µm and eliminated crosstalk completely. Without a magnetic field, alternatively, the gap could be increased so that most scattered electrons do not hit the PDA anymore.
Second, capacitive crosstalk (maximum allowed value of 3%) resulted from a poor connection to the bias voltage. This was reduced by coating the diodes with a 20 nm Al layer.
However, the high reflectance of Al (90% compared to 40% for Si) then caused optical crosstalk (maximum allowed value of 4%). This is produced by reflected photons which hit the cathode at a different position (γMdk, where d stands for diode). Therefore, an antireflective coating (14 nm a-Si:H) was applied, whose thickness was optimized for a minimal reflectance at 520 nm.
Regarding the pulse shape, the diode thickness was lowered to 200 µm to obtain a shorter pulse width (20 ns required) and length (30 ns required).
A readout rate of 40 MHz and thus a timing resolution of 25 ns were required.
The DCR depends on the red sensitivity of the cathode and ranged between 0.5 kHz/cm² and 1 MHz/cm² in prototypes; a maximum of 50 kHz/cm² was required.
Furthermore, micro-sparking was observed, which resulted in light emission. This could be measured with an external PMT through the window, and it was found that the rate depended on the HV.

The CMS HCAL HPD is excluded for LSNDs due to its low QE, large time resolution, high DCR, and small size.

TOM HPD[194]

The TOM HPD[195] is a larger version (10") of the 5" Pad HPD and was developed by CERN and INFN[196] for the CLUE cosmic ray experiment. For this work, the focus shall be put on the bialkali prototypes, since for the final version solar-blind Rb_2Te cathodes were foreseen, which are not relevant here.

The used glass tubes were originally designed and fabricated for the AQUARICH experiment and have a spherical segment window with a diameter of 250 mm and a

[194] Content based on [557, 653, 654].
[195] Named in memory of Tom Ypsilantis, one of the inventors of RICH detectors. In [557] the field distribution of the TOM HPD and a photo of the sensor are shown.
[196] Istituto Nazionale di Fisica Nucleare

total height of 275 mm. The 160 mm diameter base plate is connected to the tube with a cold indium press seal.
The bialkali cathode has an active diameter of ca. 228 mm and a peak QE of 24% at 350 nm (at 420 nm ca. 21.6%). The QE was uniform up to 60 mm radius and quickly dropped above this value, since a too tight Sb source mask was used during evaporation.
The TOM HPD uses fountain focusing with a demagnification factor of 4.02 for up to 100 mm radius. This is achieved by four concentric ring electrodes, including a bleeder electrode very close to the window.
An operating voltage of −20 kV was foreseen, but the voltage could in fact be raised to −30 kV. This resulted in a gain of 8100, an R_σ of 5% and a TE of about 95%. A perfect linearity between gain and voltage was observed with about 1 keV losses in the dead layers (Al metalization and n^+ implant). Furthermore, a point spread function (pe FWHM spot size on the Si sensor) of 1.2 mm was found.
As in the Pad HPD, a round Si sensor with 2048 p^+n diode pads and 50 mm diameter was used. However, this limited the detectable active area to 100 mm radius. Therefore, for the final sensor, a larger Si sensor with 65 mm diameter and ca. 256–400 pads with 4–3 mm size was planned. The 3 mm pads would have provided a photon position resolution of ≈ 12 mm.
An analog integrated readout of 16 VA-prime chips with 350 ns shaping time was used. In the final version, 2 self-triggering VATA chips were foreseen, and a faster version with 35 ns peaking time in the trigger channel was under development.
The total DCR was about 150 kHz (ca. 450 Hz/cm²) at −25 kV.

A 10” bialkali prototype was used in CLUE and a 5” Rb_2Te device was fabricated and tested as well [653], but no further progress was reported.

The very high DCR (306 Hz/cm² cross-section) and a slightly too low PDE at 420 nm (QE 21.6%, CE likely close to 100%, ME reduced by backscattering, TE 95%) exclude the 10” bialkali TOM HPD for LSNDs.

PET-HPD[197]

The PET-HPD[198] was also developed by CERN in the mid-2000s based on the experience gained with the Pad and TOM HPDs. In this case, the HPD was foreseen to read out a matrix of scintillator crystals for a novel positron emission tomography (PET) scanner concept.

The PET-HPD has a diameter of 127 mm, a total length of 67 mm, and a flat sapphire window with 105 mm diameter and 1.8 mm thickness.
The window is covered with a bialkali cathode (QE 22% at 330 nm, ≈ 20% at 400 nm [496]).
Proximity focusing with 12 kV is used to produce a gain of 3000 in a silicon sensor that is segmented into 208 (13×16) diode pads of 4×4 mm². The point spread function on the Si sensor is in the order of 0.3 mm.
The diodes are mounted on two integrated 128 channel front-end chips (VATA-Gp5) with self-triggering capability. Each channel contains a preamplifier [655] and two parallel chains. One uses a fast shaper with 25 ns peaking time, a discriminator and

[197] Content based on [645].
[198] For schematic and photo see [645].

a coincidence between multiple channels to trigger the readout. The other contains a slow shaper with 150 ns peaking time for the charge output.

Three prototypes were produced, where the first used the Pad HPD readout electronics, but none worked as foreseen. For the first prototype, surface discharges from the ceramic rings of the tube body occurred above 8 kV and the full HV could not by applied. For the other prototypes, the solder paste was not compatible with the vacuum bake-out and destroyed the readout. As possible solutions, a surface treatment of the ceramics to reduce the resistivity and the use of a compatible paste were suggested, but no further progress was reported.

With neither working prototypes nor continued development, this sensor is excluded for LSNDs. It is doubtful whether the achievable TTS would be sufficient, and the sapphire window likely leads to high costs.

C2GT HPD[199]

In addition, a large fountain focused HPD with 380 mm diameter was developed by CERN for the C2GT (CERN To Gulf of Taranto) conceptual study, using the experience gained from the Pad HPD. The C2GT project planned to intercept the CNGS (CERN Neutrinos to Gran Sasso) neutrino beam with an underwater WCD in the Gulf of Taranto to measure θ_{13}. The development of the photosensor was not finished – only an incomplete half-scale prototype was built – but was continued in form of the X-HPD (see p. 657).

The C2GT HPD (figs. 4.13e and 4.13f) consists of a 380 mm diameter borosilicate tube with 5 mm thickness and a spherical shape up to a zenith angle of 120°, which is closed by a metal baseplate with feedthroughs.
The glass is covered with a bialkali cathode (ca. 25% QE at 400 nm) with 120° viewing angle and beyond this angle with a shaper electrode.
A field of −20 kV focuses the electrons onto a grounded anode that consists of 5×4 Si diodes with 20 pF capacity each. These diodes cover five faces of a cube-shaped ceramic anode with 15 mm side length and are read out externally. Simulations showed a uniform angular acceptance up to ca 110°.
The anode is surrounded by a spherical field cage of ca. 30 mm diameter, which is mostly transparent to the electrons, to reduce the field gradient near the diodes (otherwise up to 10 kV/cm) and prevent discharges.
The EB gain would be ca. 4500 with a backscattering probability of about 18%.
A TTS of < 2 ns with photon counting up to at least 5 pe, a dynamic range of 6 pe, an ENC below 500 e^-, and a TE of 93% were planned. This was aimed to be achieved through low diode capacitances, 20–300 ns peaking time and waveform sampling with 40–300 MHz, at the cost of noise rates of up to 1 MHz. Field simulations showed a CTS below 1 ns, a focusing effect for the lateral pe momentum due to the strong field gradients, and insensitivity to the GMF.
Despite the segmented anode, no position/direction resolution was foreseen.

A half-scale prototype with 208 mm diameter and a cubic metallic anode was constructed in collaboration with Photonis-DEP (fig. 4.13f). First tests were planned

[199] Content based on [499; pp. 16–19] [645, 656].

to optimize the evaporation process, study the QE uniformity and test the HV stability. For this prototype, 5 unsegmented 10×10 mm² Si diodes were planned, yet their large capacitance of 36 pF would not have allowed to reach a TTS of 2 ns.

An alternative design planned to use Si APDs in place of PDs (HAPD type) to achieve a total gain of $1–3 \cdot 10^5$ (avalanche gain ca. 50, ENF 2–5). [617, 656] Due to the high capacitance of APDs (300–1500 pF/cm²) and the unavailability of devices with sizes above 5×5 mm², this was not realized, however.

Although the targeted sensor properties were quite good, neither of both variants produced functional prototypes and the development was stopped. Therefore, the C2GT HPD is not included in the short list.

Timepix/Medipix2 HPD

Several HPDs have been developed which are based on the Medipix-CMOS chip family as readout for Si diodes.

One design [500, 657] [200] foresaw a cross-focusing sensor with a 5 cm diameter cathode and sub-ns CTS, for which a position resolution of 0.5 mm FWHM was found in simulations.
The used CMOS sensor was a Timepix ASIC with 256×256 pixels with 55 µm pitch, 300µm thickness, and a clock of up to 100 MHz. The chip was bump-bonded to Si diodes [201]. Each cell contains a charge-sensitive preamplifier, a discriminator and a counter, which in timing mode counts the clock cycles until the frame end. This provides a time resolution which is equal to the clock period. During readout, a dead time of ca. 9 ms occurs.
Only a proximity focused prototype was produced, which had a CsI cathode (chosen since it is stable for about 30 min in air) and was tested with 6–20 kV voltage. At 20 kV, the main peak in the TTD showed a strong falling flank of LPs of up to >200 ns lateness. These resulted mostly from multi-pixel hits (likely from diffusion to neighboring pixels), which caused time walk. To reduce this effect, thinner diodes, a larger pixel pitch, or higher thresholds (at the cost of a moderate PDE reduction) could be used. For single pixel hits, a σ_T of 10.5 ns was found, which agrees with the clock frequency of 100 MHz. The CsI cathode caused a high DCR of 83.25 kHz/cm² at 20 kV.
Based on these findings, the development of an Photopix ASIC with good time resolution was planned for use in HPDs [202] and APP. This foresaw thinner diodes (100 µm), larger pixels (110×110 µm), a higher clock frequency (500 MHz), a shorter shaper rise time (ca. 40 ns instead of 130 ns), parallel readout (serial in Timepix) and two counters per pixel for dead-time free acquisition. This would allow digitized time stamps, 1% non-linearity at 40 MHz detection rate, and signal processing (e.g. zero-suppression).

A second design [203] [566, 658, 659] (in part by the same people) planned to use a proximity focused layout and a Medipix2 chip for spe detection.

[200] See references for a schematic.
[201] This makes this sensor an HPD and not an EBCMOS.
[202] The Timepix was actually developed for Micromegas GDs.
[203] A schematic layout is shown in [658].

Here, the CMOS has 256×256 pixels with 55 µm side length, is bump-bonded to 300 µm thick Si diodes, and is placed on a ceramic carrier. The chip has preamplifiers with ca. 13 mV/1 ke^- and a peaking time of 150 ns, which allows a count rate of 1 MHz. A 13-bit counter per pixel is used to provide a large dynamic range. With a clock frequency of 100 MHz and parallel readout, the complete sensor can be read out in 266 µs, so frame rates of > 3000 Hz are possible.
A prototype with a multialkali cathode (peak QE 10% at 450 nm, ca. 9.5% at 400 nm; HQE cathodes were foreseen for future devices) was constructed. This resulted in a high DCR of ca. 30–40 Hz/cm². The EB voltage was 5–10 kV, and frame rates above 3 kHz could be shown. However, significant IBF limited the attainable position resolution. The IBF could be reduced through higher bake-out temperatures or by use of a focusing field layout with ion traps.

While the Photopix ASIC is a very promising design, the final state of development of both the Timepix and Medipix2 variants falls short of the TTS, DCR, and QE (for Medipix2) requirements of LSNDs. Furthermore, the Medipix2 HPD had unresolved problems with iAPs. Thus, these HPD types are not included in the shortlist. Since the Medipix/Timepix series has made great advancements since then (see p. 647), future developments of HPDs that are based on newer chips should be closely watched, however, because they yield the promise of an excellent photosensor.

ReFerence (HPD variant)

See p. 679.

TransReFerence (HPD variant)

See p. 616.

Electron bombarded CCD (EBCCD)[204]

An electron bombarded CCD (EBCCD) is an HPD which contains a photocathode whose photoelectrons are accelerated onto a CCD (fig. 4.10d). It is thus a CCD with extra EB gain and combines the spe sensitivity of phototube sensors with the low position resolution of CCDs [612].
The first EBCCDs were developed in the 1980s for LLL detection in astronomical and military night visions applications. The first commercial device became available in 2000 from Hamamatsu.
Current uses comprise microscopy, biological low-light imaging, optical spectroscopy, and radiography.

Usually, proximity focusing is used with gaps of about 1 mm, which results in flight times of only a few 10 ps between the cathode and CCD. In [501] also a sensor with a large 10 cm diameter cathode and *cross* focusing was reported. This variant had a bialkali cathode, −12 kV, a gain of 2400, a demagnification factor of 5, a spatial resolution < 250 µm, and a frame rate of 30 Hz.
The CCD is normally bombarded on its thinned back side, since the front side layers of electronics would cause too large energy losses. It is also covered with a thin Al layer, which introduces a threshold energy for the detection of photoelectrons. The Si chip is wire-bonded, must fulfill cleanliness and outgassing requirements, and has

[204] Content based on [566].

to be compatible with vacuum bake-out. The pixels are typically large, since due to the EB gain their potential well can already be filled by a few photons. This limits the dynamic range.
The QE is lower than for CCDs but can reach about 50% with GaAsP cathodes. The FF is close to 100%.
An EB voltage of typically a few kV (maximum HV 14–15 kV) is used, which results in similar gains as for HPDs. As for the latter, spe detection is possible, the SER shows narrow peaks due to the EB gain, and backscattered electrons produce undersized pulses.
The electrons which are created in the diodes diffuse to the front of the pixel, during which charge sharing into adjacent pixels occurs. However, the position can be reconstructed with sub-pixel resolution through centroid algorithms, and the position sensitivity is thus limited only by the photoelectron point spread function. With iterative fitting algorithms, even better resolutions can be reached.
The readout speed is limited by the CCD and potentially the complexity of the position reconstruction algorithm. The maximum frame rate is up to 200 Hz. In [566] a possible timing improvement was presented which is based on exploiting the low charge resolution: An EB voltage sweep could allow to determine the photon arrival time from the pulse height with a time resolution below the frame length.
The DCR is determined by thermionic emission from the cathode.
Furthermore, large iAPs can occur.
The sensor lifetime is limited by damages to the chip from the energetic photoelectrons. These also produce bremsstrahlung and ions, both of which reduce the cathode lifetime as well. Compared to sensors with MCPs, on the other hand, the comparatively small volume and inner surface increase the cathode lifetime. In sum, sensor lifetime values of 10^{12} counts/mm^2 were reported, which is a factor of ten higher than for MCP devices.

Due to the slow readout, EBCCDs are not suited for LSNDs.

Electron bombarded CMOS (EBCMOS)[205]

Electron bombarded CMOS (EBCMOS) sensors (fig. 4.10d) have the same layout and similar properties as EBCCDs, except that here a back-bombarded CMOS sensor is used instead of a CCD.
EBCMOS sensors were first developed for military night vision devices in 2001, first used for science (particle physics) in 2007, and are not yet widely available.

Since each pixel (or row of pixels) has its own amplification and readout, faster frame rates than with CCDs can be achieved. Up to 500 Hz have been reported and 1000 Hz is planned. However, in the EBCMOS the photon arrival times cannot be directly retrieved either.
The pixel noise can be as low as $1\,e^-$ [612].

The slow readout prohibits a use in LSNDs.

Hybrid avalanche photodiodes (HAPDs)

A hybrid APD (HAPD) is a phototube that contains a photocathode and an electron-bombarded Si avalanche diode or array of APDs (see fig. 4.14)[206].

[205] Content based on [566].
[206] Content based on [478; p. 270] [515; pp. 426–428] [553] [194; p. 304] [153; p. 300].

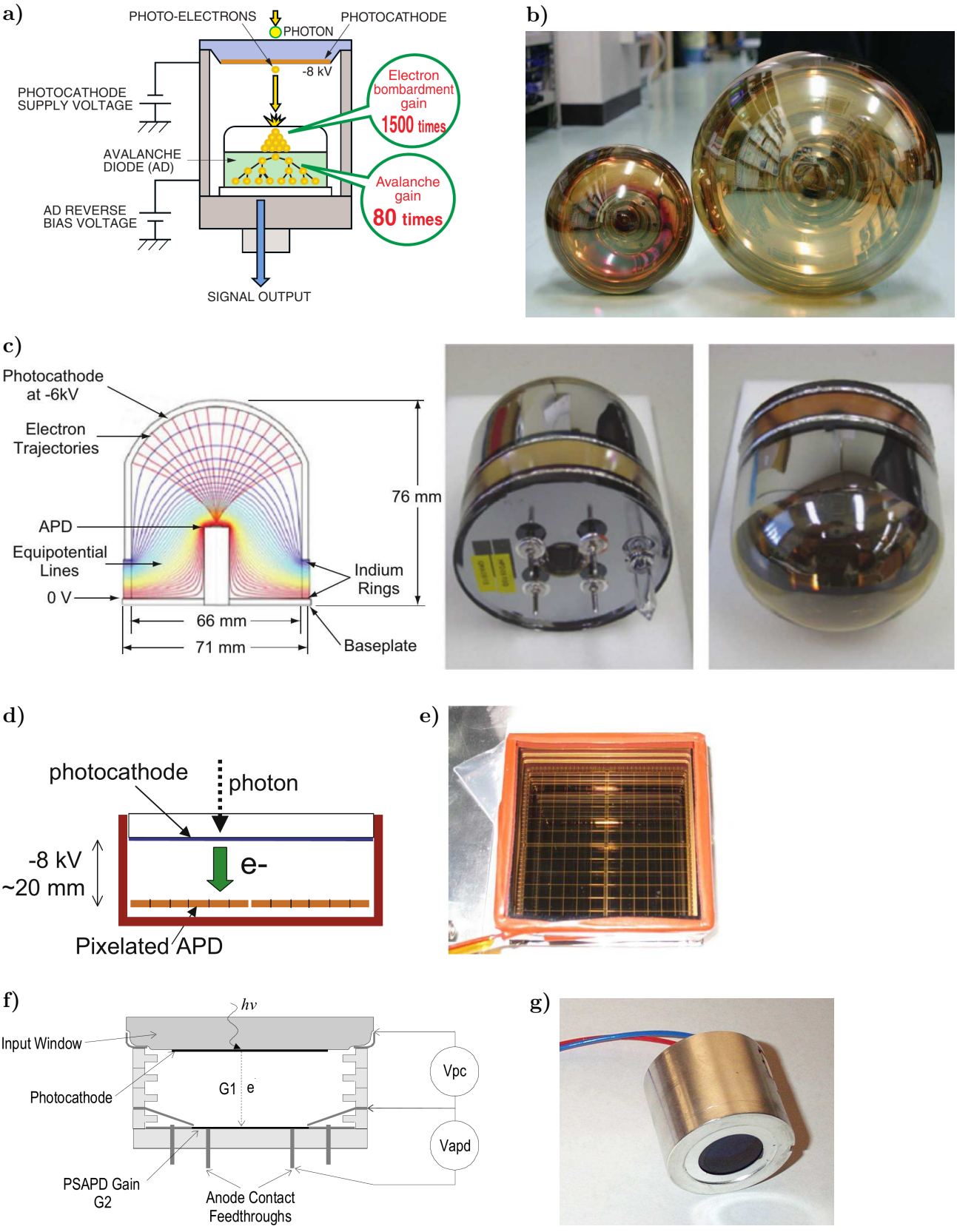


Figure 4.14: Hybrid photosensors based on electron bombardment of avalanche photodiodes (HAPDs). Continued on next page, discussed in text.
a) Principle of operation of an HAPD: Shown for a Hamamatsu hybrid photodetector (HHPD). From [169; p. 230].
b) HHPDs: 8" (R12112) and 13" prototypes, from [660].
c) QUartz Photon Intensifying Detector (QUPID), from [661].
d)–e) Belle II HAPD: d) Layout. e) Photo. Both figures from [662].
f)–g) Hybrid position sensitive avalanche photodiode (HPSAPD): f) Layout. Compare the PSD (figs. 4.9b and 4.6g). g) Photo. Both figures from [663].

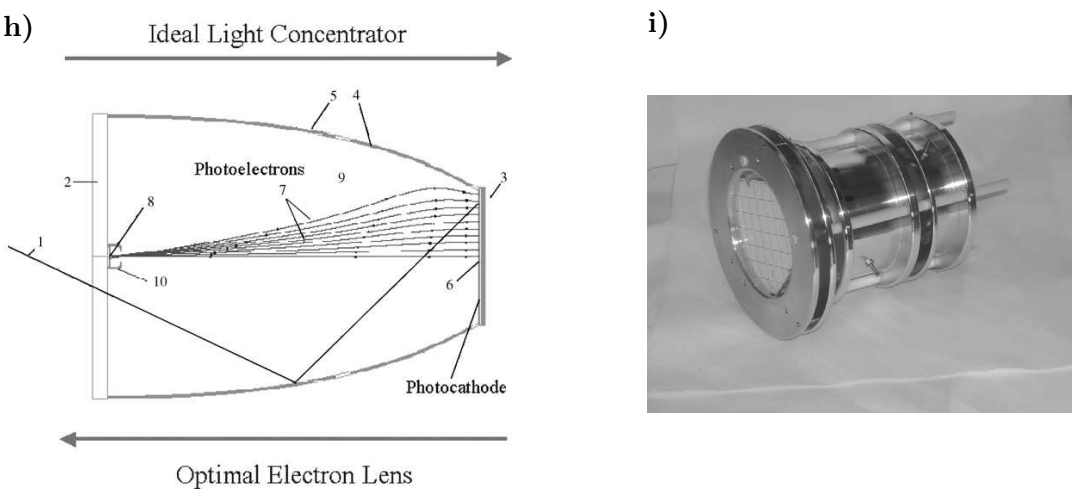


Figure 4.14: Hybrid photosensors based on electron bombardment of avalanche photodiodes (HAPDs). Continued from last page, discussed in text.
h)–i) ReFerence: h) Schematic: Incident photons (1) are focused by a compound parabolic concentrator (CPC, 4) from the entry aperture (2) to a smaller exit aperture (3) which is covered by a reflective photocathode (6). The produced photoelectrons (7) are focused onto an APD or PD (8) by two electrodes with the same shape as the CPC (4, 5), which are separated by a small gap (9). The APD is surrounded by an ion-collecting electrode (10). i) Photo of prototype. Both figures from [664].

The first HAPDs were developed in the 1990s for particle physics [478; p. 270]. While initial devices still had very small active diameters of only 8 mm [640, 665], by now HAPDs with up to 20" diameter have been produced (see Hamamatsu HPDs on p. 683).

The diodes contain of a thin p-doped layer that faces the cathode. This is followed by an intrinsic region, through which the electrons drift that are produced through ionization. A p–n junction is placed at the rear where the avalanche multiplication occurs.

A large negative EB voltage of typically 6–8 kV is used, which results in an EB gain of up to about 2000. Together with an avalanche gain of ca. 50–300, a total gain of several 10^4 up to ca. $2 \cdot 10^5$ results.
Due to the high first stage gain, the charge resolution is better than for PMTs, but because of the high ENF of avalanche multiplication (see p. 625) the charge resolution is worse than for HPDs. In sum, the npe-peaks can be well distinguished up to a few photons.
Since the gain is still comparatively low, additional amplifiers have to be used for photon counting. Due to the avalanche gain, external amplifiers are sufficient and no internal readout chips have to be used, as is commonly the case for HPDs with the mentioned disadvantages.
HAPDs also have high CE, shorter transit times than PMTs, good TTS, low AP rates and a comparatively low susceptibility to magnetic fields – especially for proximity focused layouts, which can operate in over 1 T.
While [553] states that HAPDs are expensive, because a transfer technique is necessary for fabrication, [666] says that the simple layout makes them cheaper than PMTs. Furthermore, Hyper-K started the development of large HAPDs as PMT replacements

with the explicit goal of *lower* production costs, since no dynodes are required [41; p. 124]. Therefore, it is assumed that in sum the costs are comparable or lower than for PMTs.

INTEVAC IPD (HAPD variant)

See p. 667.

ReFerence [207]

The ReFerence (figs. 4.14h and 4.14i) is an HPD/HAPD/HS-SiPM sensor which was developed between 2001 and 2007 by D. Ferenc and the University of California. The goal was an application for APP, in particular for WCDs and IACTs.

The centerpiece of this sensor is a LC (compound parabolic concentrator, CPC) with an acceptance angle of 30°, which reflects photons that enter through a window in the entry aperture onto a reflective photocathode. In addition, the CPC is made of two electrodes, which are separated by a small isolating gap and also focus the *photoelectrons* onto a small diode sensor on the entrance window. Here, either a PIN diode (HPD class) or an APD (HAPD class) is used. To reduce the IBF, this secondary sensor is surrounded by an ion collecting electrode. The IBF protection has the highest voltage, followed by the diode, the electrode close to the window, the electrode next to the cathode, and the cathode. The optimal gap position was determined through electron optics simulations. It was planned to incorporate a magnetic shield as well.
In another variant, the photoelectrons instead hit an Al-coated scintillator film, which is fiber-coupled to an SiPM (HS-SiPM class). To exceed the charge range that is affected by the DCR of the SiPM, a dpy of at least 30 pe was foreseen (achievable with an EB voltage of over 15 kV).
In addition, it was planned to construct a photosensor that consists of a flat panel which contains a matrix of hexagonal ReFerence sensors (hexagonal entry window, circular cathode). This layout maximizes the FF and provides a high pressure and shock resistance. For a fully developed industrial production, a very coarse cost estimate assumed a price of $< \$1000/m^2$ for this panel-ReFerence.

One of the targets of the ReFerence is to reach a high PDE. This is achieved by a large FF, a CE close to 100%, and the use of a reflective instead of a transmissive cathode, which allows to reach a higher QE that also extends to shorter wavelengths. Furthermore, it was planned to achieve photon counting and a very good TTS (despite the flat cathode surface). The LC provides a limited field of view, which is desirable for many detectors. In addition, the DCR was supposed to be reduced through a thermally isolated cooled cathode. Through the double focusing (photons and electrons), a concentration factor of about 1000 from window to secondary sensor was targeted.
However, it should be pointed out that the LC leads to photon collection losses and contributes a notable optical jitter to the TTS (see p. 618). In addition, the flat cathode leads to a non-negligible CTS, which is visible in fig. 4.3.1.2 in form of the tiny squares that mark isochronous positions of the photoelectrons. Furthermore, significant IBF should occur from the region between IBF protection housing and cathode as well.

[207] Content based on [519, 641, 664, 667].

These drawbacks might be the reason, why the construction of the ReFerence was discontinued and instead the development of ArcaLux and ABALONE were pursued by the same group of people.

Only partially functional prototypes were constructed.
The pe focusing was tested with an unsealed tube (3" entry and 1.5" exit aperture, Al LC) by replacing entry window and diode with a phosphor screen (as in an image intensifier) and using a movable point-like cathode. The latter was realized by substituting the photocathode plane with an Al plate with a 2 mm pinhole with 45° inclination (to avoid direct illumination of the phosphor). The plate was illuminated from the back with UV light, and a clean Al surface in the pinhole served as photocathode (see p. 56). This tube was tested in a vacuum chamber for 9.5 and 23.5 kV EB voltage for pe emission positions between center and brink. For optimized voltage settings, the position of the pe spot image on the center of the phosphor was nearly independent of the cathode position (< 1 mm movement). A concentration factor of ca. 1500 was measured. The pe focusing could also be demonstrated for hexagonal entry windows.
A first 5" panel prototype was produced, which contained 7 pixels in an Al plate and a glass vacuum enclosure but had no cathode yet. The development of a fully functional all-glass panel prototype was started.

Since no functional prototypes were produced and development was stopped, the various ReFerence subtypes are not considered for LSNDs.

TransReFerence

The TransReFerence [561] (fig. 4.8a) consists of 2–3 coupled ReFerence sensors with PDs or APDs with the target to achieve wavelength-resolved photon detection over an extended spectral range. This sensor was already discussed on p. 616.

The TransReFerence only reached the concept study stage and no prototypes were produced, likely because the development of the underlying ReFerence was never finished.

For lack of prototypes and ongoing development, this sensor is excluded for LSNDs.

C2GT HAPD

See p. 673.

ArcaLux (HAPD variant)

See p. 660.

QUartz Photon Intensifying Detector (QUPID)[208]

The QUartz Photon Intensifying Detector (QUPID) is a 3" ultra-low background fountain-focused HAPD (see fig. 4.14c).

[208] Content based on [504, 661, 668].

This sensor was developed by the University of California, Los Angeles (UCLA) and Hamamatsu for ton-scale noble liquid-based DM and double beta decay experiments. Since these detectors are limited by the intrinsic radioactivity of their photosensors (mostly PMTs), the goal was to create a sensor with a background that is at least an order of magnitude lower than for PMTs. The vacuum tube of the QUPID thus is made entirely of quartz (ultra-clean synthetic fused silica). At the same time, the photosensor should provide high QE and CE with good uniformity and a relatively large sensitive area, photon counting, large linear range, and low TTS.
In a later stage a 6" QUPID was foreseen [669].
Development of this sensor was terminated by Hamamatsu in 2012 [556; arXiv version] after 7 fully functional prototypes were manufactured in an early production line. Nevertheless, the QUPID lives on in two HSiPMs which are based on its design: The SiGHT (p. 692) has a very similar layout, where an SiPM replaces the APD to lower the EB voltage, and the same UCLA group was involved in its development [670]. And the VSiPMT (p. 691) [671; journal version].

The top part of the QUPID consists of a cylindric quartz vacuum tube (71 mm diameter) with a hemispherical window (radius 37 mm). The cylinder is coated with Al and the window is covered with a photocathode (effective diameter 64 mm). A fused silica baseplate is joined to the tube by a quartz ring with indium seals, which are also used to apply the EB and ground potentials. A cylindrical quartz pillar with a 3 mm back-illuminated APD on its top is placed on the baseplate. The total height is 76 mm.
A low temperature bialkali photocathode is used. This cathode type was developed by Hamamatsu for low resistivity in liquid xenon (LXe) and liquid argon (LAr) and a peak QE $> 30\%$. The LXe version reaches the highest QE for 178 nm (Xe scintillation light), while the LAr cathode is optimized for wavelength-shifted Ar scintillation light (peak QE 39% at 380 nm, ca. 37% at 420 nm). A output uniformity of ca. $\pm 20\%$ across the cathode was measured.
The CE was $> 80\%$ for the majority of the cathode with a uniformity of less than $\pm 10\%$. The design was optimized to achieve pe focusing independent of the cathode voltage.
An EB voltage of -6 kV resulted in a typical gain of 750 (max. 800) with a threshold energy of about 3.6–4 kV [209].
The APD was specifically developed by Hamamatsu for the QUPID and has 11 pF capacitance, a breakdown voltage of 360 V at 25 °C and an avalanche gain of 200 (maximum 300, rises with falling temperature). The total gain is typically $1.5 \cdot 10^5$ (max. $2.4 \cdot 10^5$).
A good charge resolution was found along with the usual ca. 18% pe backscattering, which allowed to discern up to 3–4 pe in the SER. The P/V was about 6 at room temperature for -6 kV (gain $1.15 \cdot 10^5$) and rose to 7.5 at -8 kV (gain $1.7 \cdot 10^5$).
For a light pulse with 1 µs width (used to simulate scintillation light), a linear range (5% deviation) of ca. 10^5 pe was found.
External readout is used, and for the timing characterization even a 4 m coaxial cable was put between QUPID and electronics to simulate the conditions in a large detector. The resulting rise time was 1.8 ± 0.1 ns, the fall time 2.5 ± 0.2 ns, and the

[209] For Si sensors, the EB gain shows a linear dependence on the acceleration voltage except for energy losses in the insensitive surface layer. The threshold voltage/energy is defined as where the extrapolated linear curve crosses the voltage/energy axis [169; p. 230].

pulse width 4.20 ± 0.05 ns. An excellent T_{FWHM} of $< 160 \pm 30$ ps (including the laser trigger jitter) was measured.
At 20 °C, the DCR amounted to 4 kHz at a threshold of 1 pe, ≈ 7.5 kHz for ca. 0.75 pe (still sufficient for spe detection with high TE), and ≈ 55 kHz for 0.5 pe. The DCR dropped to 70 Hz at −150 °C for 1 pe threshold.
The intrinsic radioactivity was 17.3 mBq for ^{238}U, 0.3 ± 0.1 mBq for ^{232}Th, 0.4 ± 0.2 mBq for ^{40}K, and < 0.18 mBq for ^{60}Co.

Planned next steps included to optimize the production process to obtain a higher EB voltage, since this would provide a higher gain, improved charge resolution, and better S/N. Furthermore, it was targeted to analyze the APs and crosstalk and to further reduce the radioactivity by a screening of the components.

The QUPID shows excellent properties except for its high DCR, which likely is caused by the low temperature bialkali, and the negative EB voltage, which could be problematic in liquids. It is thus included in the shortlist.

Belle II HAPD [210]

Another HAPD [211] was developed for the aerogel RICH detector of Belle II in collaboration with Hamamatsu between 2008 and 2016 (figs. 4.14d and 4.14e). The sensor requirements included a low height, low noise, immunity to 1.5 T axial field, radiation hardness, spe detection at 400 nm, and 5 mm 2D position resolution. For this purpose, several photosensors were developed, including a proximity-focused HAPD with APD arrays and external readout, which in the end was selected for the detector. From 2013 to 2016, over 500 such HAPDs were produced by Hamamatsu, of which 420 were installed in the detector.

The Belle II HAPD consists of a square ceramic vacuum tube of 73×73 mm^2 and 28 mm height. A super bialkali (SBA) cathode with 70×70 mm^2 effective area is placed on a quartz window and showed a mean QE of 31.7% at 400 nm (improved during fabrication). The total FF was 65% when including the APDs.
The photoelectrons are proximity-focused (gap ca. 20 mm) with high CE by typically −8 kV EB voltage onto a pixelated 144-channel APD. The latter consists of four APD arrays with 6×6 pads, 4.9×4.9 mm^2 pad size and 5 mm pitch. This results in an EB gain of about 1500. Together with an avalanche gain of typically 50 (bias voltage 250–500 V), a total gain of over $7 \cdot 10^4$ resulted.
The charge resolution is better than in PMTs but worse than in HPDs ($R_\sigma = 17.4\%$ for an early prototype [545]), and in the SER up to at least 4 pe can be discerned. As for other sensors with pixelated EB diodes, pe backscattering causes crosstalk.
The avalanche gain relaxes the requirements for the four readout ASICs (36 inputs each). These are placed outside the vacuum tube (attached to the HAPD back) and contain a preamplifier, shaper (about 100 ns shaping time) and discriminator for each channel.
The pulse height usually ranges from 5 to 10 mV and the pulse width amounts to a few 100 ns.
In a part of the devices, the dark current increased a few months after production, and for several devices extremely large noise rates occurred with irregular periods and unclear origin. Unstable or noisy sensors were rejected.

[210] Content based on [612, 662, 672].
[211] Here the acronym HAPD stands for Hybrid Avalanche Photo Detector.

An axial field of 1.5 T reduces the crosstalk from backscattering significantly, which is visible as a strong reduction of the events between the peaks in the SER. This field also considerably reduces the image distortion from non-uniform fields near the edges. Magnetic fields which are parallel to the cathode must be avoided, since they would shift the image and cause a part of the active area to be lost. In the foreseen magnetic field, some HAPDs showed large (500 mV), long (> 100 µs) signals, which produced dead times in the readout electronics. The frequency of occurrence of these signals was smaller for lower HV. As origin, sparking from the ceramic tube was suspected, and a reactivation of the tube getters to adsorb residual gas significantly reduced these issues.

Unfortunately, no information could be found regarding TTS, TTD, DCR and AP rate. This renders it difficult to arrive at a conclusive assessment of the aptitude of the Belle II HAPD for LSNDs.
Due to the proximity focusing, an initially excellent TTS is expected, which could be impaired somewhat by the shaper. Since the APD size is small, the detector capacitance of each channel should not be high, however. Backscattering will introduce a significant amount of LPs. In addition, the DCR should be quite high due to the HQE cathode. Furthermore, the negative voltage could be problematic for an operation in liquids. Additional measurements should be carried out to ascertain the missing parameters, as this is a thoroughly tested, commercially available sensor with promising properties.
Since the Belle II HAPD otherwise shows good performance, it is preliminarily included in the selection.

Hamamatsu hybrid photodetectors (HHPDs)

Hamamatsu has been developing and commercializing HAPDs with ever increasing sizes since the 1990s [665]. While the first device had an effective cathode diameter of only 8 mm, the product range now spans 3 mm to 50 cm [606, 673] and the Belle II HAPD and the QUPID were also developed in collaboration with Hamamatsu.
In most publications, the manufacturer refers to its HAPD products as “hybrid photodetectors” (abbreviated as HPDs). To distinguish these sensors from hybrid photodiodes, while keeping the denomination chosen by the inventor, they will be called “Hamamatsu HPDs” (HHPDs) here.
Apart from particle physics, applications of HHPDs include laser microscopy, fluorescence lifetime imaging microscopy, fluorescence correlation spectroscopy, and light detection and ranging (LIDAR).

In general, HHPDs allow photon counting and show low charge resolutions, good to excellent TTS, high linearity, high CE, and good PDE uniformity [212].
A negative bombardment voltage of several kV up to about 10 kV is applied.
The breakdown voltage of the APD is defined as the point of 1 µA leakage current. Near this voltage, strong gain variations are observed, so a voltage of at least 10 V lower is recommended. Breakdown voltage and leakage current decrease with falling temperatures, while the avalanche gain rises.
The time response is determined by the junction capacitance of the APD, which is minimal for full depletion. Therefore the rise and fall time are nearly constant vs. voltage variations as long as the diode remains fully depleted. Due to the large

[212] Paragraph based on [169; p. 46, ch. 12].

number of charge carriers which are produced by EB, the TTS contribution from the APD can be neglected.
Ionic APs produce large pulses but have a very low rate due to the simple layout and stable high vacuum.
EB can also produce bremsstrahlung and characteristic X-rays, which emit a large number of electrons if they hit the cathode (X-ray feedback). In small fast HPDs with crystalline cathodes (GaAsP(Cs) and GaAs(Cs)), this AP type is observed with ca. 500ns delay but only 0.01% probability. The AP charge and rate depend on the cathode type, and the probability is lower for the thin alkali-based cathodes.
The sensor lifetime is determined by EB gain degradation and cathode deterioration (e.g. from IBF) and is lower for crystalline cathodes than for alkali cathodes.

For this work, the focus is on the largest HHPD series, which are most promising for LSNDs – in particular the 20" R12850, which is currently being developed. 20" HHPDs should also have the lowest cost per sensitive area [165] and cost less than PMTs due to the simpler layout [666]. The other types will be addressed only shortly.

R12850[213]

The R12850 (sometimes called R12850-HQE) is a 20" fountain-focused HAPD which is being developed for use in the Hyper-K WCD. The requirements include photon counting, low TTS, high PDE, low DCR, high pressure tolerance, long lifetime, and the absence of discharges.
Unlike most HAPDs and HPDs, the R12850 has a *positive* bombardment voltage with a grounded cathode, since it is to be placed in water. This requires to capacitively couple the APD output to the readout electronics.
The main design difficulties came from the large EB voltage and the high junction capacitance of the APD.
This sensor has produced fully functional prototypes with excellent properties, ongoing improvements and long-term stability testing in a 200 t WCD (EGADS[214]) since 2017. However, a mass production technique remains to be established.

The phototube of the R12850 has a diameter of 508 mm and a length of 622 mm plus 87 mm for electronics.
The cathode has a minimum effective diameter of 460 mm and a peak QE of around 36% (34–37.5% at 360–390 nm for four samples) with ca. 33% at 420 nm.
Simulations showed a CE of 97% for 46 cm diameter and 80% for 50 cm.
An acceleration voltage of +8 kV is applied (reduced from 12 kV in smaller prototypes), which results in a bombardment gain of 1600.
The APD has 20 mm diameter (smaller sizes would reduce the CE), a capacitance which was reduced from 800 to 400 pF through a larger diode thickness, and an avalanche gain of 50–100 (bias voltage 400–500 V). This results in a total gain of 10^5. Since the APD capacitance determines the spe time response and the charge resolution, in addition, segmented APDs with 2 and 5 readout channels (chs) were developed to further lower the capacitance (200 pF/ch for 2-ch, 160 pF/ch for 5-ch). For the 5-ch APD (one central channel surrounded by four annulus sectors), a dependence between photon hit position and APD channel was observed. This could allow a limited position-sensitivity and CTS correction.

[213] Content based on [41, 165, 279, 440, 606, 674].
[214] Evaluating Gadolinium's Action on Detector Systems

The charge resolution R_σ is 15% for a 1-ch and only 10% for a 2-ch or 5-ch APD with individual readout channels, which allows to clearly discern up to 6 pe.
A 5% gain decrease was found at an event rate of 1.2 MHz for light pulses with 50 detected photons.
The readout electronics is external and is placed at the rear of the sensor. It contains a specifically developed fast low noise preamplifier, a digitizer, and optionally a built-in power supply for the EB and APD voltages. To minimize heat production and increase the lifetime, alternatively a watertight connector to a coaxial cable that carries both HV and signal can be used.
The T_{FWHM} is 3.6 ns when using a leading edge discriminator, 3.2 ns with time walk correction, and is limited by the noise and rise time of the preamplifier. The properties of the amplifier have less influence for signals with many photons, whose TTS is significantly lower than expected from statistics alone (compare eq. (2.131)). Extrapolation down to 1 pe shows a TTS of only ≈ 1.8 ns before the amplifier. Thus, to lower the spe TTS, an improvement of the preamplifier is foreseen. The CTS for a 5-ch device showed that the cathode–diode transit times decrease for growing zenith angles, where the maximum difference to the center value of −6 ns was reached at 70°. This could be lowered by an adapted cathode curvature.
In the TTD, a small trailing flank of LPs is visible, which extends to a lateness of ca. 18 ns. This likely originates from pe scattering.
The pulse height is ca. 10 mV and the pulse length about 100 ns.
The DCR is ca. 8 kHz (1-ch, 0.5 pe threshold, 25 °C) but might be further reducible.
An AP rate of $< 2.3\%$ is reported.
Simulations showed a pressure tolerance of up to 250 m water column if the most sensitive part – the tube stem, which can withstand only 88 m – is encapsulated. Three tubes were subjected to a pressure of up to 1.25 MPa without damage.

The known properties of the R12850 are mostly excellent and fulfill the LSND requirements. The T_{FWHM} is slightly above the demanded 3 ns, but might be improvable through a better amplifier or an adapted bulb shape. The DCR is larger than the required 4378 Hz but within a factor of 3, which is the expectable reduction from long-term operation in dark. Therefore, the R12850 is included in the shortlist as a strong candidate.

13" prototype[215]

Before the R12850, a 13" (332 mm diameter) prototype without series name with fountain focusing, a 5 mm diameter APD [677] (40 pF detector capacitance) and a metal baseplate was developed for Hyper-K (see fig. 4.14b).

This sensor had an effective cathode diameter of 300 mm [678] and a QE of 20%. A positive acceleration voltage of 10–20 kV resulted in an EB gain of 4500 at 20 kV (3.5 keV threshold energy). An avalanche gain of ca. 50 (390 V bias) was observed. This resulted in a total gain of about $(1\text{–}2) \cdot 10^5$ with a uniformity of $< 2\%$.
At 20 kV the simulated CE was 97% (95% at 15 kV).
In the pulse height distribution (PHD), the resolution is 11% σ (24% FWHM) and pe backscattering is visible.
Furthermore, a dynamic range of 3000 pe was reported.

[215] Content based on [675, 676].

The TTS σ_T is only 190 ps with a maximum CTS difference of ≈ 500 ps. The rise time is ca. 1 ns.
An external low noise charge-sensitive preamplifier with 3400 e^- ENC is used and the waveform is sampled with an analog memory cell (lower power consumption than FADCs) [679]. The front-end electronics also contain a HV supply and digital signal processing. The total power consumption of the sensor is significantly lower than 700 mW.

The PDE is slightly below the LSND requirement of 20% (especially when considering the backscattering losses), but the TTS and charge resolution are very good. The 13" HAPD is therefore included in the short list for comparison.

R12112 [216]

The R12112 (fig. 4.14b) is a fountain-focused 8" HAPD, whose development (for Hyper-K as well) was started between the 13" prototype and the R12850 and apparently still is continued. Eight samples were produced for long-term testing in EGADS.

The sensor consists of a 20 cm diameter borosilicate glass tube with a 5 mm diameter APD and has a total length of 30 cm including electronics.
The bialkali cathode showed a peak QE of 35.7% at 380 nm and the QE was ca. 33% at 420 nm. The CE is 95%.
The bombardment voltage ranges from +8 to +10 kV (gain ≈ 1600). The APD has a bias voltage of 200–370 V, which led to a total gain of $(4\text{–}9) \cdot 10^4$.
In the SER, the R_σ was 12.8%, the P/V was between 3.6 and 5.4 and a backscattering fraction of 19.2% was found.
Hamamatsu states a σ_T of 620 ps with an expected CTS difference of 210 ps from 0–70°.
A dynamic range of $1.5 \cdot 10^4$ was reported.
The pulse height was ca. 8 mV and the length amounted to about 90 ns. Before the preamplifier, the pulse shape had a rise time of 1.7 ns and a fall time of 2.7 ns.
The DCR was around 1 kHz.
With a newly developed preamplifier, a pulse height of about 10 mV and a pulse length of ca. 20 ns resulted. [680] With this amplifier, an R_σ of 10% and a T_{FWHM} of 1.1 ns were measured. The DCR ranged from 300 Hz (8 kV) to 1000 Hz (9.5 kV) for 1 pe threshold at 30.6 °C.
The R12112 is designed for 0.9 MPa pressure.

The DCR value of 1 kHz is slightly higher than the necessary ≈ 680 Hz and was measured for the old, SQE cathode. It will be higher in the newer HQE version, but still might comply with the LSND requirement after long times in the dark. This should be verified. Apart from this, the known properties are quite favorable and place the R12112 in the shortlist.

5" prototype [217]

The first proof-of-principle HAPD design for Hyper-K had 5" (128 mm) diameter (effective cathode diameter 80 mm), used fountain focusing and a 3 mm diameter APD, and already showed good properties.

[216] Content based on [165, 228, 676, 680, 681].
[217] Content based on [682–684].

The newest prototype HY0010 had an APD with a thicker depletion region to lower the capacitance from 130 to 30 pF and thus obtain a faster pulse shape (5 ns width, 18 mV height). However, this required to use a higher bias voltage of 350 V (gain ca. 53) due to the lower field strength.
A negative bombardment voltage of −8.5 kV (3.5 kV threshold voltage) resulted in a total gain of $4.4 \cdot 10^4$, a S/N of about 21 and allowed to distinguish up to 6 pe, which indicates a very good charge resolution.
The σ_T was about 0.4 ns and a DCR of 2.56 kHz was measured.

Many properties are missing, which makes it difficult to assess the suitability of the 5" prototype for LSNDs.
The DCR (19.9 Hz/cm² cross section) is too high, but still within the range, which could be lowered sufficiently through extended periods of operation. The negative HV could be problematic in a liquid, but the polarity could be altered.
Since no strict exclusion criterion is met, the sensor is placed in the shortlist until it can be fully characterized.

Fast HHPDs[218]

Hamamatsu offers also several very small fountain-focused HAPDs with extremely low TTS, 33 mm diameter and 1 mm APDs.
The R10467U-06 has a bialkali cathode with 6 mm effective diameter and 28% QE at 350 nm, which is placed on a plano-concave quartz window. This sensor achieves a T_{FWHM} of only 50 ps for full illumination, 36 ps for 5 mm spot size, and 34 ps for 1 mm – including 30 ps setup jitter. Such a low TTS is competitive with MCP-PMTs.
The R10467U-40 and R11322U-40 have GaAsP cathodes with 3 and 5 mm diameter on a flat borosilicate window and reach a QE of 45% at 500 nm. Here, a T_{FWHM} of 170 and 90 ps is reported.
All three sensors have the following in common: An EB voltage of −8.5 kV, a bias voltage between 300 and 500 V, a total gain of $1.2 \cdot 10^5$, a rise time of 400 ps, a fall time of 400 ps, and a pulse width of 600 ps. Up to 6 pe can be discerned in the SER.

Despite their excellent timing, these sensors are far too small and expensive (particularly GaAsP, which also has a limited lifetime) for LSNDs.

MAGIC-II HHPD[219]

A small hexagonal fountain-focused HAPD (Hamamatsu R9792U-40) with 28 mm size, a 18 mm diameter GaAsP cathode, and a 3 mm APD was developed by Hamamatsu and the MPP for the MAGIC-II[220] IACT.

This sensor reached a QE > 50% at 500 nm and > 30% between 350 and 700 nm. A relative degradation of the QE by 20% was observed after 10^4 h of operation at 300 MHz photon rate, which corresponds to 10 a IACT operation. The QE is comparatively low between 300 and 400 nm, but this can be improved with a WLS coating (POPOP, butyl-PBD and Paraloid B72) of the window.
For −8 kV EB voltage (gain about 1550) and 400 V APD bias voltage (gain ca. 50), up to 5 pe peaks can be resolved in the SER and the CE is close to 100%.

[218] Content based on [181, 673] [169; p. 232].
[219] Content based on [217, 545].
[220] Major Atmospheric Gamma Imaging Cherenkov Telescopes

To compensate the large temperature variation of the gain, a thermistor was added to control the APD bias voltage.
The pulse width was 2.1 ns.
Due to the very high vacuum and an ion deflector close to the APD, a low AP rate was observed.

The small size and high costs exclude this sensor for LSNDs despite its very good QE. The negative HV might also be problematic.

Imaging HHPDs[221]

Apart from the 2-ch and 5-ch versions of the R12850, Hamamatsu has also developed a small imaging HAPD with an APD array (16×16mm, 8×8 pixels, 6 pF per pixel) for particle physics.

This sensor has a 25 mm diameter multialkali cathode with 20% QE at 410 nm. It is proximity focused (2.5 mm gap) with −8 kV EB voltage. This resulted in a bombardment gain of 1270 with 4 kV threshold voltage. Combined with a bias voltage of 340 V, this produced a total gain of $5 \cdot 10^4$ with ± 12.5% gain uniformity.
In the SER, the R_F was 28% and up to 4 pe were resolvable.
The σ_T was 80 ps, including an electronics jitter of 26 ps and a light pulse jitter of 29 ps.
Regarding the pulse shape, a rise time of 1.1 ns and a fall time of 1.3 ns were reported.
The sensor showed a crosstalk of 6% and could be operated in up to 1.5 T.

Due to the low QE and small size, this sensor is not further considered despite its otherwise very good properties. The negative operating voltage could also pose a problem in liquids.

Hybrid position sensitive avalanche photodiode (HPSAPD)[222]

The hybrid position sensitive avalanche photodiode (HPSAPD)[223] was developed together by Radiation Monitoring Devices (provided the PSAPD), ITT Night Vision (cathode), and Science Wares (system integration) in 2003 (figs. 4.14f and 4.14g).
In this proximity focused HAPD, a PSAPD (see p. 626) is installed instead of an APD, which enables a continuous 2D position reconstruction with the centroid algorithm shown in eq. (4.4).

In the HPSAPD, a vacuum tube with a GaAs photocathode was used. In an improved version, a peak QE of 29% was reached at 550 nm and the QE was about 23.5% at 420 nm (before: peak QE 14.6% at 650 nm and ca. 0.5% QE at 400 nm).
The cathode was placed a few mm in front of a 14×14 mm² PSAPD with four readout contacts on the back resistive layer. A negative EB voltage of 1–10 kV was used, the top of the PSAPD was at ground potential, and a positive voltage of ≈ 1750 V was applied to the bottom contacts for the reverse bias. This resulted in an EB gain of $(1–2) \cdot 10^3$, an avalanche gain of ca. 10^3 and a total gain of about 10^6. Due to the high gain and the relatively low noise, photon counting was possible at room

[221] Content based on [685].
[222] Content based on [249, 569, 663].
[223] Originally called imaging HAPD (IHAPD) [249].

temperature. Usually the top surface signal is used for charge and timing and the bottom contacts for position information, but the bottom contacts could also be used to measure the charge.
A spatial resolution of 320 µm was reported for illuminations with 5 photons/pulse (lowest measured intensity).
The pulse shape had 50 ns rise time, 350 ns fall time and showed a second peak 50 ns after the primary pulses, which indicates the presence of LPs or APs.
Due to the proximity focusing and the large EB voltage, the sensor showed high magnetic field tolerance.

Several crucial properties are not known, the QE is relatively low, GaAs cathodes age faster, and the sensor size is too small to be practical. For these reasons, this sensor is not included in the shortlist.

Hybrid silicon photomultipliers (HSiPMs)

In a hybrid silicon photomultiplier (HSiPM), the photoelectrons from a photocathode in a vacuum tube are accelerated by a bombardment voltage of several kV onto an SiPM (see fig. 4.15) [224]. Here the incident pe creates e–h pairs, which start a Geiger-mode avalanche.
Since even for 10 kV impact energy the pe range in Si is only 1.5 µm [686], usually only a *single* cell is triggered. Thus, the output charge corresponds to that of 1 pe in the SiPM. Due to the high SiPM gain of 10^5–10^6, this signal is sufficient for photon counting.
If the impact energy is above the threshold energy, the charge response does not depend on the bombardment voltage and the number of produced e–h pairs, because the triggered SPAD is in Geiger-mode. Higher impact energies could, however, improve the SiPM TTS, since the avalanche is started by the electrons which are closest to the multiplication region. If more e–h pairs are produced, the variation of the closest electron's distance to the multiplication region is lower, which reduces the drift time jitter.
Because of the small size of the pe interaction volume, the FF of the SiPM enters as a factor in the PDE.

As the SiPM here is used as an *electron* and not a photon detector, the layout can be optimized to minimize the energy losses.
This can be done by omitting the epoxy/gel protection layer and by reducing the thickness of the oxide layer. [555] Furthermore, a p-over-n configuration is beneficial, since in this case the pe does not have to cross the high-field region before reaching the depletion region.
Because of the precisely known pe range in Si, which is shorter than the photon absorption length, the thickness of the depletion region can also be reduced. This lowers the DCR and TTS, and increases the ME.

HSiPMs were first proposed in 2008 by [686] [225]. This publication also introduced the VSiPMT (p. 691), which is the only HSiPM type that has produced fully operational prototypes so far.

[224] Content based on [558, 686].
[225] According to [558].

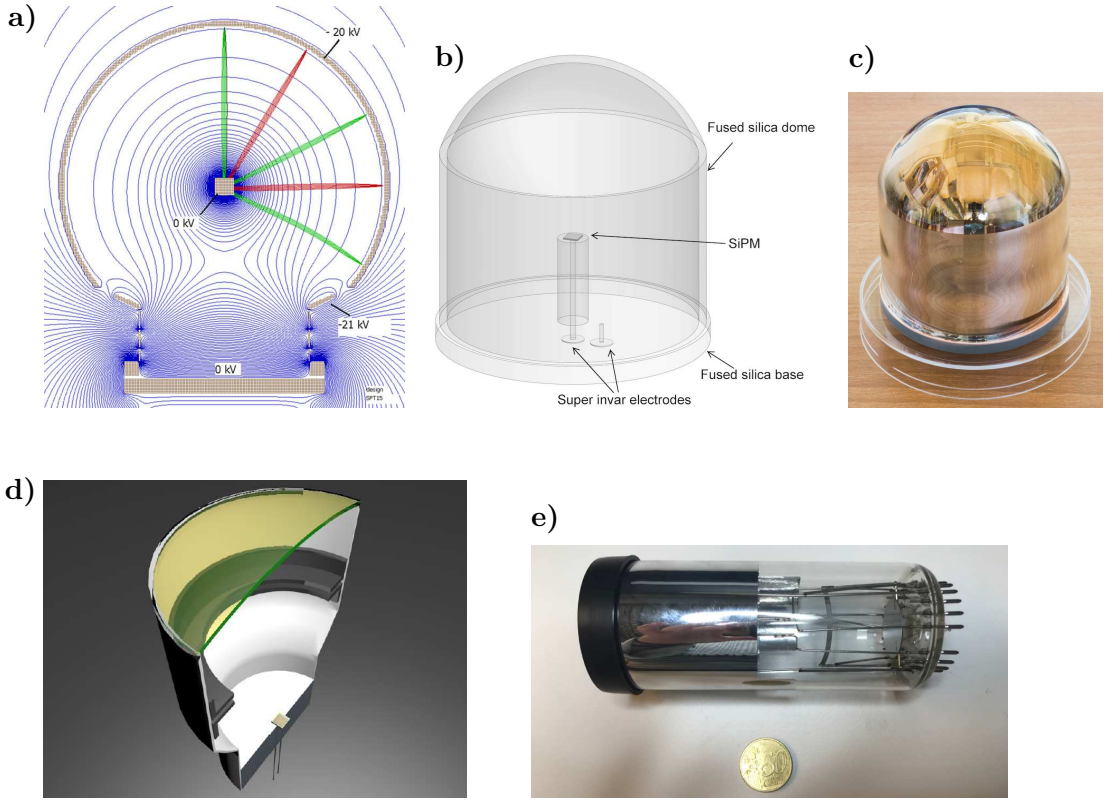


Figure 4.15: Hybrid photosensors based on electron bombardment of silicon photomultipliers (HSiPMs). Discussion see text.
a) X-HPD: SiPM variant, from [558].
b)–c) Silicon Geiger Hybrid Tube (SiGHT): b) Layout. c) Photo. Both figures from [556].
d)–e) Vacuum Silicon PhotoMultiplier Tube (VSiPMT): d) Layout. e) Photo of 2" prototype. Both figures from [687].

In general, HSiPMs have good CE, very low charge resolutions, very good TTS (the excellent TTS of the SiPM is dominated by the CTS of the cathode, however) and low AP rates. The magnetic field sensitivity depends on the electron optics layout but is lower than for PMTs due to the high voltage.
Interestingly, the large backscattering probability on Si – which significantly deteriorates the npe separation in the SER of HPDs and HAPDs – is strongly suppressed in the SER of HSiPMs. This is because the charge response is independent of the energy that is deposited in the pixel before scattering, as long as an avalanche is triggered – either the incident pe is detected or not. Therefore, the ME is quite high. However, the present author notes that this behavior should *also* lead to a large number of EAPs with regular charge, which result from inelastically scattered photoelectrons that hit other pixels; frequent pileup can be expected.
As for HPDs and HAPDs, the material compatibility of the very different components can be problematic during fabrication and afterwards when the HV is applied. Micro-discharges can lead to high DCRs, the bake-out can damage the SiPM, and the internal cathode deposition exposes all components to alkali vapors.
The biggest problem, however, arises from the DCR of the SiPM, which – despite the measures mentioned above – still is extremely high near room temperature. Since the EB does not lead to an additional gain (as in an HS-SiPM), the signal charge of 1 pe

is *identical* to that from thermal carrier generation and can not be separated through a charge cut. The resulting high DCR of the HSiPM eliminates this sensor class for many applications, including LSNDs. This requires either operation at cryogenic temperatures, where the SiPM DCR is unproblematic, or with high photon fluxes, which allow to raise the threshold and in this manner discriminate dark noise.

X-HPD (HSiPM variant)

See p. 657.

Vacuum Silicon PhotoMultiplier Tube (VSiPMT)[226]

The Vacuum Silicon PhotoMultiplier Tube (VSiPMT) is an HSiPM which is being developed since 2007 by the University of Naples[227], the INFN and Hamamatsu.
The sensor is intended for use in (astro-)particle physics, e.g. DM search, Cherenkov telescopes, proton decay, astrophysical neutrinos and space experiments.
The idea is to combine the performance of an SiPM with the large sensitive area of a PMT photocathode by replacing the dynode chain with an SiPM. The target is to provide a sensor with high linearity, gain and QE which is capable of photon counting at moderate operating voltages (ca. 2 kV). The concept was based on the QUPID (see p. 680) [671; journal version].
Initial versions had 3 mm and 1" size and still suffered from technical problems. These were overcome in the latest 2" version, of which four prototypes were constructed in 2018 that will be discussed in the following. This variant is believed to be mature enough for commercialization.

The 2" VSiPMT (figs. 4.15d and 4.15e) contains a bialkali photocathode of 46 mm diameter on a borosilicate glass window. A focusing ring accelerates and concentrates the photoelectrons onto an SiPM, which is used as a single stage amplifier operating in Geiger mode. This provides low statistical fluctuations.
A specialized SiPM is used which is called SiEM (Silicon Electron Multiplier) and has a modified structure for low energy electron detection (no epoxy resin layer, a thinner SiO_2 layer and a p-over-n junction). The EB voltage is not needed for gain but only to focus the pe onto the SiEM and provide it with enough energy to reach the active layer and trigger an avalanche. Therefore, a comparatively low voltage of about 2 kV (typically 2.2 kV) is sufficient, and unlike other sensors with EB the voltage stability is not critical.

A QE of up to 35% was measured. Assuming a CE and ME of 100%, a PDE of as high as 27% was predicted (lower than the QE due to the FF of the SiPM). This is in line with the measured PDE of $30 \pm 5\%$.
The SiEM has a size of $3 \times 3\,mm^2$, a pixel size of 25–50 µm (FF 47–74%), and a gain of $(3\text{–}8) \cdot 10^6$ for bias voltages of 52–55 V. This is equivalent to the total gain of the sensor.
A very low charge resolution was found, which allowed to clearly distinguish up to 15 pe in the SER.

[226] Content based on [505, 555, 634, 687].
[227] Including the inventor G. Barbarino.

The linear range was over 800 pe and is determined by the number of pixels; gain and linear range are *independent* in this sensor.
The TTS was 2.47 ns at 2.2 kV EB voltage and 55 V bias and was dominated by the CTS. In the 1" prototypes a plano-concave cathode was used, for the 2" design the cathode shape was not mentioned. The transit time amounts to 41.5 ns, which is surprisingly high for a 2" detector with high EB voltage and without dynode chain. In the TTD only the main peak was visible with a very small trail of LPs with up to 3 ns lateness.
The pulse shape had a length of about 200 ns and a height of ca. 10 mV.
The SiEM DCR was above $9 \cdot 10^5$ at 0.5 pe threshold ($> 2.5 \cdot 10^4$ for 1.5 pe, > 20 Hz for 2.5 pe) and is identical to the DCR of the complete sensor.
A far lower temperature dependence than for bare SiPMs was found, and the device had a low power consumption.

Most known properties are very good, although the TTS is still quite high for such a small device and the CTS likely admits of improvement. However, the enormous DCR for spe detection (0.5 pe threshold) at room temperature firmly excludes this sensor for LSNDs.

Silicon Geiger Hybrid Tube (SiGHT)[228]

The Silicon Geiger Hybrid Tube (SiGHT) is an ultra-low background HSiPM which is designed for operation in LAr/LXe (figs. 4.15b and 4.15c).
It was developed together by several institutes [229] for future large DM and neutrinoless double beta decay experiments [230].
The SiGHT is an advancement of the QUPID (HAPD class), whose development was terminated by Hamamatsu in 2012, and has a very similar design with an SiPM in place of an APD. This allows lower bombardment and bias voltages, provides a good charge resolution, and retains the favorable mechanical characteristics in LAr/LXe.
As of 2017, a feasibility study had been carried out with positive results, and a first 3" prototype with a CsI cathode was being developed at UCLA. At a later time, a bialkali cathode and tests of the electron detection efficiency in SiPMs were planned. Since then no publications or talks have been presented and it is unclear whether the development is still ongoing.

The SiGHT has a diameter of 3" and consists of a dome (a cylinder capped with a spherical section) and a base which are both made of ultra-pure fused silica. A $3 \times 3\,\text{mm}^2$ SiPM is placed on top of a quartz pillar on the baseplate.
The SiPM is used as an electron counter and should have a SiO_2 passivation layer with minimum thickness and no protection layer. Back-illumination is beneficial, since then the electrons do not have to pass the metal wiring. Furthermore, the SiPM should have a low DCR, high FF and high gain.
A thin-film (several µm) indium seal is placed between dome and base. For a better adherence on quartz, additional layers are applied before the seal (first 100 nm Cr,

[228] Content based on [556, 670].
[229] UCLA, Institute of High Energy Physics (IHEP), INFN Napoli, Laboratori Nazionali del Gran Sasso (LNGS), Princeton University, UC Davis, and University of Houston.
[230] So far, in these experiments cryogenic PMTs are used, whose achievable radioactive background is limited by the material composition of their dynode chains. For large experiments with many photosensors, the sensors can become the main background source. Thus, replacing the dynode chain with a Si sensor promises significantly lower backgrounds.

then 200 nm Cu). To minimize the required seal material [231], the surfaces are polished or lapped for < 5 µm surface roughness.
A negative bombardment voltage of about -4 kV is applied to the cathode via the vacuum seal and a metal layer that coats the inner cylinder surface and part of the baseplate. The ground potential (on the anode pillar) and the SiPM bias are provided through two super invar electrodes on the bottom of the base (similar seal structure as for the dome).
The photoelectrons hit a single pixel of the SiPM, and the signal is read out through one of the electrodes.

For the final sensor, a low temperature bialkali cathode is foreseen. Here, a high QE of $> 30\%$ at the wavelengths of LXe (178 nm) and LAr scintillation light [232] is targeted. Furthermore, a low DCR at cryogenic temperatures and a good linearity are desired. To provide a better conductivity, a transparent conductive underlayer is foreseen. For the prototypes, however, CsI or Au cathodes were used, which are easier to fabricate.
The simulated CE was nearly 100%, but the FF of the SiPM also enters the PDE, which reduces its value.
For the first tests, a SensL C30035 SiPM (35 µm pixels, 64% FF, 24.7 V breakdown voltage) was used. This could reach a gain of above 10^6 between LAr and room temperature (the breakdown voltage falls with dropping temperature).
At least up to 50 pe could be counted in the SER at 87 K, which was possibly limited by the electronics.
The DCR amounted to ca. 1 MHz at room temperature, ≈ 2 Hz at 165 K (LXe), ≈ 0.5Hz at 87 K (LAr), and showed an exponential drop down to ca. 150 K where the DCR leveled out.
A total radioactivity of < 0.04 mBq was estimated for a 3” SiGHT.

No functional prototypes have been produced so far and it is unclear whether development continues. In any case, the huge DCR unmistakably prohibits a use in LSNDs.

4.3.2 Gas-based

Gas-filled phototube

A phototube (see p. 636) can also be filled with a gas with low pressure (often Ne or Ar) to achieve amplification through a Townsend avalanche. [233]
In this sensor, the gain rises with the applied voltage between cathode and anode. However, usually the voltage is kept low enough to avoid the onset of a glow discharge, which could damage the electrodes and would lead to an undefined response. Typically, a voltage of 50–100 V is used, which results in a gain of above 5. The voltage dependence of the gain can introduce measurement uncertainties.
The transit time of gas-filled phototubes is longer than in vacuum phototubes.

[231] Indium has a comparatively high radioactive load.
[232] The emission lies at 128 nm, where quartz is not transparent, so the wavelength is shifted to 420 nm with tetraphenyl butadiene (TPB).
[233] Paragraph based on [583].

Avalanche UV detectors can be considered a subtype of gas-filled phototubes where the glow discharge is not avoided but *used* (discharge tube). [234]
These sensors are used to detect flames, since most flame types emit light between 180 and 300 nm, where sunlight is largely absorbed by the atmosphere (below 250 nm) or glass windows and where electric lamps have nearly no emission. A photosensor with a narrow spectral response in this region thus allows to detect flames with a high S/N and a higher sensitivity than smoke detectors, especially outdoors. Use cases include fire and arson alarms, flame detectors, burner combustion monitors, and the detection of discharges or UV light leakage. Commercial products exist from Spectrex, ESP safety, and Hamamatsu (UVtron).
In the UVtron, for example, a UV-sensitive reflective metal photocathode is used in a quartz phototube that is filled with low pressure Ar. A voltage of 350 V is applied between the cathode and an anode which has the shape of a bent wire. Incident UV photons then release photoelectrons from the cathode, which trigger a glow discharge (Geiger mode). The latter is quenched by an external resistor and produces a voltage pulse of 10 V height, which is independent of the incident flux (binary response). For lower voltages, the sensor operates like a proportional-mode gas-filled phototube.
The borders between a gas-filled phototube with a wire anode and an SWC SPGD are blurred, and the described layout could be considered either of both types.
The UVtron is sensitive between 185 and 260 nm and can detect fluxes down to 1 pW (corresponding to $0.9–1.3 \cdot 10^6$ photons/s, depending on wavelength). The duration of the glow discharge is about 1 ms, which limits the maximum counting rate. Since the sensor produces UV radiation, it can cause crosstalk in neighboring sensors.
Studies have shown that a CsI or CsTe cathode can improve the sensitivity of avalanche UV detectors by almost a factor of 1000, which allows operation in the proportional mode. [196; pp. 454, 456] This enables to discriminate cosmic rays and to recognize sparks.

None of the abovementioned sensors is suited for LSNDs. The low gain excludes gas-filled phototubes, whereas avalanche UV detectors are insensitive near 420 nm, cannot detect single photons, and have a slow response with long dead times.

Hybrid-gas photomultiplier [235]

The hybrid-gas photomultiplier is a sensor concept which was developed at the Technische Universität München (TUM) from 2009–2010 for the LENA project (see 1.3.1). The target was the detection of single photons at 430 nm with ns TTS, large sensitive area and low costs.
This device attempts to solve the VPGD challenge – to create a photosensor with a bialkali photocathode and gas multiplication (see p. 701) – by placing the photocathode *in a vacuum* which is separated from the gas through a thin membrane. The photoelectrons are accelerated by a high voltage onto and through the foil and produce several 100 ion pairs in the gas, which are then amplified with a TGEM.
Despite the name of the sensor, the combination of a phototube and a GD is not an HPS in the classification used here, since the GD is no photosensor.

[234] Paragraph based on [489; pp. 547–548] [196; pp. 454–458] [688, 689].
[235] Content based on [690].

Since two transport media are used – first vacuum, then gas – it is difficult to situate the hybrid-gas photomultiplier in the photosensor tree. For now, it is considered a gas-based EPE sensor. It could, however, also be placed in a new group of vacuum phototubes that use electron bombardment of GDs. In addition, the foil likely has a TSEY and thus will act like a single tynode. In this case, it would be most correct to locate the sensor in a GD subgroup of transmission dynodes.

The hybrid-gas photomultiplier consists of a cylindrical ceramic [236] vacuum tube (23 mm height) with a flat bialkali cathode (85 mm active diameter). The cathode is deposited on the window in 10^{-10}–10^{-11} mbar and the baked out tube is indium-sealed onto the window in ultra-high vacuum. For the first prototype, a solar blind CsI cathode was foreseen due to the far simpler production method.
In the center of the baseplate, a plane membrane with 2–10 mm diameter is placed, which bulges upward by 2 mm from the pressure difference. The membrane thus must have high tensile strength but also low thickness to minimize the energy loss of the electrons. While a larger diameter of the foil increases the CE, a small thickness limits the reachable diameter. For the membrane, metalized Mylar and silicon nitride (Si_3N_4) were considered as materials. Mylar with 2 µm thickness and 1 cm diameter withstood a pressure difference of 1 bar but outgasses H_2O and it not bakeable. Therefore, this material is only suited for CsI prototypes. On the other hand, Si_3N_4 membranes are etched, can achieve a minimum thickness of 30 nm, show no outgassing and are bakeable.
The baseplate and side walls are sputtered with a focusing electrode (likely at cathode potential).
The bottom part of the sensor consists of a cylindric gas tube with an Ar/CH_4 (95/5) mixture at 1 bar and contains 2 stacked TGEMs. To minimize the probability of discharges to the housing, only the inner part of the PCB has a TGEM pattern. Compared to GEMs, TGEMs provide a higher gain and better CE, and are robust, spark resistant and cheaper.

Simulations of the pe transmission for 2 µm Mylar and 10–12 keV EB voltage showed large losses and very low fractions of transmitted electrons. However, for 300 nm Si_3N_4 and 8–12 keV most electrons were transmitted with losses between 1 and 2 keV. For 12 kV and Si_3N_4 the mean pe energy after the foil was 9 keV, which is sufficient to ionize over 500 gas molecules. To achieve a total gain of 10^5, which allows photon counting, thus only an avalanche gain of 200 is needed, for which a single TGEM is sufficient. In principle, the foil could also act as a transmission dynode due to the production of SEs, which might potentially yield an additional amplification step.
In field simulations for 12 kV acceleration voltage and 1 eV emission energy, all photoelectrons hit a 1 cm diameter Mylar foil, although with a very large CTS of 50–100 ns. For a Si_3N_4 membrane with 2 mm diameter, on the other hand, only electrons with starting positions up to 26 mm radius ended on the membrane, which reduces the usable diameter.

No fully functional prototype was produced and the feasibility of this sensor concept remains to be shown.

[236] Ceramic does not outgas, is bakeable and is insulating.

Compared to VPGDs, IBF to the photocathode is *not* a concern in the hybrid-gas photomultiplier and the full gas gain of the HTGDs can be used. In addition, the QE of the photocathode is not reduced by photoelectron back diffusion as in VPGDs [237], and thus high PDE values could be reached for an optimized CE.
The first stage gain is high, which should result in a very good charge resolution.

One of the main problems of the current design is the fountain focusing from a flat cathode onto a small central membrane with a cylindrical focusing electrode. This results in a low CE and a very large CTS.
In addition, since the photoelectrons are fountain-focused onto the membrane without an unambiguous mapping, the photon position information is mostly lost in this step and the position resolution of the TGEMs can not be used. The present author therefore would like to suggest to use an *array* of foils instead of only one. This would restore the 2D position capability (at the cost of needing additional readout channels), increase the CE, and significantly lower the CTS. The cathode–foil distance could also be optimized for these properties, as far as sparking allows. For a single membrane, alternatively a curved cathode and optimized electrode shape would also improve the CE and TTS and might allow to reach a limited position resolution.

In sum, this concept of a different VPGD design seams feasible and with some improvements it should be possible to obtain good sensor properties.
Since no functional prototypes exist and this sensor is currently not being developed, it is, however, not included in the shortlist.

Solid photocathode gaseous detectors (SPGDs) [238]

Overview of photosensitive gaseous detector types

Photosensitive gaseous detectors (PGDs) are photosensors which contain gas as transport and amplification medium.
Three classes can be distinguished based on the state of matter of the photocathode: Gaseous [239] for GPGDs, liquid [240] for LPGDs and solid for SPGDs.
For SPGDs two important subclasses exist, namely solid photocathodes which are VUV and UV-sensitive only and those whose sensitivity extends into the visible range. This difference is of crucial importance for many applications and the requirements to construct SPGDs that are sensitive to visible light are vastly different. For this reason, the latter are treated as a different sensor type: visible light PGDs (VPGDs). Their UV-sensitive counterparts are referred to as UV-SPGDs in the following.

Since PGDs are based on gaseous detectors, they can be made position sensitive with the right anode structure and readout method (see p. 601).

[237] Scattering of a newly emitted, low-energetic photoelectron on the gas molecules can return it to the cathode, where it is reabsorbed. This pe back diffusion has a considerable probability and significantly lowers the QE in VPGDs.
[238] Content primarily based on [196] – the definite compilation about all PGD-related developments – as well as [538] and [194; ch. 11, pp. 305–306]. The initial PGD overview mostly uses pp. xi–xviii, 498–509 and ch. 15 of [196].
[239] Gas mixture with photosensitive component.
[240] Typically a thin liquid layer that is adsorbed to a solid photocathode.

History

Of the three PGD classes, GPGDs were developed first, initially without position resolution (SWC, 1955) and later with position sensitivity (MWPCs, 1977 and 1978, independent developments).
LPGDs were studied shortly in the years after 1986 in the hope of resolving some of the limitations of GPGDs but never found widespread use.
Gas-filled phototubes (see p. 693) can be considered the first SPGDs without position resolution, whereas the earliest position sensitive SPGDs were developed starting 1980 (CuI) and were exclusively UV-sensitive. Functional CsI cathodes were first reported in 1989 and quickly found widespread use, because they eliminated many shortcomings of GPGDs. Today, SPGDs increasingly replace GPGDs in many applications, but still mostly use CsI cathodes, which are insensitive in the visible range.
In the past two decades, many efforts and great progress have been made in the development of VPGDs, which culminated in the production of first functional prototypes (see below), although many problems still remain to be solved.

Applications

PGDs are mostly used in RICH detectors for particle physics, but applications also include VUV spectroscopy, hyperspectral imaging, readout of UV scintillators (gaseous/noble liquid/solid), detection of X-rays, gamma rays or particles with solid cathodes, astroparticle physics[241], flame detection, smoke detection, and plasma diagnostics.

Advantages

PGDs have several attractive advantages. They can be constructed with very large sensitive areas[242] (up to over $10\,m^2$) and low-background materials. Their photocathode can be gaseous, liquid (or adsorbed layer), or solid (reflective or transmissive). Furthermore, PGDs achieve high gas gains of up to ca. 10^6 (proportional mode, typically $> 10^5$; some GPGDs in Geiger mode up to 10^8), which is sufficient for spe detection. In addition, they reach position resolutions down to $\approx 30\,\mu m$ RMS, can show sub-ns TTS, can be operated in magnetic fields of up to 4 T, and are relatively inexpensive.
For GPGDs, DCR also only is caused from radioactive background and can be discriminated by the pulse amplitude.
MPGD-based SPGDs show particularly good properties, such as high gain, excellent TTS and position resolution, efficient suppression of photon and ion feedback, and a high rate capability of $> 10^6\,Hz/mm^2$. Therefore, current research is focused on this class of PGD sensors.

Disadvantages

However, there are also several important disadvantages and problems.

For all PGD types, photon feedback from the avalanche to the cathode reduces the usable gain range. In addition, IBF can lead to electron emission from the cathode,

[241] TMAE-filled MWPCs were successfully tested for air shower detection via Cherenkov light. Since this gaseous cathode type is solar blind, these sensors could operate in moonlight and possibly even daylight.
[242] When operated at 1 atm pressure, there are no mechanical constraints on the window size, and thus very large sensitive areas can be realized.

which triggers a second avalanche at the same position (iAP) [196; pp. 14–15]. For small total gains, it is furthermore necessary to use low noise electronics.

For gaseous and liquid cathodes, it is difficult to find materials which have low ionization potentials[243], high vapor pressure and are not extremely chemically aggressive. Therefore, all gaseous and liquid cathodes are UV-sensitive only (TMAE and TEA are most commonly used for both types).
The cathode material usually is a liquid or solid, which for GPGDs results in low values of the vapor pressure, photon absorption and QE. To improve these quantities, it can be necessary to heat the complete detector.
In addition, impurities in the gas can absorb photoelectrons and thus reduce the QE. The high reactivity is a direct consequence of a low ionization potential, which leads to weakly bound outer electrons. The gas handling is therefore complicated and it is difficult to find detector materials which are compatible.

Gaseous cathodes also must be thick enough for efficient photon conversion, so the ionization position varies greatly (especially for low pressure and a large mean free path of the photon). This significantly worsens the time and position resolutions.

Solid cathodes, on the other hand, have a well defined conversion position. This results in a far better TTS (down to < 1 ns) and spatial resolution (30–50 µm can be reached). Furthermore, solid cathodes do not involve chemically aggressive gases, and with suitable materials (Cs_3Sb or bialkali) the spectral response can extend to the visible range.

However, for solid cathodes almost exclusively UV-sensitive types are used (mostly CsI), since they are more robust – these materials can even withstand exposure to air for several ten seconds. Here, problems arise from photon feedback from the avalanche, which limits the maximum gain, and pe back diffusion to the cathode, which reduces the QE. However, both issues can be solved with adapted fill gases (e.g. CH_4).

Solid cathodes which are sensitive to visible light have short lifetimes in GDs. This occurs due to even minute impurities of oxygen or water in the gas[244], which requires to use ultra-clean gases and sealed gas chambers with very low outgassing. In addition, the lifetime is lowered by IBF, which can not be reduced in this way.
Feedback from photons and ions[245] furthermore limits the gain that is achievable before breakdown.
A thin coating layer of e.g. CsI can protect the cathode from impurities and ions, but this significantly reduces the QE, since the photoelectrons lose energy on their way to the surface.
Adapted GD structures, which limit photon and ion back-flow have proven to be a better solution. Recent cascaded HTGD types have shown particularly good suppression of both feedback types, which makes high gains possible even in poorly quenched gases.

[243] The upper limit of the sensitive spectrum is inversely proportional to the ionization potential. Liquid cathodes have lower ionization potentials than in their gas phase, which is one of the reasons why they were studied.
[244] The photocathode acts as getter and gets poisoned increasingly, especially if the gas is flushed and new impurities are introduced into the sensor continuously.
[245] Photocathodes with low photoemission thresholds also have large ISEYs.

Since only SPGDs can be sensitive in the visible range – with the abovementioned problems – PGDs are currently practically limited to the UV range.

Operation

Most PGDs are operated in the proportional mode, since here low operating voltages are sufficient and it is possible to discriminate spe signals from background. Geiger mode is used only in rare cases (especially GPGDs).
It was found that methane (CH_4) and ethane (C_2H_6) are better suited for spe counting than He or Ar (pure or mixed with quenchers).

UV-sensitive solid photocathode gaseous detectors (UV-SPGDs)

An UV-SPGD is a gaseous detector in which the primary electrons are not created through ionizing particles but are emitted from a transmissive or reflective solid UV photocathode placed on one surface of the chamber (see e.g. fig. 4.5c).
The photoelectrons typically first drift in an electric field before starting a Townsend avalanche in a high field region. The created electrons (possibly also the ions) induce signals in the anodes (cathodes in case of ions), which usually are structured to allow the reconstruction of the 1D or 2D photon position in the readout electronics. The exact layout of the drift, amplification and readout regions depends on the GD type on which the sensor is based (see p. 587 and fig. 4.5).
Since UV-SPGDs are UV-sensitive only (as opposed to VPGDs), the whole sensor class is irrelevant for LSNDs and will be treated only shortly. Note that this is different for WCDs, which *do* use UV Cherenkov light.

The following cathode materials have been studied for UV-SPGDs, mostly between 1987 and 1992: [196; chs. 5, 6]

- **CsI:** Sensitive up to $\approx 210\,\mathrm{nm}$ with a QE of 46% at 150 nm. Typically a $\approx 0.5\,\mu\mathrm{m}$ thick CsI layer is used. The QE is only slightly lower than TMAE vapors over the complete spectral range. CsI, however, is not chemically aggressive, can be exposed to air shortly, shows no strong aging from IBF, is stable for common gas mixtures (CH_4, CF_4) in a continuous gas flow, and can be applied to large surfaces. It is thus the most cost- and performance-effective solution for large detectors.

- **Cs_2Te:** Sensitive up to $\approx 360\,\mathrm{nm}$ with a peak QE of ca. 5% at 240 nm in $\mathrm{He} + 0.8\%\,\mathrm{CH_4} + \mathrm{EF}$ at 1 atm.

- **CuI:** Sensitive up to 230 nm. The QE is 5.8% at 140 nm but varies depending on gas composition, prior heat treatment (with bake-out up to 13% at 150 nm), previous contact with air/humidity, and the extraction field strength. Adding TMAE to the gas increases the QE by a factor of over 2–3. CuI was the first solid photocathode material which was studied (1980).

- **TMPD** (tetramethyl-p-phenylenediamine): Sensitive up to $\approx 235\,\mathrm{nm}$. The maximum QE is about 0.6% at 200 nm and room temperature and ca. 4% at 200 nm, $-40\,^\circ\mathrm{C}$ and flushed with methane and neopentane (NP). This material is applied as a ca. 1 µm thick layer on a Cu substrate.

- **TMAE + NP**: Sensitive up to 260 nm with a QE of up to ca. 4% at 232 nm and −20 °C with CH_4. This is a solid mixture of TMAE and NP and must be cooled.
- **Diethylferrocene:** Sensitive up to 260 nm with a QE of up to ca. 0.2% at 200 nm and −40 °C with CH_4. This material is applied on a Cu substrate and must be cooled.
- **CPIHMB** (cyclopentadienyl-iron-hexamethyl-benzene): Sensitive up to 295 nm with a QE of up to ca. 0.1% at 200 nm. This substance is applied on a Cu substrate.
- **Diferrocenyl mercury:** Sensitive up to 210 nm with a QE of up to ca. 0.5% at 185 nm, with EF sensitive up to 220 nm with a QE of 1.3% at 185 nm. This material is applied on a Cu substrate and is nonreactive with air, water or polar solvents.
- **Diferrotriphenylphosphinegold ferrocene:** Sensitive up to 225 nm with a QE of up to ca. 0.35% at 190 nm.
- **d,d-dimethyl-trimethylene-ferrocenephane:** Sensitive up to 230 nm with a peak QE of ca. 1% at 200 nm.

Metalorganic compounds, such as the last four entries, combine a small ionization potential with low chemical reactivity, which prompted their study. For these materials, the QE can be increased by a factor of 2–3 through short contact with TMAE or EF vapors, even if the sensor is subsequently operated with a different gas for prolonged times.

So far, the following solid cathodes and GD types have been combined in UV-SPGDs: [196]

- **CsI:** SWC, MWPC, PPAC, RPC, MSGC, CP, GEM (also [538]), TGEM, RETGEM, S-RETGEM, MHSP, COBRA (flipped COBRA + 2 GEMs [529]), THCOBRA [531], Micromegas
- **Cs_2Te:** CP + mesh PPAC (stacked)
- **CuI:** MWPC
- **TMPD:** MWPC
- **CPIHMB:** SWC
- **d,d-dimethyl-trimethylene-ferrocenephane:** PPAC (with adsorbed EF layer)

The properties of UV-SPGDs strongly depend on the GD type and the specific sensor. The achievable characteristics of the different PGD types were already listed in the GD overview starting on p. 587. For many types spe detection with high gain, very good spatial resolution and comparatively low TTS is possible.
Due to the multitude of realized UV-SPGD types, it would go beyond the scope of this work to list the individual properties for all of them; aside from the fact that this sensor class is not sensitive at the wavelengths of liquid scintillators with wavelength shifters and thus is irrelevant here. Please refer to the extensive review in [196] for detailed properties.

Visible light photosensitive gaseous detectors (VPGDs) [246]

Visible light PGDs (VPGDs) are SPGDs which employ semitransparent or reflective solid photocathodes that are sensitive to the visible spectrum (figs. 4.16a and 4.16b). Possible cathode materials include bialkali (K_2CsSb), monoalkali (Cs_3Sb), multialkali (Na_2KSb:Cs), GaAs:Cs or AgO:Cs (compare pp. 54–55).
Recent VPGD prototypes focus on cascaded HTGDs or micromeshes in a sealed housing.
So far, only scientific and commercial prototypes have been produced, but these sensors promise the same advantages as UV-SPGDs for visible light: extremely large sensitive areas, spe detection, very good position resolution, relatively small TTS, high rate capability (MHz/mm²), very low sensitivity to magnetic fields, low costs, and a compact layout [692].

Problems

While SPGDs with UV-sensitive solid photocathodes can achieve high gains and long lifetimes [247], unfortunately this is completely different for the visible range, and the following problems occur:

- **Gas impurities:** Cs-based cathodes are highly reactive and extremely sensitive to even tiny amounts of oxygen, water and other impurities, which poison them and reduce the QE. This requires to use high purity gases and special detector materials, and the sensors must be sealed to avoid introducing new impurities.

- **Gain:** Furthermore, APs from ion and photon feedback limit the achievable gain before breakdown occurs through discharges. The latter is the case, when the product of gas gain, the fraction of ions/photons that reach the cathode, and the probability of electron emission by ion or photon hits exceeds 1.
  For CsI, the SE emission probabilities are $< 10^{-6}$. Therefore, even for GDs with open geometries (e.g. MWPC, PPAC), where nearly all ions can drift to the cathode, gas gains of 10^5 can be reached, which allows spe detection.
  For visible light cathodes, however, the emission probabilities are in the order of 1% due to the low work function, which limits the maximum gain to ca. 100 if all ions reach the cathode.

- **Lifetime:** In addition, IBF rapidly damages the cathode and reduces its lifetime. The QE degrades after an integrated ion charge of a few µC/mm² [693].

- **QE:** Lastly, if gases with small quencher concentrations are used (to allow lower operating voltages), the QE is notably reduced by pe back diffusion compared to the value in vacuum.

Coating the photocathode with a thin protective layer can prevent or slow the diffusion of gas, while being partially transparent to photoelectrons. This allows the use of gases with ordinary quality, operation in the presence of micro-leaks, and even exposition to air or water vapors for 5–10 minutes.

[246] Content based on [196; ch. 7, pp. xi–xvi, 498–509] [523, 538, 691] [194; pp. 305–306].
[247] CsI suffers some damage from ion impacts, but the SE emission probability (ISEY) is low and does not limit the achievable gain [523].

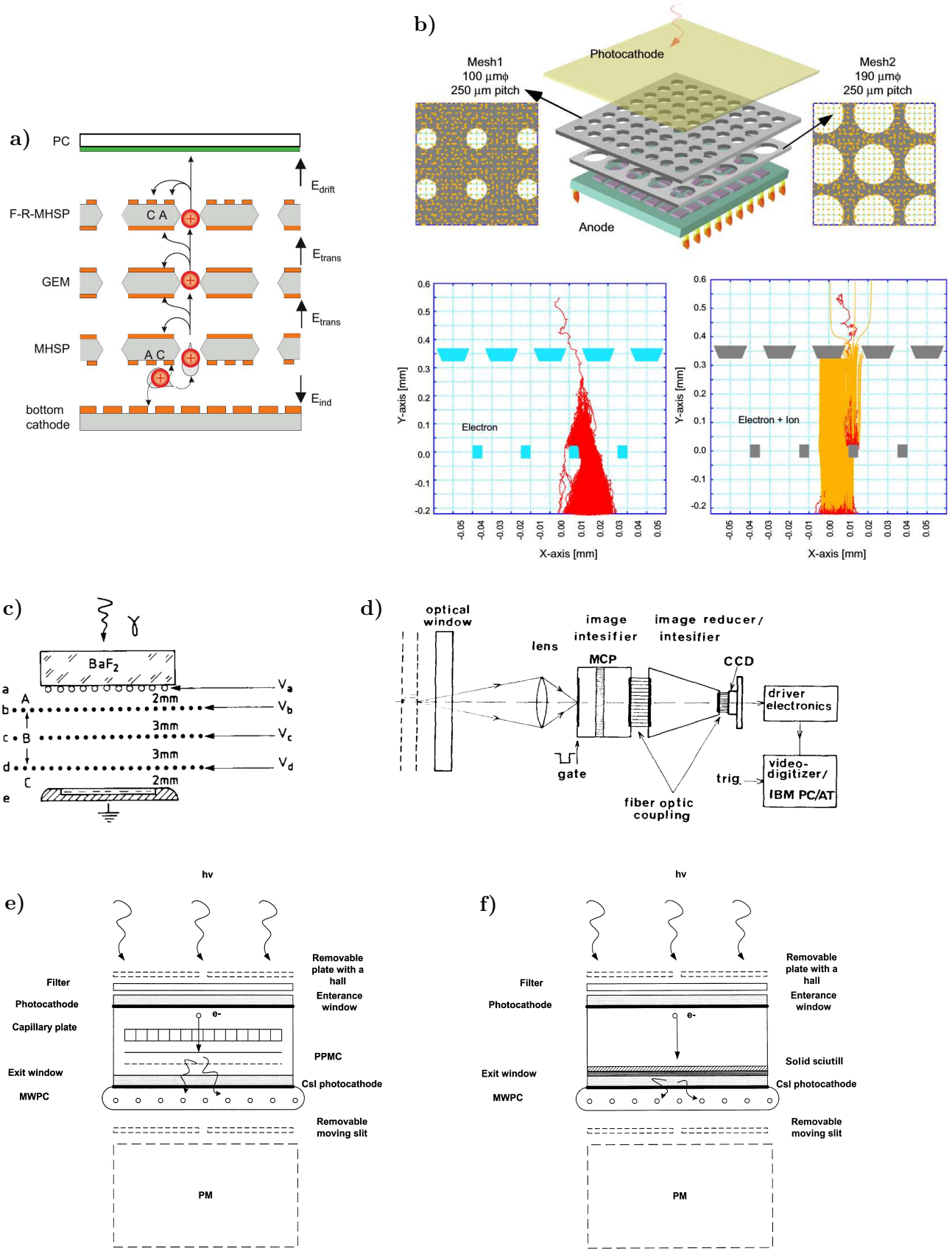


Figure 4.16: Gas-based photosensors. Discussed in text.
a)–b) Examples of visible light photosensitive gaseous detectors (VPGDs), from [523] and [694].
c) Liquid photocathode gaseous detector (LPGD), from [695]. A BaF_2 scintillator that is illuminated by gamma rays emits UV light. This is detected by an MWPC (b + d cathode wires, c anode wires) with a 1–3 mm thick liquid photocathode made of TMAE (tetrakis (dimethylamino) ethylene). The photocathode is located inside a tub formed by an Al cathode.
d) Optical readout GPGD (gaseous photocathode gaseous detector), from [696].
e) Dual gaseous photomultiplier (DGPM), from [636].
f) Hybrid gaseous photomultiplier (HGPM), from [636].

CsI was found to provide good protection against oxygen if the coating is sufficiently thick but severely reduced the QE above 250 nm. NaI also provided high protection from oxygen, while a layer of 30 nm CsBr prevented ion and photon feedback.
To improve the QE, it was tried to use removable films, which only coat the cathode during exposition to unclean gases or air and can be removed for operation. However, the heating during the removal procedure markedly reduced the QE as well.

Ion back-flow is particularly problematic, and for early prototypes high-gain operation was only possible in gated mode. To reduce IBF, it is necessary to use closed geometries such as HTGDs, which can suppress photon and (depending on layout) ion feedback.
For cascaded GEMs, the IBF was found to scale linearly with the drift field. [523] This suggests to minimize the field strength, but the necessity for efficient pe extraction from the cathode requires a minimum value.
Recently, it was also suggested to use a suspended graphene monolayer to block IBF while possibly showing a large transparency for electrons with a few eV kinetic energy. However, it remains to be shown whether this is effective and can be realized for large areas.

So far, it has proven highly difficult to obtain a high QE, large CE, sufficient gain for spe detection, and a long lifetime at the same time, which is substantiated by the fact that this has been attempted for over three decades now.

Prototypes

In the following, the most relevant VPGD prototypes which have been produced so far – all based on HTGDs and micromeshes – shall be shortly summarized.

The earliest HTGD-based types used one or two CPs with transmissive cathodes (Cs_3Sb, GaAs:Cs or AgO:Cs) in $Ar + 5\% CH_4$. [697] The maximum obtained PDE was 0.7% for Cs_3Sb at 250 nm (sensitive up to 550 nm). This value is quite low due to pe back diffusion and the FF of the CPs; the CE was at most 30–40%. For a single CP, a gain of over 10^3 was reached before feedback appeared [248]. For two cascaded CPs, a gain of almost $5 \cdot 10^4$ was achieved. This was possible, since the avalanche photons were effectively shielded and most ions were collected on the CP cathodes. Several CP-based VPGD prototypes were constructed, including commercial ones by Hamamatsu.
In [697] also a combination of a (removable) CP with inclined capillaries and a mesh PPAC (20 mm diameter, 0.6 mm gap, 80% transparent) was reported. The CP blocked all light that was produced in the PPAC and had a CE of up to 43%. A gain of 10^3 was reached.

Another prototype used two cascaded GEMs and a bialkali cathode in $Ar + 5\%$ CH_4. [698] With this, a peak QE of 13% at 405 nm (25% in vacuum; the reduction is mainly due to pe back diffusion) and a gain close to $2 \cdot 10^4$ was reached, which indicates a strongly reduced IBF. This gain was sufficient for spe detection, although at a reduced threshold efficiency; a TE of 100% would require a gain of $(1–2) \cdot 10^5$ with common electronics.

[248] The onset of feedback is visible as a deviation from an exponential gain–voltage dependence.

To improve the ion and photon feedback suppression, a cascade of F-R-MHSP, GEM and MHSP (fig. 4.16a) was combined with a transmissive cathode (Cs_3Sb, K_2CsSb or Na_2KSb) in $Ar + 5\% CH_4$. [699, 700] This reduced the IBF to $3 \cdot 10^{-4}$ and a gain of 10^5 could be reached; with a double GEM, feedback was already visible above a gain of 100. An ISEY of $3 \cdot 10^{-2}$ was measured for all three cathode materials, from which it could be concluded that an IBF below $3.3 \cdot 10^{-4}$ would be sufficient for a gain of 10^5 [523]. The CE of the F-R-MHSP at the top was about 100%.
In a cathode production study [700], for bialkali cathodes QE values above 30% were reached between 360 and 400 nm in Ar/CH_4 (95/5) for some samples. In vacuum, the mean peak QE was 28% and was reached around 400 nm, but for some cathodes the peak value came close to 50% QE at 360–380 nm. Cs_3Sb cathodes reached 10–40% peak QE around 400 nm in vacuum, while Na_2KSb cathodes had 15–25% peak QE around 370 nm in vacuum. The bialkali cathodes which were used for the VPGD prototypes described above had lower peak QEs of ca. 12–24% (partly estimated) at 370–400 nm in gas and 27–43% in vacuum.
The QE reduction in gas was assumed to be caused solely by pe back diffusion, from which the pe extraction efficiency was calculated to $\approx 43\%$ at 313 nm and 79% at 546 nm.
For bialkali, the QE decreased by ca. 20% after an accumulated ion charge of 1–2 $\mu C/mm^2$ on the cathode. This would correspond to 46 a operation at 5 kHz/mm^2 photon rate when assuming 10^5 gain and $3 \cdot 10^{-4}$ IBF.

The COBRA structure was developed to even further improve the ion deflection compared to the F-R-MHSP through a large fraction of area covered by cathodes. [523] With a flipped COBRA and two GEMs, the IBF was reduced to $3 \cdot 10^{-6}$ at 10^5 gain (the IBF rose with gain to a value of 4% at $4 \cdot 10^5$ gain). In the gain–voltage curve no feedback was visible up to $3 \cdot 10^5$ gain.
However, the CE was only about 20%. This might be improvable by an optimized geometry.

On the other hand, [701, 702] found that alkali cathodes were incompatible with organic materials such as the polyimide (Kapton) substrate of GEMs, MHSPs and COBRAs. Polyimide absorbs alkali vapors during the photocathode activation, which results in a strongly reduced QE, which is only 0.03 times as high as when using glass MB-GPs (see below). A liquid-crystal polymer was tested as alternative substrate material, but almost the same QE decrease was observed. The authors concluded that even if a transfer method is used (as was the case for [700]), the alkali metals on the cathode will be absorbed eventually by organic GD substrates. To achieve long-term stability, thus other materials have to be employed.

Based on these findings, in [701, 702] [249] instead a transmissive bialkali cathode was combined with two Micromegas (metal) or an MB-GP (glass) and a Micromegas. As gas, $Ar + 10\% CH_4$ / $Ne + 10\% CF_4$ was used.
For the double-Micromegas sensor, a peak QE of 14% at 350 nm in Ne, 12% at 420 nm in Ar, and 20% in vacuum was reported. After 1.6 a storage in gas, the QE dropped to about 80% of its initial value. Due to the open geometry, a gain of only up to about 10^3 could be achieved. The QE decreased by 17% at a gain of 670 in Ne after 0.4 $\mu C/mm^2$ accumulated cathode charge (5 mC anode charge) [703]. The

[249] Collaboration between Tokyo Metropolitan University, Yamagata University and Hamamatsu.

sensor could tolerate high magnetic fields, and in the most sensitive direction a gain drop of 50% was observed at 0.75 T.
For the MB-GP + Micromegas sensor, a maximum gain of merely ≈ 400 was reached (the Micromegas was not used for gain). Furthermore, the single MB-GP was not sufficient to suppress the IBF, which was already present at gains above 10.

These designs were further improved by the same group by arranging two metallic micromeshes with rounded holes of different diameters in a staggered pattern (see fig. 4.16b). [694] Here, conical etched holes were used which had 80 µm diameter in the first and 120 µm in the second mesh. Both layers had a pitch of 250 µm. This was combined with a semitransparent bialkali cathode and Ne + 10% CF_4. A gain of 10^4 without feedback and an IBF of $6 \cdot 10^{-4}$ could be reached, since the top mesh collects more ions due to the smaller hole diameters. The maximum gain was limited by breakdown at high voltages. Since no signs of feedback were visible, this likely occurred due to internal defects in the GD, so higher gains could be reached with an improved design.

Conclusion

In conclusion, recent VPGD prototypes have consistently achieved QE values of about 12% in gas. One group reported values of over 30% in gas, but not in a functional sensor. The QE is significantly reduced by pe back diffusion (at 420 nm by a factor of ca. 1.6–2). To reach the LSND requirement of a PDE above 20%, thus, extremely high vacuum QEs are required even when assuming a high CE. On the other hand, the flexible geometry of VPGDs could allow to achieve an optical coverage of the detector walls of close to 100% instead of the targeted 30%. This could by far compensate a lower QE.
The CE can be quite low for some designs due to the FF, but with optimized layouts such as the F-R-MHSP one can reach values close to 100%.
So far, high gains which are sufficient for efficient spe detection have only been reached with GEM-based GDs; CPs came close but are expensive. However, GEMs with standard materials (polyimide) are not compatible with visible light cathodes, and instead glass- or metal-based GDs which can strongly suppress ion and photon feedback are needed. Glass-based GEMs and MHSPs might be promising candidates [196; p. 267].
By now, photocathodes have been shown to be stable in gas for a few years. However, before operation in LSNDs can be considered, further studies are required to analyze the long-term stability. It is difficult to assess the lifetime requirement regarding ion bombardment of the cathode. In LSNDs the occurring photon rates are far lower than in particle physics experiments (for which PGDs are designed), and the results of cathode aging that are reported in [692] are very encouraging. However, so far no aging studies were carried out in a fully functional sensor with a gain of 10^5.

As of now, no prototype has been produced which fulfills all target properties of LSNDs. It remains to be seen whether a VPGD can be created which has a low IBF, high gain, QE and CE, as well as long lifetime at the same time. Efforts are ongoing [693], and it will be interesting to see what the future holds.
For now, VPGDs do not fulfill the gain and lifetime requirements of LSNDs (possibly also the iAP rate) and are not considered as candidate sensors.

Hybrid gas scintillation photosensitive gaseous detectors (HGS-PGDs)

Dual gaseous photomultiplier (DGPM) [250]

The dual gaseous photomultiplier (DGPM, fig. 4.16e) was proposed in the same paper as the HGPM (fig. 4.16f, p. 664) [636]. It has the same layout and basic concept (light amplification with a PGD as secondary sensor), except that the scintillation light here is produced by the gas avalanche in a VPGD. This is thus an alternative attempt to achieve a sensor with VPGD properties (e.g. large sensitive areas, very low sensitivity to magnetic fields, and low costs).

The DGPM consists of two parts, the scintillation chamber and the readout detector. The former is a 30 mm diameter disk with 4.5 mm thickness that is filled with ultra-clean (6N) Xe at 1 atm and is sealed. It contains a semitransparent Cs_3Sb or bialkali photocathode (Cs_3Sb provided the best results), a CP, and a mesh PPAC (called PPMC in the plot) with 1 mm gap and 700–800 V operating voltage. To improve the QE and stability of the cathode, a sealed chamber with only 2–4 feedthroughs and no outgassing materials was used. The photoelectrons produce weak VUV scintillation light in a low-gain avalanche in the PPAC. These photons are emitted through an exit window and are detected by a PGD, while the CP suppresses light feedback.
As readout detector, an MWPC was used. This had a CaF_2 window, a CsI cathode, and CH_4 or $Ar + CH_4$ gas at 1 atm. Here, an operating voltage of 1530 V was applied.

A peak QE of ca. 6.7% at 310 nm (5.6% at 420 nm) was measured. The PDE was lowered by the CE of the CP and reached up to 2.8% at 360 nm (2.4% at 420 nm). The QE remained stable for over 40 d, but it is not clear whether the test was performed under operating conditions. A position resolution of ca. 1 mm was reported, and a rate capability of above 10^4–10^5 Hz (mainly limited by electronics) was reached.

A variant with only a CP and no PPAC was also successfully tested but showed a narrow dynamic range. Here, the light amplification took place inside the capillaries.
In a preliminary study, the MWPC-PGD was replaced by PMTs with MgF_2 windows and CsTe or multialkali cathodes ("PM" in figure). This resulted in hybrid sensors of the hybrid gas scintillation photomultiplier (HGS-PMT) class.
The authors of [636] note that instead of the CsI-MWPC also a GPGD with TMAE, TEA or TMA (trimethylamine) could be used.

For LSNDs, the low PDE and uncertain lifetime excludes the DGPM.

4.4 Photoionization sensors

4.4.1 Gaseous photocathode

Gaseous photocathode gaseous detectors (GPGDs) [251]

A gaseous photocathode GD (GPGD) [252] is a gaseous detector (see p. 585 and fig. 4.5) whose fill gas contains a photosensitive gas or vapor, usually an organic compound.

[250] Content based on [636, 637] [196; pp. 249–250].
[251] Content based on [196; ch. 3, pp. 2, 162] [538] [194; pp. 305–306] [153; pp. 302–303].
[252] Also called photoionization detector [153; pp. 302–303].

The latter acts as a gaseous UV photocathode and thus the GPGD needs to have an UV-transparent window. Incident photons create ion pairs through photoionization and the generated photoelectrons are then amplified and read out in the GD.

Problems

The ionization potential of the photosensitive gas should be as low as possible to raise the cutoff wavelength.

However, gases with low ionization potential have several crucial disadvantages. Often, the vapor pressure is too low for efficient photon conversion at room temperature, which requires to use thick absorption regions (which is detrimental for the TTS) or to heat the detector.
Furthermore, the gases are frequently chemically aggressive and can react with detector materials and amplification gases. This leads to aging even if the other sensor components are carefully selected or protected.
In addition, some gases degrade through contact with oxygen (already for ppm concentrations), which requires to use high-purity gases and makes the gas handling more complicated.
The gases also can be poisonous or hazardous.

As a consequence, only few gases are suitable for operation in a GPGD. Despite extended efforts, no photosensitive gas with sensitivity in the visible range was found and all materials are limited to the UV spectrum.

Photosensitive gases and vapors

The following gases or vapors have been used in GPGDs so far:

- **TMAE**[253] (tetrakis (dimethylamino) ethylene, $C_{10}H_{24}N_4$): Sensitive up to 231 nm with a peak QE of up to 60% at 130 nm and around 143 nm. The QE exceeds 30% between 130 and 190 nm. Due to the low vapor pressure (melting point at −4 °C), heating to 40–60 °C is required. TMAE is chemically aggressive: It reacts rapidly with oxygen, requires special sensor materials (it reacts with e.g. vacuum grease and O-rings), causes aging and leakage currents, and penetrates into some materials (e.g. Al), which leads to slow outgassing over several months. Cleaning is required before operation and the gas handling is very complicated. Despite these significant drawbacks, its excellent QE makes TMAE the material of choice.

- **TEA** (triethylamine, $(CH_2CH_3)_3N$): Sensitive up to 165 nm, the peak QE is ca. 32% at 150 nm. The vapor pressure is sufficient for a use in the saturated mode without heating; at 20 °C already a 2 mm gap is sufficient for full photon absorption. TEA is less reactive than TMAE. Using CH_4 or C_2H_6 as amplification gas allows to reach gains that are sufficient for spe detection without photon feedback. This was the most frequently used compound before TMAE.

- **TMA** (trimethylamine, $(CH_3)_3N$): Sensitive up to 160 nm, the peak QE amounts to ca. 30% at 133 nm in He. In CH_4, the absorption leads to a cutoff

[253] Usually pronounced “tammy”.

at 135 nm, so here the peak QE is 28% at 145 nm. TMA is gaseous at room temperature and is often used as a water solution. The high vapor pressure makes it difficult to adjust the concentration and requires monitoring to avoid rising concentrations, which would lower the gas gain. The material is slightly reactive with metals like Cu or brass.

- **EF** (ethylferrocene, $C_{12}H_{14}Fe$, organometallic): Sensitive up to 204 nm, peak QE ca. 19% around 192 nm. EF is chemically inert and can be exposed to air. It has a very low vapor pressure, which requires long conversion lengths or heating the sensor to 60–80 °C. This material is less popular than the first three compounds due to the lower QE and the high operating temperatures.

- **Benzene** (C_6H_6): Sensitive up to 135 nm, peak QE 60% at 124 nm. C_6D_6 even has a slightly higher QE at all wavelengths, up to ≈63% at 124 nm. Benzene was used in the first GPGDs.

- **Toluene** (C_7H_8): Sensitive up to 146 nm, with a peak QE similar to benzene. This substance also was used in the first GPGDs.

TMAE, TEA, TMA, and EF are the most frequently used materials.
In [153; p. 302] also a few materials with cutoff wavelengths between 200 and 229 nm are listed which are not mentioned in later reviews [194, 196, 538] anymore.

Used gaseous detector types

So far, the following gaseous cathodes and GD types have been combined in GPGDs: [196]

- **TMAE:** SWC (also commercial)[254], MWPC (also [538]), RPC, GEM[255], RETGEM[255]

- **TEA:** MWPC (also [538])

- **TMA:** SWC, MWPC

- **EF:** SWC (also commercial), MWPC, GEM, RETGEM, S-RETGEM

- **Benzene:** Ionization chamber [704], MWPC (also [538])

- **Toluene:** Ionization chamber [704], MWPC (also [538])

Properties

The properties of GPGDs were already discussed briefly on p. 696.
The operation is limited to the UV range, but the QE values there can be extremely high. The spatial uniformity is excellent, and the size of the sensitive area is almost unlimited, since for an operation at atmospheric pressure there are no mechanical constraints on the window size.
Typically, gas gains lie above 10^5, which allows spe detection. For low gains, the PHD follows an exponential distribution (Furry distribution), which leads to large

[254] In a flame sensor prototype by CERN, at low temperatures the TMAE condenses and forms a *liquid* photocathode to allow an operation in this temperature range as well.
[255] A modified layout is needed to avoid leakage currents.

threshold losses. Above gains of about 10^4–10^5, the PHD becomes a Pólya distribution (see p. 104) with a defined peak, a lower standard deviation, and therefore a higher TE.
GPGDs usually have a large TTS, since the extended conversion region introduces a strong CTS (for MWPCs in the µs range). However, with the right gas and GD type (small conversion gap), sub-ns TTS values can be reached.
Depending on the GD type, good 2D position resolutions in the low µm range are possible, but thick conversion volumes increase the position uncertainty [534; p. 91].
Photon feedback can be problematic for some gas mixtures. This can be suppressed, however, by using a cascaded GD [256], amplification gases with longer emission wavelengths (such as CH_4 or iso-C_4H_{10}), or complicated blind structures which reduce the solid angle for feedback from the emission sites.
Ion feedback – i.e. ions that drift to the cathode and emit SEs upon recombination which create APs – usually is unproblematic. First, the probability that ions reach the cathode is low. Second, only the photosensitive molecules are ionized, which is due to their low ionization threshold. For these molecules, the probability of SE emission from the cathode is very low or actually zero (again due to the low ionization potential).
GPGDs are also practically insensitive to magnetic fields.

Applications

Previous uses include RICH detectors, scintillators, plasma diagnostics and flame detectors. However, in the vast majority of cases, these were scientific devices and the commercial availability is very limited.
GPGDs continue to be widely used – mainly due to the ease of constructing large sensors – but due to the mentioned disadvantages, they are being replaced in many applications by SPGDs, diode-based sensors and MaPMTs.

Due to the lack of sensitivity in the visible range, this sensor class is not relevant for LSNDs.

Hybrid gas scintillation photodiodes (HGS-PDs)

Optical readout GPGD [257]

In a PGD, the electrons which are produced in the gas avalanche are usually read out using structured anodes. Alternatively, also the scintillation light which is generated in the avalanche can be extracted through a window and detected with a photosensor (optical readout).
In case of the DGPM (p. 706), an SPGD is used for this purpose (fig. 4.16e). An optical readout GPGD, in turn, employs an image intensifier and a CCD (fig. 4.16d). The latter layout provides a very good 2D spatial resolution and a low cost per channel [708]. In addition, it allows to separate the readout electronics from the device, in which the Townsend avalanche and possibly sparking is produced.
The drawback is that to produce sufficient light, higher gains of about 10^7 are required, which makes sparking all the more probable.

[256] For example, one or two PPACs as preamplifier stages with a subsequent long drift distance to an MWPC.
[257] Content based on [696, 705–708].

An optical readout GPGD consists of an UV-sensitive GPGD with TMAE vapors and an optical readout system.
The GPGD uses 2–3 mesh PPACs with a gate (up to 8 meshes in total), where the last PPAC is optimized for photon emission. This is achieved by a lower field strength, which results in a smaller or absent charge gain while significantly enhancing the probability to excite the TMAE molecules. These emit photons isotropically upon deexcitation (peak emission at 480 nm). Up to 5–20 photons can be emitted per electron that enters this last stage, and up to about 1000 photons are passed to the next stage, where they are amplified.
This light-amplifying GPGD is coupled to an optical readout system that consists of 2–3 image intensifiers (MCP-based and electrostatic) and a CCD.
In sum, this complex design results in a high-gain light amplifier (up to over 10^7) for a CCD.

The first optical readout GPGD prototype was constructed for HELIOS and CERES. [691, 709] This device had $20\times20\,\mathrm{cm}^2$ size, a quartz or CaF_2 input window, a triple mesh PPAC with a gate, and was filled with 40–100 Torr $C_2H_6 + 20\%$ Ar with 2.4% TMAE vapors at 34 °C [705, 706]. The last PPAC stage produced almost only light, since the field was below the threshold for charge amplification. A quartz exit window of 18 cm diameter was coupled with a lens to two image intensifiers (MCP-based and electrostatic, phosphor decay times 250 and 100 ns, total light gain $1.4 \cdot 10^4$). The image intensifiers were connected with fibers to each other and a CCD with 144×208 pixels. A total gain of over 10^7 allowed spe detection with a S/N of about 100. The light spots extended over about 100 pixels, so a centroid could be used to obtain a position resolution of ca. 300 µm at 100 Torr. Sparking would, however, have limited the gain to 10^5 for an operation in CERES, and therefore the originally foreseen optical readout was replaced by a regular pad readout in the end [710].
A similar prototype was produced at CERN with $27\times27\,\mathrm{cm}^2$ size, a 16 cm diameter quartz window, a 2-stage PPAC with $He + 3\%\,C_2H_6$ and TMAE at 1 atm and 35–45 °C, and a glass (Pyrex) exit window (see fig. 4.16d). [696] Two image intensifiers (MCP-based and electrostatic, total light gain about 1000) and the same CCD as above were used. A total gain of 10^5–10^6 and a spatial resolution below 220 µm RMS were reported.
Finally, one such sensor was designed and used in the NA35 experiment. [691, 709] This device had $50\times50\,\mathrm{cm}^2$ size, a Suprasil quartz window, a gated double mesh PPAC (total voltage 7.7 kV), used $He + 7\%\,CH_4$ saturated with TMAE at 48 °C, and had a Pyrex exit window. The GPGD was coupled with a lens to a three-stage image intensifier (electrostatic, MCP, electrostatic; all connected by fiber plates, total light gain $1.6 \cdot 10^6$) and a CCD (576×384 pixels with 23 µm, 8 bit signal, about 20 ms readout). The product of CE, ME and TE was 70–75% and a spatial resolution of below 800 µm was achieved. At the high gains which were required for spe detection, however, frequent sparking was observed, which introduced about 15% dead time due to breakdown and recovery time.

Optical readout GPGDs are irrelevant for LSNDs, since they are UV-sensitive only, are unstable, and have slow readout.
However, the optical readout of GIDs is continued to be used for particle detection [711], with current developments e.g. for the field of DM search [712].

4.4.2 Liquid photocathode

Liquid photocathode gaseous detectors (LPGDs) [258]

Photosensitive gas components like TMAE are absorbed on the walls and windows of GPGDs if these are not heated, where they form a thin photosensitive film. A liquid photocathode GD (LPGD) exploits this effect by using a thin layer of photosensitive gas on a metal cathode (plate or wires) as a liquid photocathode (see fig. 4.16c and overview of PGD classes on p. 696).
The photoionization threshold of liquids is lower than for their gas phase due to the (negative) polarization energy of the positive ion and the formation of the conduction band (energies can be lower or higher). This extends the sensitive range to longer wavelengths than for the corresponding gas. Nevertheless, liquid cathodes are still limited to the UV range.

The QE increases with the thickness of the film, which can be increased by cooling the cathode.
For thin layers, the film is mostly transparent for photons and the underlying metal is used as a photocathode. The absorbed film significantly enhances the QE of the solid cathode and extends its sensitivity to longer wavelengths. For very thin absorbed layers (one to a few monolayers), this is due to the gas molecules forming dipoles on the metal surface, which lowers the work function and improves the electron extraction [259]. For thicker (but still thin) condensed layers, a lower work function can be caused by a lower CB in the liquid than in the metal.
A thick liquid layer, on the other hand, absorbs all photons and the QE is determined solely by the properties of the liquid photocathode. The pe emission process is similar to solid photocathodes (see p. 48). Impurities (mostly oxygen and CO_2) can trap the photoelectrons, and thus purification is required to achieve the necessary low concentrations of impurities.
The photosensitive vapor can also be mixed with a solvent with high drift mobility and thermalization length, such as tetramethyl silane (TMS), tetramethyl pentane (TMP) or NP. This can increase the QE significantly due to a longer thermalization length, a higher electron temperature, and a lower CB than in the pure liquid.
Applying a strong electric field across the liquid layer enhances the pe escape probability as well, since the electrons gain kinetic energy during their drift. This raises the QE considerably.

Apart from the higher cutoff wavelength and the larger QE at long wavelengths than in GPGDs, LPGDs have a well-defined conversion position (like SPGDs), which lowers the TTS and position resolution.
However, the handling of liquid cathodes is complicated.

Liquid photocathodes were first proposed by Anderson in 1982 and were intensely studied in the late 1980s. They never gained widespread use, however, since CsI cathodes were developed shortly afterwards and showed very favorable properties.

The few liquid photocathode materials which were studied are summarized in the following list (compare the list on p. 707 for the properties of the gas phase).

[258] Content based on [196; ch. 4, pp. xv, 4, 499].
[259] The dipole bond also makes it hard to remove the molecules.

The handling of TMAE was too complicated for large particle physics experiments, which prompted the study of TEA, EF and diethylferrocene, which are non-toxic, do not react with oxygen or common fill gases, and can be operated at room temperature.

- **TMAE:** Sensitive up to 350 nm.
  - QE for absorbed (very thin) layer at room temperature (independent of fill gas and pressure): On Cu $\approx 0.013\%$ at 200 nm down to $10^{-4}\%$ at 300 nm. On Al 0.02% at 200 nm down to $4 \cdot 10^{-5}\%$ at 300 nm. The TMAE layer is difficult to remove from cathode and detector and reforms from outgassing from every surface.
  - QE for condensed (thin) layer on a cooled Cu cathode in CH_4 at 0 °C: 0.8% at 230 nm down to 0.05% at 245 nm. Isobutane reduces the QE by a factor of 2–3, likely due to an absorbed isobutane layer on the cathode that impedes electron emission.
  - QE for liquid (thick) layer of 1–3 mm: higher than for condensed layer.
  - QE for solution: For a flat cathode, the highest values are reached with 1.5 bar CH_4 + TMS on Cu at −5 °C: 2.3% at 230 nm down to 0.06% at 260 nm. The values are slightly lower if TMAE is solved in NP or TMP. With a wire cathode, the extraction fields are far stronger and with TMS at room temperature a QE close to 60% (!) at 235 nm was reported.

  However, TMAE and its solutions react strongly with oxygen, are corrosive for many materials, lead to severe aging effects in MWPCs, require to work in very clean conditions, and the photocathode surface must be cooled to achieve a liquid phase.
  Realized LPGD types: SWC [260], MWPC.

- **TEA:** Sensitive up to 240 nm. The peak QE is about 32% at 150 nm for 1.05 Torr of TEA and room temperature, but the QE is only up to ca. 1% above 170 nm.
  Realized LPGD types: MWPC.

- **EF:** Sensitive up to ca. 225 nm.
  - QE for adsorbed layer on Cu: $\approx 3\%$ at 186 nm down to 0.01% at 224 nm. The QE is slightly lower on In and Al.
  - QE for condensed layer with electric field of ca. 6 kV/cm in 0.5 atm CH_4: About 6.5% at 193 nm.
  - QE for a 1 mm thick liquid layer: 5% at 192 nm down to 2% at 200 nm.

  EF does not react with oxygen or common materials in GDs. Since the QE at long wavelengths is higher than for TEA, EF was continued to be studied for possible applications.
  Realized LPGD types: SWC (commercial flame sensor like for TMAE), PPAC.

[260] Including commercial outdoor flame detectors, where at low temperatures a gaseous cathode condenses to a liquid one to extend the range of operation.

Possible applications of LPGDs include the readout of UV crystal scintillators and the abovementioned flame sensors. Interest in their development has mostly stopped and it is difficult to even find sources that mention them, let alone properties of produced photosensors.

Since these sensors are limited to the UV range, they play no role for LSNDs.

4.5 Selection of promising types and series

After this extensive review of photosensor types, it is now possible to draw a conclusive statement about the suitability and prospects of presently known photosensors for LSNDs and similar detectors.
The following section is divided in three parts. First, the candidate sensors which were placed in the short list, because they fulfill LSND requirements or come close, are discussed in 4.5.1. The full compliance with target properties is checked, problems and exclusion criteria are addressed, and the most promising existing alternatives to PMTs are identified. After this, in 4.5.2 auspicious ongoing sensor developments are covered, which so far have not produced prototypes that fully comply with LSNDs but might do so in the future (the watch list). The chapter closes with a glance at possible future sensor types and ideas in 4.5.3 (the wish list).

4.5.1 Comparison to requirements

Exclusion criteria

LENA requirements

The compliance of photosensors with the requirements of a LSND were studied using the example of the LENA detector (see 1.4.2 and table 1.1).
In summary, the necessary properties comprise a PDE $\geq 20\%$ at 420 nm, a P/V ≥ 2, a gain sufficient for spe detection, and a dynamic range $\geq 0.3/\mathrm{cm}^2$ (per sensor cross-section). Furthermore, the requirements demand a $T_{\mathrm{FWHM}} \leq 3\,\mathrm{ns}$, an EP probability $\leq 1\%$, an LP probability $\leq 7.3\%$, and a short pulse shape (rise time, fall time, pulse width). Additionally, the physics agenda requires a DCR $\leq 2.16\,\mathrm{Hz/cm}^2$ (cross-section), an AP rate $\leq 5\%$ for 5–100 ns delay and $\leq 5\%$ for 0.1–200 µs, low light emission, and severe radiopurity requirements ($\leq 10\,\mathrm{ppb}$ ^{232}Th, $\leq 30\,\mathrm{ppb}$ ^{238}U, $\leq 2.4\,\mathrm{ppb}$ ^{40}K). A compressive strength of 1.3 MPa would be beneficial but is not required, since optical modules (OMs) can be used. The same holds for GMF tolerance, where magnetic shields can compensate deficiencies. The required properties must be achieved at 22–23 °C and for a lifetime of at least 30 a. Whether compatibility with the buffer liquid is necessary, depends on the sensor radioactivity. Furthermore, organizational aspects such as a sufficient yearly production capacity and affordable costs must be fulfilled.
Compared to other LSNDs, these requirements are modest for the most part, except for the DCR, TTS, radiopurity, lifetime, compressive strength and EAP rate. Nevertheless, they exclude the vast majority of alternative photosensors which were

analyzed in 4.2–4.4, which already was a preselection (visible-/UV-sensitive photosensors capable of room temperature operation) of the types discussed in 4.1.1.1.

Frequency of the reasons for exclusion

The suitability of each sensor for LSNDs was already shortly summarized at the end of the respective description.
The most frequent reasons for exclusion of a sensor (multiple reasons are possible) were: A significantly too high TTS (26 sensors), incapability of spe detection[261] (18), a considerably too high DCR (14), absence of fully functional prototypes (12), very small active areas[262] (12), too low PDE (9), high costs (7), and insensitivity in the visible range (5). Other reasons included insufficient lifetime (3), long dead times (3), too high LP probability[263] (2), inadequate dynamic range (2), too low P/V (1), and that the sensor is not produced anymore (1). In addition, one sensor each required operation at cryogenic temperature (VLPC) or in magnetic fields (ISPA-tube).

Solvable problems

There were also several hybrid sensors in which a (large) negative voltage is applied to the cathode. This could be problematic for operation in a liquid[264], but was considered a problem rather than an exclusion criterion, because it might be possible to e.g. adapt the sensor layout for reverse polarity or to include safety gaps in an OM. This would have to be addressed if a sensor with negative HV is selected.

Exclusion criteria for sensor classes

When considering groups and classes of sensors, common exclusion reasons become apparent.

Diode-based sensors without gain (PD, PDA, PSD) or which use impact ionization[265] either cannot detect single photons due to insufficient gain or have too high DCRs at room temperature. This not necessarily extends to diode-based hybrid sensors (see below).

Sensors that employ conversion gain[266] in general have a too high TTS and cannot count photons without additional gain steps or sensitivity-boosting readout structures (CCD, pnCCD).

Devices with transistor amplification[267] or conservation gain (photoresistor) usually have either a too high TTS or a gain which is insufficient for spe detection.

[261] Sensors with no or too low gain; see p. 597 regarding what is considered "too low" for the respective sensors.
[262] This might be considered a (solvable) problem rather than an exclusion criterion, but no sensor is eliminated due to small size alone.
[263] The LP requirement might have to be reevaluated, but both sensors are excluded due to other reasons as well.
[264] A potential gradient between window and cathode can also poison and destroy the photocathode (see p. 60).
[265] Linear mode: APD, APD array, PSAPD, EMCCD, EMCMOS, VLPC. Geiger mode: SPAD, SPADA, SiPM, dSiPM.
[266] CCD, pnCCD, skipper CCD, QIS, EMCCD, IEBCCD, ICCD, EBCCD.
[267] Phototransistor, photo-Darlington, CIS, sCMOS, IEBCMOS, ICMOS, EBCMOS.

For the phototransistor, photo-Darlington and photoresistor one has to choose between both drawbacks. The newest members of the Timepix CMOS family and the hybrids based on them are a notable exception.

Phototubes with no amplification (phototube) or very small gain (gas-filled phototube, VPT) cannot detect single photons.

Regarding sensors with continuous dynodes, the CEM is limited by dead time and a dynamic range of 1 pe. This is resolved in the MCP-PMT through parallel channels but comes at the cost of a low PDE (MCP open-area ratio) and high costs (lead-glass MCPs). Newer MCP-PMT variants (LAPPD, fountain-focused MCP-PMTs) promise to solve these problems and enter the short list.
HM-PDs and HMS-PDs also can have high costs due to the inbuilt MCPs, but their exclusion criteria are mostly determined by the secondary sensor (CCD or CIS), and future CMOS-based sensors might be eligible.

The properties of photosensors with electron-bombarded scintillators (HS-PMTs, HS-PDs, HS-SiPMs, HS-PGDs, HMS-PDs) crucially depend on the used scintillator. Backscattering can cause large LP probabilities and high EAP rates, reduces the P/V and lowers the ME (and thus the PDE); the higher the Z_{eff} of the scintillator, the higher the backscattering coefficient k.
A short decay time τ and a high ly produce a narrow, well-defined pulse shape with low fluctuations from the emission times of individual photons. This reduces the TTS (τ and ly) and the charge resolution (ly). Otherwise, it can also be very difficult to extract the time of overlapping pulses from a recorded waveform [268]. The quantity ly/τ can be used as a figure of merit (similar to the one presented in [627]) and should be maximal.
The following list summarizes the properties of important scintillators: [239, 281, 283, 553, 558, 621, 627]

- **YSO** (Y_2SiO_5:Ce, "P47"): $\tau = 30$–$60\,\text{ns}$, $\text{ly} = 10\,\text{photons/keV}$, $Z_{\text{eff}} = 33$, $\text{ly}/\tau = 0.17$–$0.33/\text{keV ns}$
- **LYSO** ($Lu_{2(1\text{-}x)}Y_{2x}SiO_5$): $\tau = 42\,\text{ns}$, $\text{ly} = 26$–$30/\text{keV}$, $Z_{\text{eff}} \approx 63$–$66$, $k = 50\%$ at $20\,\text{kV}$ [558] / 30–40% at $25\,\text{kV}$ [283], $\text{ly}/\tau = 0.62$–$0.71/\text{keV ns}$
- **LSO** (Lu_2SiO_5:Ce): $\tau = 40\,\text{ns}$, $\text{ly} = 30/\text{keV}$, $Z_{\text{eff}} = 66$, $\text{ly}/\tau = 0.75/\text{keV ns}$
- **YAP** ($YAlO_3$:Ce): $\tau = 25$–$30\,\text{ns}$, $\text{ly} = 16/\text{keV}$, $Z_{\text{eff}} = 36$, $\text{ly}/\tau = 0.53$–$0.64/\text{keV ns}$
- **SBO** ($ScBO_3$:Ce): $\tau = 30\,\text{ns}$, $\text{ly} = 16/\text{keV}$, $Z_{\text{eff}} = 17$, $\text{ly}/\tau = 0.53/\text{keV ns}$
- **ZnO:Ga**: $\tau = 0.4\,\text{ns}$, $\text{ly} = 16$–$40/\text{keV}$, $Z_{\text{eff}} = 28$, $\text{ly}/\tau = 40$–$100/\text{keV ns}$

The figure of merit reproduces the qualitative TTS behavior found in the QUASAR variants (see p. 654).
According to this list, ZnO:Ga would be the best scintillator by far, but as of now only small scale prototype sensors with low quality could be produced [348, 628]. For the remaining materials, a compromise must be made between low backscattering or a good TTS and charge resolution. Usually LYSO or LSO are chosen at the cost of a high k.

[268] Compare fig. 7 in [502] (QUASAR-370 with YSO) with fig. 2 in [283] (ABALONE with LYSO).

HS-PMTs (SMART, QUASAR-370, X-HPD) and HS-SiPMs (Reference, ArcaLux, ABALONE, SiPMT) both show good properties. However, HS-PMTs are currently not being produced or pursued, while two HS-SiPMs are under development.
The single HS-PD sensor (streak camera) is not designed for large-area detection and an HS-SiPM would have better performance than an HS-PD.
The one existing HS-PGD (HGPM) had a very low PDE, likely due to coupling losses, which might be a general problem of this sensor class.

For sensors with electron-bombarded semiconductor sensors, the HAPD class seems to be the ideal middle course between HPDs (no spe detection or too large TTS) and HSiPMs (extreme DCRs). The properties of HAPDs can even be slightly better than for PMTs (smaller charge resolution, high CE, good TTS, low AP rates, better magnetic field tolerance) but due to their lower gain, front-end preamplifiers are mandatory. Currently, for HPDs one has to chose between spe detection or low TTS, but this predicament might be removed with next-generation CMOS readouts. HPDs in most cases have internal amplifiers (except IPD, CMS, C2GT), whereas HAPDs and HSiPMs can use external ones.

Most photosensors with gas amplification are UV-sensitive only (GPGDs, LPGDs, UV-SPGDs, HGS-PDs). So far, no sensor has been created which is sensitive in the visible range and shows comparable performance to PMTs. Currently, VPGDs either have short lifetimes (and high iAP rates) due to IBF or are limited to gains that are insufficient for photon counting. The hybrid gas photomultiplier has not produced working prototypes, and the HGS-PGD class (DGPM) likely is not more promising than VPGDs.

Shortlist sensors

After a loose application of the LSND requirements, the following sensors were considered to be most relevant for LSNDs and were placed in the short list:

- PMTs (chs. 2 and 3)
- IEBCMOS (p. 646)
- GDB-6201 and GDB-6203 (p. 643)
- LAPPD and Argonne LAPPD (p. 641)
- QUASAR-370(LSO) (p. 654)
- ABALONE (p. 661)
- QUPID (p. 680)
- Belle II HAPD (p. 682)
- HHPDs: R12850, 13” prototype, R12112, and 5” prototype (pp. 684–686).

These sensors are highlighted in fig. 4.1 in bold font.

Properties

The properties of the shortlisted photosensors as stated in their respective section are summarized in table 4.1.

Here, missing PDE values were estimated based on QE, CE and FF values where possible. An unknown ME cannot be predicted, however, since the fractions of backscattered electrons that result in late pulses and losses cannot be assessed without measurements or simulations. The same holds for the TE. Because the ME and for low gains also the TE will lead to PDE reductions whose extent is not exactly known, the estimated PDE values are marked by a prefixed "≈" where no large reductions through ME and TE are expected, and otherwise a preceding "$\lessapprox$".
The ME was not specified for any sensor, although it is relevant for all of them. The TE is not known either, and possibly is notably below 100% for sensors with low gain and comparatively high ENC (i.e. a low S/N). This might be the case for some HAPDs and possibly one IEBCMOS subtype.
The R12850 is the only fountain-focused sensor that might be able to achieve a coarse position resolution, which is possible through the use of a segmented APD with 2–5 channels [269].
The radiopurity values for PMTs that are given in Bq/kg are from [309] and were measured for glass samples. Therefore, their compliance with requirements cannot be stated, since the amount of glass in the PMT is not known. The QUPID radiopurity values (in mBq for the whole sensor) cannot be compared with the limits either, since the weight of a QUPID could not be found in literature.
Some properties (EP probability, EAP rate, mean life) unfortunately are known for no sensor except for some PMTs series.

This table can be compared to the characteristics of relevant PMT series which were summarized in table 3.7.
To facilitate the comparison, the value range that is spanned by these PMT series is listed in the first column of table 4.1. To determine the range, for each series the mean value was taken; where no mean was available, the minimum/maximum numbers were used. Independent measurements were preferred over manufacturer data. The HV range of PMTs is specified for a gain of 10^7, and the HV polarity of PMTs can be changed by using a different voltage divider. For the DCR, the lowest available value of each series was used to estimate the achievable range. The mean life is listed according to 2.7.3. The resulting values of each property not necessarily represent all PMT series, since only large PMTs with LF and B+L dynodes were considered and for many properties values are only known for a few series.

Many of the photosensors which were discussed in 4.2–4.4 and also several sensors in the shortlist have a position resolution. While this provides more information than a mere point sensor, for a large neutrino detector with thousands of m² instrumented surface, which consists of several 10000 large photosensors, the position, time and track resolution likely will not benefit much from image or position sensors. The limiting factor in the employed reconstruction algorithms is not the position resolution of the sensors but statistics (the low number of detected photons). This is evidenced by the tendency to make the PMTs as large as possible in these experiments.

[269] The segmentation was originally introduced to lower the capacitance and thus improve the TTS and pulse shape.

Table 4.1: Properties of photosensors that were shortlisted for liquid scintillator neutrino detectors.
Values listed if known, for concept studies in brackets. Numbers in red if not fulfilling requirements (see LENA values in table 1.1); dark count rate (DCR) in red if more than 3 times the required value. The used sources are listed in the respective photosensor sections. Table continued on next page. Discussion in text.
In the high voltage (HV) values, for HAPDs the electron bombardment voltage and the APD bias are listed where known. Since all shortlist sensors are based on vacuum tubes, the focusing type (see p. 572) is listed as well. Where the photodetection efficiency (PDE) was unknown, it was estimated based on values of the quantum efficiency (QE), collection efficiency (CE) and fill factor (FF) without taking into account unknown and incalculable factors (multiplication efficiency, threshold efficiency). CE and FF are summarized to one value where both are relevant (prefixed by "<" where one of both is unknown). The charge resolution R_σ (p. 119) and the TTS σ_T (p. 168) were converted to R_F and T_{FWHM} using eq. 2.68, which then is marked by a preceding "≈"; for one sensor only the value for the pulse height distribution (PHD) was available. For the linear range, the specified gain decrease at which the value was measured is listed in parentheses. The position resolution is given in mm/µm FWHM (full width at half maximum) or as the number of channels. The DCR for the R12112 was only available for a prototype with a standard QE (SQE) cathode and is likely higher now due to the use of high QE cathodes. For the ionic afterpulse (iAP) rate, in some cases only upper limits are available. Radiopurity values listed in Bq/kg or mBq were not assessed for compliance, since the sensor masses were unknown. The value ranges for PMTs are based on the series in table 3.7, the HV range of PMTs uses [160, 213, 263, 276, 306, 309, 335, 439, 450].

Sensor	**PMT**	**IEB-CMOS**	**GDB-6201**	**GDB-6203**	**LAPPD**	**Argonne LAPPD**	**QUASAR-370(LSO)**	**ABALONE**	**QUPID**	**Belle II HAPD**	**R12850**	**13" HHPD**	**R12112**	**5" HHPD**
Size [cm]	⌀ 7.6–50.8	1.5×1.5–2.8×2.8 [4.0×3.5]	⌀ 50.8	⌀ 50.8	23×22		⌀ 37	⌀ ≈ 10	⌀ 7.1	7.3×7.3	⌀ 50.8	⌀ 33.2	⌀ 20	⌀ 12.8
Sensitive area [cm]	⌀ 6.5–46				19.5×19.5	6×6			⌀ 6.4	7×7	⌀ > 46	⌀ 30		⌀ 8
HV [kV]	1.14–2.15	3	1.74	1.9	2.0–2.4		25	25	6 + 0.36	8 + 0.25–0.5	8 + 0.4–0.5	10–20	8–10 + 0.2–0.37	8.5 + 0.35
HV polarity	+/−	+	+	+	+	+	+	+	−	−	+	+	+	−
Focusing type	Fountain	Proximity	Fountain	Fountain	Proximity	Proximity	Fountain	Fountain	Fountain	Proximity	Fountain	Fountain	Fountain	Fountain
PDE at 420 nm [%]	23–28		30.9 (400 nm)	≈ 25.5 (400 nm)	≈ 14.1/17.5				≲ 29.6	≲ 20.6 (400 nm)	≲ 32.0	≲ 19.4	≲ 31.4	
QE at 420 nm [%]	22 (420 nm)–37 (408 nm)		31.3 (400 nm)	30 (400 nm)	≈ 21/26	12.5	20–24		≈ 37	31.7 (400 nm)	≈ 33	≈ 20	≈ 33	
CE / FF [%]	90–95		98 [98.8]	85	≈ 67			≈ 100	> 80	< 65	97	95–97	95	
Max. gain g [1]	$(1–10) \cdot 10^7$	$\approx 10^4$ [few 10^4]	10^7		10^6–10^7	$(3–14) \cdot 10^6$	$(2–5) \cdot 10^8$	$\approx 6 \cdot 10^8$	$(1.5–2.4) \cdot 10^5$	$> 7 \cdot 10^4$	10^5	$(1–2) \cdot 10^5$	$(4–9) \cdot 10^4$	$4.4 \cdot 10^4$
Charge resolution R_F [%]	59–106		≈ 73	≈ 94			70–80 LSO: 35			≈ 41	≈ 24–35	24 (PHD)	≈ 24	
P/V [1]	2.1–4.1		7.1		1.5–2		2.5	≈ 2.9	6–7.5				3.6–5.4	
Linear range [pe]	250–370		1400 (10%)		500 kHz/cm² (10%)				$\approx 10^5$ (5%)					
Dynamic range [pe]	> 3000											3000	$1.5 \cdot 10^4$	

Sensor	PMT	IEB-CMOS	GDB-6201	GDB-6203	LAPPD	Argonne LAPPD	QUASAR-370(LSO)	ABALONE	QUPID	Belle II HAPD	R12850	13" HHPD	R12112	5" HHPD
T_{FWHM} [ns]	1.5–4.7	0.13 [≈ 0.05–0.07]	20.0 [< 3]	4.3	≈ 0.12–0.17	≈ 0.08–0.16	2 LSO: 1	0.55	< 0.16		3.2	≈ 0.45	1.1	≈ 0.94
Transit time [ns]	40–95							1.5–2.3						
EP probability [%]	0.6–2.7													
LP probability [%]	2.3–7.9		[< 2]		≈ 4									
Pulse height [mV]	6.5–19.8							≈ 6		5–10	≈ 10		≈ 10	18
Rise time t_r [ns]	2.7–6.9		1.44		0.85	0.44–0.54			1.8			≈ 1		
Fall time t_f [ns]	3.5–10.2		23.5			1.2			2.5					
Pulse width t_w [ns]	4.5–11.6				1.1	0.8		≈ 150	4.2	few 100				5
Pulse length t_l [ns]											100		≈ 20	
Position resolution FWHM [mm]	no	5–7 µm [5–10 µm]	no	no	2.4/0.76		no	no	no	5	5-ch	no	no	no
DCR [kHz]	0.8–8		37.8		several 100 Hz/cm²		39		7.5		8		≈ 1 (SQE)	2.56
EAP rate [%]	2.7–3.4													
iAP rate [%]	1.0–16		1.0				< 0.1	0.5			< 2.3			
Radiopurity K_{nat} [ppm], ^{232}Th [ppb], ^{238}U [ppb] (Bq/kg)	60 (0.6–4.0), 30 (0.2–0.3), 30 (0.9–1.0)								0.4 mBq, 0.3 mBq, 17.3 mBq					
Magnetic field tolerance	32–50 µT			> 100 mT		≈ 0.64 T	GMF			1.5 T				
Pressure capability [MPa]	0.2–1.25										1.25		0.9	
Mean life [a]	57–571													

A higher granularity could be reached by using a larger number of smaller PMTs, but apparently this is not a crucial issue and other properties determine the choice of sensor size.
In short, unlike for high energy physics, where track reconstruction with µm accuracies is required, for LSNDs a position resolution is optional. Nevertheless, it would be interesting to study by how much the detector performance could be improved by a sensor with extremely good position resolution and TTS, such as the LAPPD. The THEIA project (see 1.3.3) might provide valuable insights in this regard.

Comparison of sensors

In the following, the properties of the sensors in the table shall be compared to PMTs and to each other, where reasonable.

MCP-PMTs, HAPDs and the QUASAR-370 can reach sizes that are comparable to those of PMTs. The Argonne LAPPD, QUPID, ABALONE, Belle II HAPD and particularly IEBCMOS sensors are small by comparison, which might require a different instrumentation approach in a LSND.

For PMTs, MCP-PMTs and IEBCMOS sensors, only moderate voltages are necessary, whereas most HAPDs need high voltages, and the 13" HHPD and the scintillator-based hybrids (QUASAR-370 and ABALONE) even require very high voltages.
Some HAPDs have negative polarity, which might have to be adapted for use in a liquid.
Most sensors are fountain-focused, but proximity-focused sensors are present as well (LAPPDs, IEBCMOS, Belle II HAPD), whereas cross-focusing (which admittedly is seldom used) is absent.

The PDE mostly lies in two groups: One with high values around 30% (HQE PMTs, GDB-6201, QUPID, R12850, R12112) and one with about 20% or less (PMTs, LAPPDs, QUASAR-370, ABALONE[270], Belle II HAPD, 13" HHPD).
This is closely connected to the QE, for which the best values are reached for some PMTs (HQE bialkali) and the QUPID (low temperature bialkali). Many sensors have HQE cathodes, which reach 30% or more at 420 nm, whereas low QE values are found for some PMTs, the QUASAR-370 (although the double cathode effect raises the QE somewhat) and the 13" HHPD. Particularly low QEs result for the Argonne LAPPD and the ABALONE.
The CE in most cases is close to 100% (GDB-6201, ABALONE, R12850, 13" HHPD, R12112, most likely also QUASAR-370). Moderate CEs are reached for the GDB-6203 due to a tradeoff between TTS and CE and the QUPID (likely for the same reason). For some sensors the FF is quite low (LAPPD, Belle II HAPD), which significantly lowers their PDE.

The total gain is quite low for IEBCMOS sensors and HAPDs, which requires the use of front-end preamplifiers. For MCP-PMTs, the gain is comparable to PMTs, whereas scintillator-based hybrids (QUASAR-370 and ABALONE) can produce very high gains.

[270] The ABALONE currently employs a Cs_3Sb cathode; this material can reach a QE of up to 14.6% at 400 nm, see p. 55.

The charge resolution is best for HAPDs and the QUASAR-370LSO. For the GDB-6201/6203 and the QUASAR-370 it is similar to PMTs.
The related P/V is highest for the QUPID and the GDB-6201, still high for the R12112, comparable to PMTs for the QUASAR-370 and ABALONE, and very low for the LAPPD. The P/V values correspond to the quality of the charge resolution except for the GDB-6201, which only has a moderate charge resolution.

The linear and dynamic ranges are at least comparable to PMTs where values are available. For the LAPPD, no comparison is possible due to the format of the value.

The TTS is excellent for the IEBCMOS, LAPPDs and QUPID, very good for the 13" HHPD, ABALONE, 5" HHPD, R12112 and QUASAR-370LSO, and good for some PMTs and the QUASAR-370. The values are still sufficient for many PMTs series, the R12850 and possibly an improved GDB-6201 but too high for the GDB-6203, GDB-6201 and some PMTs.

For the EP and LP probabilities, almost no values are available, but in some cases this can be estimated qualitatively from published TTDs. For many sensors, high LP occurrences are expected due to backscattering, and the exact values will have to be measured to ascertain the compliance with LSND limits.

Regarding the pulse shape, the pulse height is comparable to PMTs where known. The rise time (important for the TTS) is very short for the LAPPDs and the 13" HHPD, and short for the GDB-6201 and QUPID.
The fall time is small for the Argonne LAPPD and QUPID, and large for the GDB-6201.
The pulse width is very low for the LAPPDs, similar to PMTs for the QUPID and 5" HHPD, and very large for the ABALONE and Belle II HAPD due to the used preamplifiers.
Finally, the pulse length is comparable to PMTs for the R12112 and large for the R12850, which also is owed to the amplifier.

Only some photosensors are position sensitive. The spatial resolution is excellent for IEBCMOS sensors, good for the LAPPDs and the Belle II HAPD, and very coarse for the R12850.

Not many sensors achieve DCRs which are as low as for the best PMTs. This is only the case for the R12850 and possibly the R12112 (has to be measured for HQE yet). Values which are higher but still similar to PMTs are reached by the GDB-6201 (two photocathodes), QUASAR-370 and 5" HHPD, while the QUPID and LAPPD have very high DCRs due to their photocathode materials.

The EAP rate was not measured yet for any of the shortlisted sensors, apart from some PMT series. It can only be guessed that it will be far from negligible for many types due to pronounced inelastic scattering, which requires further measurements. The iAP rate is excellent for the QUASAR-370, very good for the ABALONE and GDB-6201, and good for the R12850.

Basically all sensors – even if not listed in the table – show a better resistance to magnetic fields than PMTs due to higher voltages and/or proximity focusing. The

QUASAR-370 needs no magnetic shielding from the GMF, while the GDB-6203 has a very high tolerance of over 100 mT, and the Argonne LAPPD and Belle II HAPD even are suited for extremely high fields.

The pressure resistance is only known for some HHPDs, which achieve high values. Proximity-focused sensors will have low pressure capabilities due to their flat design, which is unfavorable for this, particularly for large areas. As an example, in the current design of the 8×8" LAPPD already large spacers, which occupy 9% of the cathode area, are required merely to withstand the ambient pressure.

Comparison of sensor classes

In addition, some general observations about the represented sensor classes can be made.

Compared to the other sensors, PMTs give a mixed picture: They have low operating voltages, a rather good PDE, high gains (likely no preamplifiers required), low DCRs, and high lifetimes. On the other hand, they cannot reliably count photons (high charge resolution and low P/V) and have a rather large TTS, no position sensitivity, and a very low magnetic field tolerance. In sum, however, they manage to (in some cases barely) fulfill all LSND requirements, which is more than can be said for most alternative sensors.

Although both the GDB-6201/6203 and the LAPPDs are MCP-PMTs, they are very different sensors and show as many similarities as differences. Both types achieve high gains with low voltages, have similarly high uncertainties at photon counting as PMTs, and show high DCRs.

Scintillator-based hybrids (QUASAR-370 and ABALONE) can achieve extremely high gains due to very large EB voltages and a large secondary gain (PMT or SiPM). Their charge resolution is comparable or better (LSO) than for PMTs, and they show very good TTS values and low iAP rates.

HAPDs have low gains around 10^5 despite high to very high EB voltages but show very good charge resolutions, which allows to reliably count up to several photons, and also a high dynamic range. The TTS is typically very good, the pulse shape strongly depends on the preamplifier, and the DCR can be as low as in PMTs. Position-sensitive layouts with large areas can be realized.

Correlations between properties

The table can also be used to analyze correlations between parameters. However, in many cases a dependence between properties is difficult to verify, since often values are missing and the characteristics of a sensor depend far more on the detection process and device design than any single other variable. The statements which can be made thus are quite general and mostly were already noted before.

A larger sensitive area increases the TTS and transit time (except for proximity focusing and higher voltages). Proximity-focused sensors usually have rather small areas, with the exception of the LAPPD. The more compact the layout, the higher the tolerated magnetic field is.

A higher voltage improves the CE and charge resolution (depends strongly on the detection mechanism), reduces the transit time (no data) and raises the magnetic field tolerance.

Proximity-focused sensors have an excellent TTS, a good to outstanding position resolution, and a very high magnetic field resistance. However, all such sensors also show significant FF losses, which leads to a lower PDE, and a low pressure capability (see above).

There is a correlation between the values of CE and TTS, since the electron optics can only be optimized for one of both, and generally a lower TTS leads to a lower CE (compare GDB-6201 and GDB-6203). However, there are also sensors which achieve very good values for both (ABALONE, 13" HHPD, R12112).

An anticorrelation is expected between the P/V and the EP, LP and EAP rates, since higher rates of the latter cause more undersized pulses in the SER, but no data was available to confirm this hypothesis.

Exclusion

For the selection of shortlist sensors, the LSND criteria were only loosely applied to allow a comparison of promising sensor types.
Values which do not meet the required limits are colored red in table 4.1 and for a strict application of the requirements, only some PMT series and the IEBCMOS, ABALONE, Belle II HAPD and R12112 remain in the selection. Due to currently missing values, however, likely even more sensors will have to be excluded.
In the following, the drawbacks of each shortlisted sensor will be reviewed, before a conclusive statement about possible alternatives for PMTs is drawn.

IEBCMOS

These sensors have a slow readout but a very good TTS. Since they cannot produce a trigger and require triggered readout, the detector electronics would have to be adapted. Because of the small sensor size (limited by the CMOS chip size due to proximity-focusing), a front-end readout would be necessary to reduce the number of channels.
The high power consumption of a few watts per sensor might introduce two practical problems in a LSND. First, the heat generation might induce convection, distribute radioactive impurities, and contaminate the fiducial volume. However, this might be manageable by a greater distance from the detector center through the use of LCs and with a separating layer between the LC apertures. Second, for large detectors the electricity supply starts to become an issue. For a detector like LENA with 9772 m^2 surface and 30% OC, and assuming an LC with an effective magnification of 1.24, the larger IEBCMOS of the concept study, and an optimistic power consumption of 2 W, this results in $1.7 \cdot 10^6$ sensors and 3.4 MW.
Finally, the cost and availability have to be determined, as MCPs can be quite expensive and for such a large quantity of devices these concerns can easily exclude a sensor. The costs might be lowered through the use of ALD-GCA-MCPs.

The currently known properties fulfill all limits, and only potentially solvable problems have been identified thus far. To allow a final assessment, however, the

missing crucial properties have to be determined. For the time being, the IEBCMOS is a PMT alternative.

GDB-6201

The TTS is significantly too high but might be lowered below the limit of 3 ns through several suggested modifications.
The DCR is 8.6 times higher than allowed, which is still within the factor of 10 decrease that is observed after long times in the dark for some PMTs. Whether this is the case would have to be verified.
Also, a high LP rate with a lateness of up to 22 ns is observed.
The production capacity should not be an issue for this variant or the GDB-6203, when considering that 15000 GDB-6201 sensors have recently been fabricated for JUNO [604].

At the moment, the GDB-6201 does not fulfill the criteria, but a version with an improved TTS would warrant further consideration.

GDB-6203

This series has a significantly better TTS (4.3 ns) than the GDB-6201; however, this value is still too high.
The magnetic field tolerance of over 100 mT is remarkable when considering that this is a 20" fountain-focused design with only 1.9 kV HV – PMTs have a tolerance that is over a factor of 1000 lower for similar cathode shapes and voltages. This is likely due to the asterisk-shaped focusing electrodes.

Presently, the TTS excludes this sensor, but similar improvements as suggested for the GDB-6201 should be able to significantly improve the TTS of the GDB-6203 as well.

LAPPD

Currently, LAPPDs have DCRs which are far too high for large LSNDs and can not be expected to decrease sufficiently with time. It is not clear from literature whether a Na_2KSb (high-temperature bialkali) or a Na_2KSb:Cs (multialkali) cathode is used. The former has a lower DCR than bialkali at room temperature, while the latter has a significantly higher dark noise; both have a far lower QE than bialkali (see p. 55). The measured DCR indicates the use of multialkali, so there likely is room for improvements by using a bialkali cathode.
At the moment, the PDE is not fulfilled either: The cathode material limits the QE, while the CE is basically equivalent to the open area ratio, and the spacer introduces 9% FF losses in the cathode area. Even when assuming the most optimistic open area ratio and a quartz window, the PDE reaches only 17.5% at 420 nm. The PDE would benefit from a (HQE) bialkali cathode, and the target PDE of 20% could be reached for a QE of 29.7%. Nevertheless, due to the quadratic sensor shape, a higher OC can be reached in a detector. This compensates the low PDE, which thus is not an exclusion criterion anymore.
The P/V is barely sufficient or too low.
Furthermore, the dynamic range can be problematic: While high rates are possible, the maximum number of coincident photons is only 4 – this will likely be significantly better for GEN II devices.

A quite high LP rate was observed, but due to the short maximum lateness of 1.7 ns, this should not be problematic. Some EPs with an earliness of up to 400 ps were found, which should be acceptable as well.
Finally, the production capacity is still only at the level of several 10 sensor per year, would not nearly be sufficient for a large LSND.

In sum, the LAPPD currently does not fulfill LSND requirements, but it is probable that this can be the case with some improvements, and the sensor holds tremendous potential (ps timing with mm position accuracy). Therefore, close attention should be paid to its future development.

Argonne LAPPD

This LAPPD variant has a very low QE. The CE and FF will be similar to the LAPPD, in which case the PDE only amounts to 8.4%. While the quadratic layout enables higher OCs, this likely cannot compensate such a low PDE, however, because the sensor has a large insensitive border [271] and in a large LSND pressure housings [272] would be required. Thus, further photocathode improvements would be needed, which should be possible.
The performance notably depends on the capillary diameter: For 20 µm, the properties are far worse than for 10 µm (higher TTS, high LP probability, longer rise time, lower magnetic field tolerance). 6µm pores are under development and could yield further improvements.
Several important properties (DCR, charge resolution or P/V, and dynamic range) are currently not known.
Furthermore, the small size would necessitate an adapted readout scheme.

Altogether, the low PDE currently excludes the Argonne LAPPD. However, this might be improved in the future and thus further progress should be watched.

QUASAR-370

For the QUASAR-370, the QE is quite low despite the double cathode effect. Combined with the expected CE (likely close to 100%) and ME (probably notably lower than 100% due to backscattering loses), the PDE likely is just under 20%.
The DCR is too high by a factor of 12.9, which makes it unlikely that a sufficient drop with time will occur.
In addition, the QUASAR-370 is not produced anymore.
The LP rate is reduced by the dual triggering method, but despite of this, the main peak in the TTD still shows a strong trailing flank up to 7 ns lateness, and a small EP contribution on the rising flank is expected. Both should be within limits, however.
The EAP rate is not known, but a notable amount of electronic APs due to inelastic backscattering can be expected.
The dynamic range will be lower than that of the UGON PMT by a factor which equals the dpy of the scintillator and thus should be determined.
For the LSO variant, the LP probability and EAP rate likely exceed the requirements.

In sum, the QUASAR-370 just fails to fulfill the LSND criteria (which could be slightly adapted) and is excluded, since the production was stopped.

[271] The PCB is far larger than the window (cf. fig. 2a in [590]).
[272] Which also introduce a FF.

ABALONE

For the ABALONE, many critical properties are missing, but the PDE currently probably is too low, since the Cs_3Sb cathode limits the QE to about 15% (p. 55) and the ME will be reduced by scattering losses.
In all likelihood, the DCR of the cathode will also be too high, as will the LP and EAP rate.
The dynamic range is reduced by the dpy and should be determined.
The small sensor size requires modified electronics, but the claimed final sensor costs are quite low.
Currently, the production capacity is too low but is planned to be scaled up.

So far, the ABALONE is only in a prototype stage, and currently the sensor most likely will not fulfill LSND requirements. Since the development is ongoing and the properties can be expected to improve notably, this sensor will be a future contender.

QUPID

Due to its low-temperature bialkali cathode, the DCR of the QUPID is a factor of 88 too high at room temperature, which is unacceptable for large LSNDs.
In addition, the development was stopped by Hamamatsu.
While the PDE will be further lowered by the ME, it will most likely still exceed 20% thanks to the large QE.
The linear range was measured with a different method than usual but should likely be sufficient, since the resulting value was very high.
The negative voltage and the small sensor size would require some adaptations in the electronics.

All in all, the QUPID can not substitute PMTs.

Belle II HAPD

Despite the high QE of the Belle II HAPD, the low FF reduces the PDE to only barely above 20%. When including expectable values for the ME and CE, the PDE is very likely lower than 20%.
The LP probability should be large, and the HQE cathode will also cause a high DCR.
The TTS likely is very good but this has to be measured, along with other missing parameters.
The negative voltage and small size pose a challenge, but are no exclusion reasons.
For Belle II, 500 devices were produced by Hamamatsu, and the production capability can likely be ramped up sufficiently.

The low PDE just eliminates the Belle II HAPD.

R12850

For the Hamamatsu R12850, the TTS of 3.2 ns is slightly too high but similar to some promising PMT series, and it is planned to reduce the TTS by a better preamplifier or an improved cathode shape. Therefore, the R12850 should not be excluded based on this alone.

The DCR is a factor of 1.8 too high, which is well within the expectable reduction through longer times in the dark.
The ME is unknown and likely notably below 100%, but the PDE should still be high.
A low number of LPs are observed with a lateness of up to 18 ns, which probably is within allowed limits.
The EAP rate and dynamic range should still be assessed.
No mass production technique was established yet, but it can be expected that Hamamatsu can accomplish this.

In sum, the R12850 is a strong alternative candidate to PMTs, although its TTS should still be improved.

13" HHPD

In the 13" HHPD prototype, the peak QE is only 20% at an unknown wavelength. Based on the expected high CE and non-ideal ME, it can therefore safely be assumed that the PDE is below 20%.

This already suffices to exclude the sensor.

R12112

For the R12112 HHPD, the DCR is too high by a factor of 1.5. This would well be acceptable; however, this value was measured for an earlier version with an SQE cathode, while now an HQE variant is used, which typically show far higher DCRs. Thus, the compliance of the DCR must be verified, since this sensor is a strong candidate.
While the ME and TE are unknown, the QE is high enough that it can be assumed that the PDE will lie above 20%.

Until the DCR can be measured, the R12112 is another PMT alternative.

5" HHPD

For this sensor, the DCR is too high by a factor of 9.2, which is still within the possible reduction with time, although one cannot safely assume that such a large decrease will be the case.
Due to the low gain, the TE might be $< 100\%$, and many crucial properties are not known.
Furthermore, the negative voltage might have to be addressed.

In sum, similar to the IEBCMOS, too little is known about this sensor to arrive at a conclusive statement. The known properties fulfill the requirements, although it is expected that the later HHPD prototypes and series are superior in performance, since otherwise the development and characterization of this sensor would have been pursued further.
Until the missing properties can be determined, the 5" HHPD prototype is thus – unexpectedly – a potential alternative sensor.

Summary

Judging from these observations, the most severe LSND requirements for shortlist sensors are (multiple reasons are possible): the PDE (excluding 5 sensors), DCR (4), TTS (2) and production capacity (2).

Based on a strict application of the requirements to the known properties, there are three sensors which might substitute PMTs: the **R12112**, **IEBCMOS** and **5" HHPD**. For the last two, many properties have yet to be determined and their compliance with demands is uncertain.
The **R12850** is a strong alternative, as well, if the preamplifier is improved or the TTS requirement is slightly relaxed.
Of these four sensors the 20" and 8" HHPDs are considered the most promising.

Several of the currently excluded shortlisted sensors show room for improvements and are continued to be developed. These sensors thus should be placed in the watch list (see below). This includes the LAPPD, Argonne LAPPD, ABALONE, GDB-6203 and GDB-6201.

Conclusion

Several conclusions can be drawn from this extensive two-tiered selection of photosensors for LSNDs and similar detectors.

First, PMTs are very good photosensors. It is hard to find sensors which are on par for the specific combination of properties that is required for LSNDs.
Of course, this might in part be due to a confirmation/expectation bias that results from the design of LSNDs, which so far were based on PMTs and thus are adapted to their capabilities and limitations.

Second, only sensors which are based on vacuum phototubes managed to be placed in the shortlist, all of which are hybrids except for MCP-PMTs (and PMTs). Unlike semiconductor sensors, phototubes allow to reach large sensitive areas with low DCRs. And other than sensors with gas amplification, they can combine photocathodes that are sensitive in the visible range with long lifetimes and gains which are sufficient for spe detection.
Hybrid sensors, in turn, manage to combine the favorable properties of phototubes with those of the secondary sensors, such as good position resolution, low charge resolution, small TTS, or high gain (boosted by the often very large EB voltage).

Third, some LSND requirements are quite severe (e.g. DCR and TTS) and might have to be adapted to allow greater flexibility in the choice of sensors and series.

In conclusion, PMTs and HAPDs are currently the prime sensor type candidates for large LSNDs which are based on small time resolutions.
The LAPPD and ABALONE, and possibly also fountain-focused MCP-PMTs, will be of relevance in the future.

4.5.2 Promising developments

Apart from the shortlisted sensors which do not fulfill the LSND criteria yet but might in the future (LAPPDs, ABALONE and GDB-6201/6203), there are several sensors which were not added to the shortlist but might become serious candidates at a later time.

The watch list thus in sum includes the following sensors:

- LAPPD and Argonne LAPPD (p. 641)
- ABALONE (p. 661)
- GDB-6201 and GDB-6203 (p. 643)
- ICMOS (p. 649)
- TiPC (p. 650)
- SiPMT (p. 663)
- Timepix HPD (p. 674)
- VPGD (p. 701)
- Hybrid gas photomultiplier (p. 694)

Compared to the short list, besides vacuum phototubes also two gas-filled phototubes are included. Nevertheless, the watch list still contains only phototubes.
The prospects of the designs that were not already analyzed for the shortlist will now be discussed.

ICMOS

Typically, ICMOS sensors have a too high TTS. However, a promising concept study was carried out for a GaAsP cathode, an YSO phosphor and a megapixel CMOS. This estimated an achievable QE of over 40% from 400–600 nm, a gain of 10^6 which allows spe detection, a TTS of 100 ps, possible counting rates above 1 MHz, and 20 µm position resolution. Unfortunately, no follow-up publications could be found, but if this or a similar design could be realized, this would result in a very promising sensor.
However, certain issues would have to be addressed. The high gain could severely limit the dynamic range and the costs could be high due to the use of MCPs.
Alternatively, a Timepix4 CMOS could be combined with a fast phosphor and a cheaper MCP that is based on ALD and a glass capillary array. This would result in a sensor which is similar to an LAPPD, except for a better position resolution and a worse TTS (unless the MCP recharge current is used for timing).

TiPC

The TiPC foresees a gain of $6.5 \cdot 10^4$ (sufficient for photon counting), a TTS in the order of 1 ps (1.56 ns with Timepix3 readout chip) and a spatial resolution that is only limited by the pixel pitch.
So far, only a proof-of-principle prototype without photocathode, merely one tynodes, a low TSEY and a Timepix1 chip has been produced.
However, the high speed of tynodes make the TiPC a promising sensor – particularly if the Timepix4 could be used – although the TSEY is lower than the RSEY of dynodes and the charge resolution thus will be very high.

SiPMT

The SiPMT is explicitly designed for neutrino detectors and aims at a large sensor with low TTS (0.8 ns CTS) and a gain of 10^7 (first stage gain ≥ 10) with a HV of a few kV, small DCR, and low costs.
Because the development of this sensor was started just recently, only a demonstrator has been created and there are no prototypes so far.
Since the HS-SiPM layout has already shown very good properties in the ABALONE, the SiPMT should be watched as well.

Timepix HPD

Previous Timepix/Medipix HPDs were based on Timepix and Medipix2 CMOS chips, whose TTS was not sufficient; a Photopix HPD with better properties was planned but not realized.
The development of an HPD that is based on the newest Timepix4 chip (once finished) with bump-bonded Si diodes should be very promising, however.

VPGD

VPGDs promise an almost arbitrarily large sensitive area, very low spatial resolution, good TTS, high dynamic range, high magnetic field tolerance and very low costs. However, as of now, the IBF is very problematic and no VPGD with high gain, QE, and CE as well as low IBF and long lifetime could be constructed. Since a very high OC would be possible, this could compensate a low QE.
If a VPGD with all target properties can be realized, it would immediately enter the final selection or even become the favored sensor for a large scope of applications, and thus the development of the VPGD should be followed closely. In the last decades, great steps towards this goal have been made, and it remains to be seen whether or when it can be reached.

Hybrid gas photomultiplier

Finally, the separation of vacuum and gas phase in the hybrid gas photomultiplier promises to achieve a VPGD-like sensor without the trouble of having to eliminate IBF to the photocathode, which affects all properties in a VPGD. In addition, the QE is not reduced by pe back diffusion and the first stage gain is high, which should provide a very low charge resolution.
While the development was stopped without working prototypes, the problems which were encountered were nothing that has not been solved for PGDs or HPDs before, and with a few improvements this sensor type might well win the race for a VPGD-like sensor.

Conclusions

Several developments – LAPPDs, ICMOS, Timepix4 HPD, TiPC, VPGD and hybrid gas photomultiplier – promise to permanently change the photodetection in APP and beyond. This is because these sensors might allow to combine a low position resolution with a small TTS, high gain, large sensitive area and low DCR [273]. The future will show which new detection concepts and measurements can be achieved with the additional information provided by such sensors.

[273] Except for a small spatial resolution and a low TTS, all of this is already the case for large PMTs.

4.5.3 Possible developments

This chapter shall conclude by casting a glance at possible future sensor developments which might prove worthwhile.
A few promising new sensor subtypes and designs were already presented in the last section and shall be revisited where appropriate.

The wish list of sensor properties is long.
For LSNDs, the most critical exclusion criteria for short list sensors [274] were the PDE (with all its components QE, CE, FF, ME, TE), the DCR and the TTS.
Furthermore, future sensors should also display unambiguous photon counting, because this would greatly improve event reconstruction, especially for low energy events.
However, apart from optimizing currently required characteristics, the next logical step is the combination with *additional* capabilities, such as a good position or direction resolution, since this will introduce new possibilities which right now are not even conceivable.
In addition to the development of photosensors, auxiliary equipment can also contribute to the improvement or extension of properties. For example, a direction-sensitive LC (see p. 614) allows to convert a position resolution into a direction sensitivity [275].

While it would be desirable to take advantage of the high QE which is provided by semiconductors, all promising sensors so far are based on the *external* photoeffect. This is because in internal photoeffect sensors a tradeoff between spe detection, DCR and TTS is necessary.
The focus should therefore lie on phototube-based sensors. Experience has shown that here a tradeoff between high QE and DCR is necessary. The double cathode effect can probably only be used in a limited manner to increase the QE, since photocathodes that cover large solid angles show a large CTS. This is evidenced by the substantially larger total TTS of the SMART and X-HPD compared to the QUASAR-370.

Regarding the focusing type, proximity focusing allows to achieve a supreme position resolution, an excellent TTS, and a high magnetic field tolerance.
Fountain focusing, in turn, enables large sensitive areas and provides an intrinsic partial direction sensitivity for position-sensitive sensors. This comes at the price of a larger TTS due to CTS and longer transit times.
Cross focusing might be used as a compromise, since it has a better TTS than fountain focusing, a good spatial resolution and allows larger areas than proximity focusing. However, the layout of the electron optics prohibits cathodes with large solid angles.
Since spe detection is mandatory, multiple gain steps are beneficial, because this allows to use lower voltages and external preamplifiers.
For proximity- or cross-focused designs, a readout which preserves the full position information should be chosen. This can be either XS, XDL, double readout strips, resistive charge division, pixels/pads, or a fast CMOS chip, preferably with in-cell

[274] Apart from practical concerns.
[275] Although at the cost of a lower effective magnification than achievable with non-directional LCs.

TDCs. For a fixed position resolution, larger sensor areas are cheaper to realize with strip-based readout schemes than with pixels, since then the number of channel only scales linearly with the sensor size.

Promising known sensor classes and types

Based on the preceding sections, several sensor classes and groups can be recommended for future development regarding LSNDs (see also p. 714).

Proximity focused sensors have been conceived for discrete dynodes, MCPs, tynodes, HPDs, HAPDs and PGDs, but many of the possible combinations of these sensors with 2D-capable readout layouts have not been realized yet.
MaPMTs might be used with a CMOS, XDL or striplines.
MCPs, in turn, could be read out with resistive charge division (although it is questionable if this will improve the properties).
Further, tynodes, HPDs and HAPDs could also be combined with XS, XDL or strips.
Whether any of these amplification–readout pairs could improve the performance versus existing combinations would require a thorough analysis or simulations, though.

PMTs still continue to improve despite the venerable age of the original concept, and nearly as many dynode types as GD layouts have been developed. Still further improvements might be possible, e.g. the aforementioned readout variants.
In addition, a cross-focusing MaPMT with mesh or metal channel dynodes might allow to obtain large sensitive areas with good position sensitivity.

For MCP-PMTs, the variants with ALD-GCA-MCPs should show the most promise. Proximity-focused designs provide excellent time and good position resolution, while fountain focusing yields a higher PDE and possibly a lower price.

HS-PMTs are a very promising class, and the QUASAR-370 likely only would need a modern bialkali cathode (higher QE and lower DCR) and a manufacturer to constitute a serious candidate. A ZnO:Ga scintillator would further improve many properties significantly, such as the TTS, LP and EP probability, EAP rate, pulse shape, and charge resolution.
The same holds for HS-SiPMs like the ABALONE and SiPMT. Furthermore, an imaging HS-SiPM (as the ArcaLux or panel-ReFerence concepts) would be interesting and should be possible with regard to DCR background (see eq. (4.8)).

HAPDs already now provide some true PMT alternatives, and this class shows promise of further improvements, in particular regarding large position-sensitive sensors.
The 5-ch version of the R12850 only has a very coarse spatial resolution but this already might allow improvements in reconstruction algorithms, also since the hemispherical shape partially correlates the incidence position and direction.
The Belle II HAPD already had a good position resolution thanks to the proximity focusing; however, the FF still would need to be improved.
Finally, the development of the HPSAPD might be pursued with the goal to implement a bialkali cathode and increase the size, if this is achievable with a small TTS and DCR.

Hybrid sensors with CMOS readout and preceding amplification stages should generally show good properties when using the Timepix4 chip. This includes the IEBCMOS (best with ALD-GCA-MCPs), ICMOS, HPD, and EBCMOS, assuming spe detection can be achieved for the latter.

As already sufficiently addressed above, the development of a working VPGD type or an improved position-sensitive variant of the hybrid-gas photomultiplier both would be very desirable.

Novel sensor classes and types

By using the photosensor tree in fig. 4.1, one can additionally have a look at what combinations are *missing* so far, which is eased through the grouping by physical interactions.
Many branches and leafs are still open for exploration, but for some regions one can already briefly assess whether they would merit the journey.

Position sensitive SiPM

A position sensitive SiPM could likely be realized via resistive charge division, but this sensor would be limited by a dynamic range of 1 pe (or slightly more). The parallel connection of the SPADs would introduce a large capacitance and deteriorate the pulse shape.
Thus, it would be better to just use a SPADA, which also provides position information for a dynamic range of 1 pe with unaltered pulse shape, or to develop a position sensitive dSiPM, which would have none of these disadvantages but a lower FF.

Position sensitive SPAD

Position sensitivity for a single (large) SPAD might be possible by resistive division of the avalanche charge, by using the arrival times of the avalanche on different output pins, or with both. However, this concept either is limited to very small devices (where a position resolution is of little use) or likely will suffer from longer pulse shapes and a larger TTS if the sensitive area is increased.

Hybrid sensors with dSiPMs and SPADAs

In hybrids with SiPMs (HS-SiPM, HSiPM), the secondary sensor could be replaced by dSiPMs or SPADAs. The digital cell output might improve the charge resolution, an in-cell or in-chip TDC could lower the TTS, and position sensitivity could be implemented, but the FF is lower. For hybrids, however, the TTS of the secondary sensor is not the limiting factor, so no large improvements can be expected. For fountain focusing, furthermore a position sensitivity of a single secondary sensor likely could be used, since its area is fully needed to cover the photoelectron spot size, whereas its lower FF would reduce the PDE. Here, only a sensor array might allow to achieve a coarse position resolution, but this could also be realized with HSiPMs. For a hybrid dSiPM/SPADA, a similarly high DCR as for HSiPMs is expected even when using fountain focusing, which excludes these combinations.

A hybrid scintillator dSiPM/SPADA could only achieve a good position resolution for proximity focusing. However, for LSNDs likely no better spatial resolution than the dimensions of a single secondary sensor (of an array) is needed, which renders a position sensitivity unnecessary.

Dark count rate versus single photon detection in hybrids

For any proximity focused hybrid sensor with SiPMs, dSiPMs or SPADAs, the extremely high DCR of the secondary sensor limits the sensitive area of the device. The achievable area depends on the gain which occurs before the secondary sensor (e.g. the dpy in a scintillator). The sum charge of the dark noise pulses of the secondary sensor in an acquisition window already reaches the level of the spe pulses for relatively small areas.
The area with an S/N of 1 can be calculated from Poisson statistics using eq. (2.11), where the mean λ is the product of the DCR f_d and window length Δt. For large values of λ, the Poisson distribution resembles a Gaussian, and for the considerations here it is thus sufficient to simply compare λ with the primary gain δ_p before the secondary sensor. Then the area A which is achievable with an S/N of 1 amounts to

$$A = \frac{\delta_p}{\lambda} = \frac{\delta_p}{f_d \cdot \Delta t} \tag{4.8}$$

For fountain or cross focusing, the achievable area can be calculated by multiplying the right side of eq. (4.8) with the demagnification factor. To reach a target S/N that is different from 1, the right side should be divided by the S/N.
For example, in a proximity-focused hybrid scintillator dSiPM with an dpy of 100 (achieved by the ABALONE), a typical DCR of $10^5\,\mathrm{Hz/mm^2}$, and a 100 ns trigger window, a sensor diameter of 11.3 cm would result in as many dark noise pulses in the window as the dpy.
When the dark noise charge equals the spe charge, the higher spe pulse height could still be used to detect photons without running into dark noise. However, then significant contaminations of the SER can be expected, which will increase the charge resolution. Furthermore, the TTS will be larger, since pileup modifies the rising flank, and the dynamic range will also be affected at some point. On the other hand, for sufficiently high numbers of DCR pulses, these become a statistic contribution, which can be reliably subtracted.
Due to these different effects, it is difficult to assess what the achievable properties of such a sensor would be. In any case, it can be concluded that high primary gains are required to avoid strong statistic fluctuations of the background, which would significantly deteriorate the performance for e.g. a gain of around 20.

Hybrid sensors with micro-channel plates

In the case of HM-PDs (IEBCCD, IEBCMOS), the MCPs could be combined with arrays of APDs or SiPMs instead of PDs to achieve a very low TTS.
For SiPM arrays (HM-SiPM), a single MCP with low gain would be sufficient, which would allow to achieve a low IBF and a good charge resolution. The SiPMs could also be operated with a low gain to reduce the DCR, crosstalk and EAP rate while still obtaining large total gains. For proximity-focused devices, the DCR of the secondary sensors (assuming $10^5\,\mathrm{Hz/mm^2}$ and a 100 ns window) reaches the gain of a single MCP (about 10^3–10^4) only for a quite large diameter of ca. 36 cm. In addition, one

could always use a chevron stack to raise the primary gain, so this should not be a concern.
Sensors with APD arrays (HM-APDs) would have large charge resolutions, since the APD increases the already not very low value of an MCP-PMT. Such sensors likely would be inferior to the SiPM array variant.

For hybrids with MCPs, a scintillator, and PDs (ICCD, ICMOS), the secondary sensors could be replaced by APDs (HMS-APD) or SiPMs (HMS-SiPM). This would allow to achieve a very good TTS (possibly limited by the scintillator) and extremely high gains, but the layout is quite complex.

Hybrid sensors with tynodes

The TiPC is foreseen with a CMOS readout. If instead a (d)SiPM or SPAD array were used, the primary gain from several tynodes could be even higher than for an HS-SiPM, but due to the amplification statistics of the tynodes, the charge resolution would be far worse nonetheless. In such a sensor, one would have to make certain that multiple cells are triggered by SEs to avoid large DCRs as in HSiPMs. Otherwise, the gain would also be identical to that of the secondary sensor. In addition, one could use fountain focusing as in the ABALONE to reduce the background at the price of a higher TTS.
An APD array as readout for tynodes would produce a truly bad charge resolution due to the low tynode gain and the unfavorable amplification statistics in the APD; likely no npe peaks would be resolvable. On the other hand, the DCR should be unproblematic and the areas could be made quite large if the tynode structure allows this.

Hybrid sensors with gaseous detectors

The combination of a GD with MCPs or tynodes would be similar to a hybrid-gas photomultiplier (GD and EB gain) with extra gain stages. However, since the gain of a working hybrid-gas PMT would already be sufficient, this is unnecessary and would needlessly complicate the layout.

Hybrid sensors with scintillators

All hybrids with a scintillator are limited by its emission time and require very fast scintillators with high dpy. The effect of a long-term bombardment of keV electrons that are focused from a large photocathode on a small scintillator crystal or phosphor should also be studied.
Hybrid sensors which incorporate an EB scintillator do not preserve the position information unless proximity focusing and a thin phosphor layer on top of an imaging secondary sensor are used. For fountain focused devices, thus a hybrid scintillator CMOS is not reasonable.

When combining an EB scintillator with diode sensors, PD variants will not be able to detect single photons due to the low primary gain (the dpy). APD sensors might barely reach spe sensitivity but will face similar problems as HPDs regarding the compromise between a low TTS and a high TE. Hybrid scintillator SiPMs should be superior to both classes, which is why so far only this variant has been realized.

Use of an EB scintillator with MCPs would require proximity focusing to obtain an imaging sensor. Due to the small distance between cathode and scintillator, the HV could not be made very large, which would limit the dpy. It is questionable, whether this could achieve characteristics which are as good as when omitting the scintillator altogether, i.e. for a large MCP-PMT like the LAPPD.

Novel amplification structures

Finally, two new amplification structures shall be introduced and will be briefly discussed regarding their suitability.

Tynode channel plates

One problem of MCPs is that the SE emission angle strongly affects the distance that is covered in the pore before impact and amplification. This impairs the charge resolution and is the reason for the large optimal length–diameter ratio (L/D) of MCP channels.

Horizontal tynodes

This might be solved by including one or more tynodes in the channels. Since all SEs have to pass the tynodes, hereby additional multiplications are introduced at defined locations, which could lower the charge resolution and might raise the gain. In addition, it would prevent IBF, though no tynode should be placed at the pore entrance to avoid shielding the field lines from the pore interior, which would lower the CE.

Several concerns, however, might prevent the practical realization of such **tynode channel plates** (TCPs).
First, the tynodes would have to be extremely thin, because with lower thickness the maximum TEY increases and the voltage for which it is reached decreases [518; pp. 83–84]. So far, for the best material (5 nm thick ALD MgO) a TEY of 5.5 was reached at 1.2 keV and the TEY exceeded 1 above ca. 600 eV. Thus, for usable TEY values, higher voltages than in MCPs are required and the channel diameter will likely have to be increased to allow the emitted electrons to gain higher kinetic energies. At this point, the construction might not be very different from a stack of tynodes. Simulations would have to show whether a better charge resolution than in MCPs can be obtained in this way at the price of higher voltages.
To avoid charge buildup, the later tynodes should also include metal films for recharging, which lowers the TEY.
Furthermore, aging is a concern, since the last tynodes might quickly degrade from the high SE fluxes.
Finally, a system of MCP channels that contain tynodes is complicated to produce, since the outermost tynodes block access to the interior, which is required for etching and coating of the channel and the inner tynodes. Likely, the easiest method of production would be to stack multiple substrates which each have one tynode at the bottom.

Vertical tynodes

A possible variant of the TCP would be to use one vertical tynode that divides each channel in half instead of horizontal tynodes. In this manner, both the REY and TEY (in the other half of the channel) of the tynode could be used to boost the gain. However, for the favored TEY material (5 nm ALD MgO) the SEY reached its peak value of ≈ 3.25 already at 350 V and showed a lower value of 2.1 at 1200 V (after annealing: peak SEY 4.7 at 350 V, 3.4 at 1200 V) [518; pp. 83, 93]. Further, the SEY decreases with a lower tynode thickness, since the interaction volume which would occur in a bulk material is only partially contained in a thin film. Thus, there will be an optimum thickness and voltage that maximize the sum of REY and TEY.
For 5 nm thick ALD MgO, the SEY curve still shows a value of 2.7 (annealed 3.1) at the lowest depicted voltage of 150 V in [518]. Therefore, with this material similar reflective gains as in MCPs should be achievable for comparable voltages of around 1 kV per stage.
For a noticeable TEY, however, far higher voltages are necessary, at which the SEY is smaller. This likely will requires to use a far lower L/D, since otherwise the kinetic energy of the electrons will not suffice to penetrate the tynode. The latter then would effectively just reduce the channel cross section.
Again, simulations will have to show if this layout can provide benefits over regular MCPs.

Cynodes

It is also possible to solely use transmissive dynodes which exploit both the REY and SEY to construct an amplification chain. Since this concept *combines* the properties of a dynode and a tynode it is called **cynode** (cn for n-th cynode) here. The gain of a cynode is the sum of REY and SEY and is dubbed combined electron yield (CEY).

A possible layout of a cynode stack would be to tilt the first stage by 45° to the cathode. The second stage is placed below [276] c1 and is arranged perpendicular [277] to c1 so that the SEs from both REY and TEY are collected on c2. Subsequent cynodes are placed below the previous ones and perpendicular to them, which results in a staggered herringbone pattern.
Between the stages, cross-focusing should be employed through a focusing electrode to reduce the interstage transit time variation which would result from different impact points of the REY and TEY electrons. Otherwise, the TEY electrons would impinge on the next cynode at a point which is closer to the stage after next. This would escalate after a few stages and would cause very broad pulse shapes. The distance between the focusing electrode and the next cynode is a compromise between a small spot size (small distance) and a lower interference with the fields that accelerate REY electrons from the next stage to the one after that. For shorter distances, the REY electrons are increasingly pushed away from the cathode, which introduces an asymmetry in the spatial impact distribution and increases the transit time variations.
The cynodes themselves have to be only a few nm thin to provide a notable TEY. They thus will be very fragile, so a robust holder is needed. At the same time,

[276] Further from the cathode.
[277] The axis of symmetry of c2 lies on the extension of the plane of c1.

the rims of the holder should be as thin as possible to reduce their interference with the focusing fields.
Since the recharge currents are only very small for the first cynodes, these should not be coated with a metal layer to maximize the CEY and thus improve the charge resolution. For the last stages, regular dynodes or tynodes could be used to avoid space charge buildup.
The angle of c1 to the cathode and between later cynodes[278] can be varied to control the amount of REY vs. TEY electrons. For glancing incidence, the REY is enhanced (cf. p. 71), while the TEY should be reduced. The optimal angle for a maximum CEY and a minimum TTS should be determined by simulations. For very large interstage angles close to 180°, the REY will dominate and the cynodes basically operate in dynode mode, while for very small angles near 0°, the cynode shields its emitted REY electrons from the next stage and only the TEY electrons can proceed (tynode mode).
Based on the data for 5 nm thick MgO from above [518], there will be an optimum thickness and HV for the maximum CEY. This would have to be determined numerically, but in any case the required interstage voltage will be far higher than for dynodes.

A phototube with such a cynode stack could be called cynode multiplier tube (CMT) or cynode phototube (CPT) and some of its potential properties can already be estimated.
It is unlikely that the CEY of a cynode can exceed the SEY of good dynodes (even without considering GaP), compare p. 72. However, it is probable that the CEY can be higher than the TEY of tynodes. For 5 nm ALD MgO at 1200 V, a CEY of 8.9 should be achievable. When assuming this value, with 6 cynodes a gain of $5 \cdot 10^5$ could be reached for 7.2 kV voltage ($4 \cdot 10^7$ with 8 cynodes and 9.6 kV). According to eq. 2.65 the charge resolution then would be about 36% (R_σ) or 84% (R_F), which is rather high and comparable to PMTs.
The size of tynodes and thus of cynodes is limited to a few mm or less (see p. 582. Because the interstage distances will likely be larger than in the TiPC, the TTS will probably be higher. Since for the TiPC the TTS before readout is expected to be as low as 1 ps, the time resolution of the CMT could still be excellent, however.
Due to the small maximum size of cynodes, multiple cynode stacks can be placed next to each other to realize a position sensitive device through proximity focusing and 2D readout (e.g. CMOS, pixels, strips, or XS). To minimize losses due to the inevitable FF, focusing electrodes should be placed in front of the first cynodes. This layout is similar to metal channel or mesh dynodes and should allow an excellent TTS for sub-mm position resolutions.
Alternatively, cross or fountain focusing from the cathode to c1 could be used. Cross focusing would still provide a very good TTS, while fountain-focusing would allow larger areas. In both cases, it might still be a good idea to use multiple cynode chains next to each other to account for finite pe spot sizes. Even for the ABALONE, which achieves nearly 100% CE for a 6×6 mm² scintillator platelet, the target size is larger than possible cynode sizes, and here very large acceleration voltages of 20 kV were used.
A HQE bialkali photocathode should be used, but the FF might constitute the limiting factor for the achievable PDE.

[278] Measured between the surface vectors of the impact sides.

In conclusion, the CMT might be a worthy future endeavor. While many properties should be similar to the (excellent) targeted characteristics of the TiPC, as far as can be appraised from this short investigation, the CMT might have a slightly better charge resolution. However, it could be more difficult or expensive to arrange the mm-sized cynodes as described above.

Recommendations

In summary, the following photosensor types or upgrades seem most promising and are proposed for future development (the wish list):

- **MCP-PMT:** Large proximity focused variants with ALD-GCA-MCPs and 2D readout are most auspicious.
- **HAPD:** This class shows very good properties and new large image or point sensors can be true PMT alternatives.
- **HS-SiPM:** Existing projects should be watched and new developments would be promising, especially if they aim to construct imaging sensors.
- **HS-PMT:** A remake of a QUASAR-370-like sensor with new scintillator types and possibly position sensitivity (by using an MaPMT for readout) should show very good properties.
- **Timepix4-based hybrids:** This refers to IEBCMOS, ICMOS, TiPC, HPD, and possibly EBCMOS.
- **Hybrid MCP SiPM (HM-SiPM):** This is a new, promising sensor class, whose complete potential still has to assessed more thoroughly.
- **Hybrid MCP scintillator APD / SiPM (HMS-APD / HMS-SiPM):** These are proposed novel classes with good predicted properties but a rather complex layout. It is recommended to analyze the expectable properties and costs more closely.
- **Hybrid tynode SiPM (HT-SiPM):** Another possible hybrid sensor class, which is based on tynodes and whose prospects should be studied in more detail.
- **VPGD:** Working sensors with all target properties could revolutionize photodetection.
- **Hybrid-gas photomultiplier:** An improved remake with position-sensitivity might allow to produce a VPGD-like sensor.
- **Tynode channel plate (TCP):** This amplification concept might warrant further studies.
- **Cynode multiplier tube (CMT):** Cynodes and sensors based on them could show excellent properties and might prove to be viable TiPC alternatives. The feasibility and attainable performance of this layout should be investigated.
- **PMT:** Last but not least, PMTs continue to be actively developed and most likely can be further improved still.

5 Optical module

5.1 Objectives

Photosensors are a crucial component for every liquid scintillator neutrino detector (LSND) and have to fulfill a multitude of critical requirements as was detailed in 1.4. Next-generation LSNDs with multi-kiloton masses pose even more severe demands (cf. table 1.1).

Photomultipliers (PMTs) so far have been the standard photosensors in LSNDs. Their properties were studied in chs. 2 and 3, while the suitability of other sensors for this detector class was analyzed in ch. 4. This showed that for LSNDs such as LENA (Low Energy Neutrino Astronomy), PMTs are still the sensors of choice, together with the newest hybrid avalanche photodiode (HAPD) series. Since of these sensors only PMTs can be mass-produced so far, they continue to be the primary sensor candidate for the time being.

However, even for the best available PMT series and other sensor types, some LSND requirements are not fulfilled.
There are two possibilities to resolve these deficiencies.

The obvious approach is to develop new sensor types or improve existing designs. This path is pursued by e.g. the JUNO (Jiangmen Underground Neutrino Observatory) and Hyper-K (Hyper-Kamiokande) experiments.

Nevertheless, even for the best sensors, some demands of future large LSNDs might not be satisfiable without considerable effort, for example pressure compliance. This suggests to instead enhance sensor properties through accessories. This is only possible for certain properties – e.g. the transit time spread can not be lowered in this way – and other characteristics can only be improved through sensor development. As it turns out, however, for all crucial unfulfilled properties of PMTs and HAPDs, the compliance with next-generation LSND requirements can be achieved through additional components.
These structures are often bundled to an optical module (OM), which is an assembly of a photosensor and the components which are required to enable and enhance its functionality. The elements of an OM are usually contained within a housing.

It is therefore the subject of this final chapter to design an OM for the inner detector (ID) of a large LSND using the example of LENA (see 1.3.1).
For this purpose, it has to be studied which OM components are needed for optimal PMT performance, what properties can be achieved with them, and how these elements can be combined to a functional OM.

Requirements

There are four requirements for PMTs in LENA which can only be fulfilled by means of accessories: pressure tolerance, exposure to water, elimination of the inner vessel and magnetic field tolerance.
In addition, while not strictly necessary for operation, the detector and sensor characteristics can be greatly improved by placing light concentrators (LCs) in front of the PMTs.
These five OM demands shall be explained in the following.

Pressure

All photosensors which are currently shortlisted for LSNDs and all promising photosensor developments are based on vacuum phototubes [1] (see 4.5). This introduces a crucial problem: In a large LSND – particularly if positioned upright – the pressure that weighs on the photosensors at the detector bottom can be quite significant and may exceed the compressive strength of the sensor.

In a large water Cherenkov detector (WCD) like Hyper-K, the newest PMT and HAPD series (Hamamatsu [2] R12860 and R12850), which are specifically designed for high pressure tolerance, can still (barely) fulfill the pressure requirements. This is due to the larger attenuation length of the liquid for this detector class, which allows a compact cylindrical detector shape.
For the large LSND JUNO, the spherical symmetry reduces the hydrostatic head.
For even bigger, upright cylindrical LSNDs like LENA and possibly THEIA, however, the PMTs/HAPDs have to be protected by a pressure housing. For e.g. LENA a maximum peak pressure on the lowest sensors of 1.3 MPa was estimated (see below), which cannot be fulfilled by any of the available series.
The other shortlist/watch list sensors have notably lower compressive strengths [3] and will require an encapsulation even in smaller detectors with more favorable geometries.

In case of the ID of LENA, the peak pressure is reached while the detector is filled with ultrapure water for cleaning, since the density of the scintillator is lower than for water. Thus, the expected maximum load on the lowest PMTs was calculated both for the water cleaning phase and for operation.
The liquid head in the detector plus the liquid handling leads to a peak pressure of 0.90 MPa during operation and 1.04 MPa during cleaning. Here a maximum height of the piping of 7 m above the tank was assumed, and the temperature-dependent density and the local gravity in Pyhäsalmi were considered.
The worst case atmospheric pressure was determined to be 0.134 MPa from the altitude above sea level, the depth below ground level, the operating temperature, the local gravity, the overpressure from the mine ventilation system (estimated at < 0.02MPa), an N_2 blanket of 1 kPa above the scintillator, and accounting for extreme weather.

[1] Except for VPGDs (see p. 701), which use gas-filled phototubes.
[2] Hamamatsu Photonics K. K.
[3] Especially flat proximity-focused layouts like the Large Area Picosecond PhotoDetector (LAPPD).

In sum, a maximum peak pressure of 1.05 MPa is expected during operation and 1.20 MPa during cleaning. With an additional safety margin, the pressure housing should be able to withstand a maximum load of 1.30 MPa.
Pressure fluctuations from a PMT implosion are not included in these figures and the shock resistance of a pressure housing will have to be determined separately with dynamic simulations and tests in a pressure tank.
In addition, a safety margin is needed to account for aging, manufacturing variations and faults in the material. This was considered in the finite element analysis (FEA) simulations reported in 5.4 by requiring a factor of safety [4] of 2.

For the outer detector, a different pressure housing design is required, since here no LC is included in the OM. Since the outer detector is filled with water, the pressure encapsulation has to be capable of withstanding peak pressures of about 1.20 MPa for the full operating time.

Protection from water

During construction, the inner detector is temporarily filled with ultrapure water for cleaning, so the pins and voltage divider (VD) of the PMT have to be protected from water.
Even if the PMT is housed, its pins and VD should be cast with polyurethane to prevent contact with water that might seep in, but it is sufficient to encapsulate and cast the region around the PMT stem and VD.

Buffer liquid

One key consideration for LSNDs is to surround the photosensors and their LCs with inactive buffer liquid.
Otherwise, the intrinsic radioactivity of the PMTs (see 2.6.2), voltage dividers, LCs, cables, scaffolding, etc. would produce background events with high dpy (detected photon yield [5]) in close vicinity to the sensors. These events would be detected primarily by only one or a few sensors. The effect is further aggravated by the focusing of the LCs and would raise the already critical background event rate to unsustainable numbers, as was shown by simulations [77].
For LSNDs it is thus necessary to surround the PMTs with buffer liquid.

In previous experiments (e.g. Borexino and Double Chooz) the solution was to include an inner vessel: a thin transparent container for the scintillator, which separates it from the inactive buffer liquid (outside), in which the PMTs are placed. This also acts as a barrier for the diffusion or convection of radioactive elements to the scintillator and fiducial volume (FV). These impurities are dissolved from the detector surfaces, such as the photosensors and tank walls.
The inner vessel can be a flexible membrane [6] or a rigid structure [7].

[4] Defined as material strength divided by applied load.
[5] The number of detected photons per energy deposited in the detector by a particle.
[6] Borexino used two concentric spherical nylon vessels with 125 µm thickness, the inner with 32 kg mass for a volume of 321.6 m^3 [80, 713].
[7] Double Chooz employed two nested 8 and 12 mm thick acrylic vessels with 350 and 1100 kg mass for 10.3 and 22.6 m^3 content [157].

However, the use of an inner vessel has several critical disadvantages for large cylindrical detectors, and the engineering and construction of a cylindrical vessel is demanding:

- **Load-bearing capacity:** For large cylindrical next-generation LSNDs, a rigid vessel is problematic, since it would have to be able to bear its own weight in an unfilled detector, which requires a very large wall thickness (e.g. acrylic is very brittle). While this solution was chosen for JUNO, its detector is smaller and spherical, and extensive mechanical studies were required; for e.g. LENA and THEIA, a rigid inner vessel likely would not be possible.
- **Buoyancy:** As buffer liquid typically either the scintillator without added wavelength-shifters (WLSs) or a non-scintillating mineral oil is used. This can cause a density mismatch with the scintillator *with* WLSs inside the vessel, and the buoyant forces may distort and strain the inner vessel, which could lead to rupture.
  For a flexible membrane, a support structure or ropes can be required to maintain the desired shape, and the mechanical strength of the membrane and its seams may not be exceeded. Despite this, the vessel may be prone to distortions and leaks, which introduces uncertainties in the shape and volume of the FV and thus can affect the event reconstruction and sensitivity for physics analyses.
  For a rigid hull, it is unavoidable to match the density of the liquids within and without, as was done for Double Chooz [159]. This can come at the cost of less-than-ideal optical properties.
  For both vessel types, the detector filling procedure is complex and critical, and it is questionable whether the vessel will not develop leaks over the course of several decades of detector operation.
- **Refractive index:** In addition, the refractive index of the vessel must match that of scintillator and buffer liquid, or reflections will occur, which reduces the dpy and complicates the event reconstruction.
- **Radiopurity:** The material of the vessel contains traces of radioactive backgrounds, which are in direct contact with the scintillator and thus can lower the FV. This is less crucial for a flexible vessel, since the mass per contained volume is far lower, and usually is not a large issue because the used materials are very radio-pure. For a massive rigid hull, however, the radiopurity of its material likely could *not* be neglected anymore.

Since it is questionable, whether the critical mechanical issues could be resolved satisfactorily, it was deemed best to eliminate the inner vessel entirely: Instead of a global buffer region for all OMs, the buffer liquid is included *in* the OMs, which then necessitates to also incorporate the LC in the OM. This has the added bonus that a dissolution of radioimpurities of the components inside the OM would be localized. Since now the OMs, cables and scaffolding are placed inside the active scintillator, convection must be avoided. This requires a temperature-stabilized detector, as was implemented successfully in Borexino in the last years [443, 714].
In addition, an intransparent light shield should be added between the entry apertures of the OMs to prevent that scintillation light which is caused by radioactive impurities of the OMs reaches the photosensors.

Magnetic field

The PMT has to be protected from the geomagnetic field (GMF) by a passive shielding (e.g. mu-metal or FINEMET).
Active compensation would require to surround the complete detector with coils which often is less favorable.

Light concentrator

Light concentrators focus light from an entrance aperture onto a smaller exit aperture. This magnification of the collection area of a photosensor comes at the price of a reduced viewing window, since the product of the spatial and angular extent of a bundle of rays is a conserved quantity (étendue) [8] [560].

Both the magnification and the reduced viewing angle lead to far-reaching consequences for the detector performance.
The effective magnification factor m allows to either obtain a fixed optical coverage (OC [9]) with less PMTs or to reach a larger OC and thus a greater mean dpy. The latter lowers the energy resolution and improves the vertex reconstruction, while the former reduces the radioactive backgrounds from the PMTs and lowers the construction costs and time. As was detailed in 1.4, background suppression is the most crucial requirement for a large LSND, for which reason a fixed OC is chosen. This leads to the following advantages of LCs:

- **Reduction of radioactive background:** LCs allow to move the PMTs further from the detector center without lowering the number of detected photons.
  The limited field of view also screens the PMTs from light that is emitted by radioactive particles in the scintillator close to the PMTs, which reduces the amount of background events that trigger the event acquisition.
  In addition, the magnification leads to a lower radioactive load per effective sensitive area. This is because the few materials which are necessary to build an LC can be selected with high radiopurity and have low density. Furthermore, with LCs less OMs and cabling are required and the scaffolding on which the OMs are mounted can be made lighter.
  All these effects allow to extend the FV and thus increase the effective target mass.

- **Improved pulse shape discrimination**: LCs were found to greatly enhance the alpha-beta discrimination due to a better dpy homogeneity and the suppression of scattered photons [77]. This also helps to lower the backgrounds.

- **Better homogeneity:** The dpy decreases towards the detector center due to absorption and scattering in the liquid. Calibration with radioactive sources can only partially eliminate this dependence due to the finite position resolution, especially for low energy events [77].
  With LCs, however, the dpy near the center is increased due to the magnification, while the limited viewing window reduces the values near the surfaces, since less photons with large incident angles are detected. This greatly improves

[8] Except for absorption and scattering – the étendue of a system can only decrease or at best stay constant.
[9] The fraction of detector surface that is instrumented with (effective) sensitive area.

the homogeneity of the dpy in the detector, which in turn lowers the energy resolution and possibly enhances the vertex reconstruction.

- **Lower dark count rate:** The magnification also lowers the dark count rate per effective sensor area. For LENA, m is only up to about 1.2–1.3 depending on PMT shape, however. Therefore, this decrease is small compared to the excess of the dark count rate limit which some PMT series or short list sensors show.

- **Improved vertex reconstruction** The reduced viewing window of each sensor provides additional information about the photon origin. This could be used in position reconstruction or tracking algorithms to improve the accuracy, provided that the scattering in the liquid is not too strong.

- **Organizational aspects:** The total detector costs are lower, since LCs are cheaper than PMTs per added sensitive area and less electronic channels are required. Furthermore, the construction time is shorter, because fewer PMTs have to be produced.

For some properties, it is unclear whether LCs have a favorable or adverse effects. The effective photodetection efficiency (PDE) might be increased by LCs through additional (reflected) photon trajectories which use the double cathode effect (p. 64). On the other hand, the photocathode hit pattern is altered by the LC, and it could be that the cathode brink – which has a smaller PDE – is favored, which would lower the PDE. Which effect prevails depends on the specific layout of PMT, LC and detector and has to be checked on a case-by-case basis through simulations and ideally measurements.
An LC might also have an effect on the early pulse, late pulse and early afterpulse rates, since some time shift effects depend on the position and angle of incidence (see 2.3.5 and 2.3.6). This effect is difficult to assess without a full electron-optical simulation of PMTs, so it would be easiest to study it experimentally with an LC prototype.

On the other hand, LCs have a few drawbacks as well: The magnification could lead to a larger required dynamic range, the reflections in the LC introduce slight transit time variations, which raise the effective transit time spread, and for fewer OMs the granularity is lower, which affects event reconstruction.
For a position-sensitive photosensor, a non-imaging LC also would destroy the position resolution (of e.g. a Cherenkov ring on an LAPPD). This is because not both the photon position and incidence angles are measured and thus the original photon trajectory cannot be reconstructed.

All in all, the advantages far outweigh the drawbacks, and the use of LCs greatly improves detector performance.

In summary, in order to allow the operation of PMTs (or any vacuum-tube-based sensor) in a LSND such as LENA, an optical module in form of a pressure reducing housing is necessary.
In the classic design, only the PMT and its base are housed – sometimes even only the lower part of the sensor – but in case of LENA it is favorable to include the buffer liquid in the OM, as well, to eliminate the inner vessel. This results in

an encapsulation, into which then also every other photosensor component (LC, magnetic shielding, voltage divider, ...) can be embedded.
Since the sensor is not operational without the OM, in a way the complete OM can be regarded as the photosensor.

Existing optical modules

A large number of (astro)particle physics experiments and projects have used or foreseen OMs to house PMTs for a multitude of reasons, and in nearly all cases for a water-Cherenkov detector:

- With glass spheres (Benthos-spheres) as OMs: DUMAND [715], Baikal neutrino telescope [502], AMANDA [311], NESTOR [716], IceCube [717], NEMO [718] and [449] (direction sensitive), ANTARES [719], and KM3NeT [629], as well as Hyper-Kamiokande [41; p. 129] for the optional multi-PMT module
- In non-spherical housings: Borexino (muon veto) [80], GERDA (muon veto, OM based on the Borexino design) [720], the LBNE water Cherenkov detector [185], and MEMPHYS [721]
- Housing only the PMT stem and voltage divider: CTF [722], SNO [162], LUX (housing extends up to bulb) [93], and Daya Bay [723]
- In two cases an OM was actually used for a deployment in scintillator: For KamLAND (only stem and voltage divider) [724] and Double Chooz (muon veto [10], Borexino OM with small modifications) [157; p. 56].

However, the existing OMs are not designed for and not easily adaptable to the demanding requirements of multi-kton LSNDs, namely the inclusion of a buffer liquid and LCs, high ambient pressure, and large photosensor sizes. Therefore a dedicated OM design is required for this detector class.

The target of this chapter thus is the development of a generic optical module for a large LSND using the example of LENA, which can be utilized or adapted for other experiments such as THEIA.

In order to give an overview, first the layout of the OM is discussed in 5.2. Afterwards, the development of the LC is presented (section 5.3). This is split into the optimization of the LC shape in 5.3.1 and measurements of the reflectance and compatibility with the buffer liquid for possible materials in 5.3.2. Following this, the design of a pressure housing which contains all OM parts and its optimization for minimum weight is discussed in 5.4. The chapter closes with a conclusion.

Due to the mutual dependencies of the individual components, the design of LC, pressure housing and the other OM components was an iterative process. The development of the OM was performed at an early stage of this work and in parallel to the PMT characterization (ch. 3) and the photosensor suitability analysis (ch. 4).

[10] Unlike in the other listed experiments, the muon veto detector in Double Chooz is not based on water but *scintillator*.

Since the final PMT candidate series were not clear (or existent) at this time, the LC and OM layout were based on the then favored series, the Hamamatsu R11780-HQE, which now is not the prime candidate anymore (see 3.3). However, the used methods and the results are in large part general and can be applied to other series, sensors and detectors.
The calculation of possible LC shapes, the LC optimization for dpy homogeneity, the LC layout, the material compatibility tests, as well as the design and the optimization of pressure housings were performed in collaboration with German Beischler and already are detailed elsewhere [725, 726].
Since the employed methods are particularized in these sources and the studied series are outdated, the results of these and subsequent works shall only be briefly summarized here.

5.2 Layout

The layout and the main components of an OM for the ID of LENA are shown in fig. 5.1.

The photomultiplier is the center piece of the OM, whose complete design revolves around it. Here, a Hamamatsu R11780-HQE is used.
The PMT pins are inserted in a socket (blue) that contains the voltage divider (green). Depending on the electronics layout, also front-end electronics such as a preamplifier [11], a power supply, or digitization can be included.
A reflective light concentrator, which consists of an acrylic glass substrate that is coated with Ag, is attached to the front window of the PMT. In the shown schematic an earlier version of the LC was used, the final version is longer (compare fig. 5.5b).
The complete photosensor is surrounded by a passive magnetic shielding, e.g. mu-metal or FINEMET (see p. 436), in form of a truncated cone. This shape can be created with spot welding, which minimizes the area that loses magnetic permeability due to heating.
All this is contained in a pressure housing that consists of a hemispherical acrylic glass window and an elliptic stainless steel housing. The window is attached to the steel part with a flange and screws, where gaskets and knife-edge seals, which cut into the acrylic glass, are used to tighten the OM.
The steel rear part has a flat base plate for the cable connector and to provide enough space to mount components. The flat section connects to a rotated ellipse [12], but this layout might have a slightly lower stability than a true ellipsoid.
For the voltage supply and signal output, a water-proof cable connector is set in the middle of the base plate, next to the VD.
The voltage divider is surrounded by a steel cylinder, into which the PMT socket is placed and that is attached to the steel housing with threaded rods, which are welded to the base plate. Due to a lack of space, it is not possible to directly weld the cylinder to the base plate. The cylinder has a flange to attach the holders for the magnetic shielding and LC.

[11] The cable length to the electronics on top of the detector can exceed 120 m.
[12] Transition with a continuous first derivative.

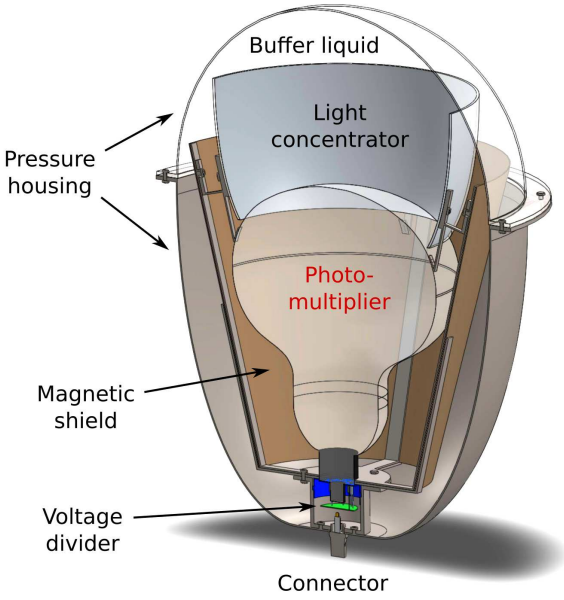


Figure 5.1: Layout of the optical module (OM). Description see text. Adapted from [725].

The cylinder is cast with polyurethane (not shown) and the region around the PMT pins is filled with silicone to prevent contact with seeped in water and to fixate the lower parts [727]. To avoid light emission from the OM, an intransparent polyurethane type which was tested for discharges should be used (see p. 417).
The rest of the housing is filled with buffer liquid (LAB [13], a mineral oil).
The liquid is inserted through a filling tube at the bottom (not shown), which afterwards is closed with silicone and e.g. a screw-cap.
The magnetic shielding and LC are fixated on the base cylinder with two holders in form of rings with four arms. These are placed outside and inside the magnetic shielding and fix it geometrically without requiring screw holes. Both holders are attached to the flange of the cylinder and also secure the PMT socket. The LC is attached (glue or screws) to extensions of the arms of the inner holder, which also fixate the PMT.

The assembly of the OM was considered and included in this design; the detailed assembly procedure can be found in [725; pp. 96–98].

For the outer detector, an optical module is required for pressure compatibility as well, but in order to maximize the viewing angle of the PMTs, here no LCs are employed, and thus a different OM design will be required.

It can be of advantage to group multiple OMs to an array which has a common front-end readout electronics and high voltage (HV) supply that are housed in a unit in the center of the array. [14]
In this way, less cabling is required (radiopurity), a digitization of the signal avoids a degradation in the cables, a prefabricated mechanical assembly can speed up the

[13] Linear alkylbenzene
[14] Paragraph based on [122, 123].

detector construction, and the electronics costs are significantly lower. Since the PMTs are regularly calibrated with laser pulses during operation, such an OM array can also include a splitter, from which a fiber leads to each OM. For a precise event reconstruction, the position and alignment of the OM arrays must be known to 10 mm and 2° accuracy (conservative estimate).
One such system was designed in the French PMm² program for an array of 4×4 PMTs, where a common application-specific integrated circuit (ASIC) called PARISROC is used for signal readout and HV distribution. This ASIC provides a high-speed discriminator, 12-bit charge and time digitization, gain adjustment for each channel, and a common HV and was tested with the MEMPHYNO prototype for the MEMPHYS [15] [728, 729] WCD. This system was also foreseen for LENA.

5.3 Light concentrator

5.3.1 Shape

Three targets should be fulfilled by an ideal LC shape: The mean dpy should be maximized, its position dependency (detector homogeneity) should be minimized, and the number of required sensors should be minimal (maximum effective magnification factor m) to reduce backgrounds and costs.
However, these goals typically cannot be reached at the same time – for example, to obtain the maximum mean dpy, the effective OC should be maximal which requires a large amount of sensors. This makes it necessary to determine which of the three targets is most crucial so that the LC shape can be optimized in this regard.

For LENA, the dpy that results from an OC of 30% with PMTs without LCs was found to be high enough to fulfill the physics agenda [77, 81]. In the present work, a dpy of 295/MeV was found for the R11780 without LCs and assuming a very conservative PDE of 20%. A value of 200/MeV is considered acceptable by [81]. After 30 a of operation, a dpy of 100/MeV is seen as still sufficient for operation, when taking into account the failure of PMTs and channels as well as the degradation of the LC reflectance and buffer liquid transparency [730]. For these reasons, it is sufficient to define a target mean dpy and optimize other quantities.
The background that was introduced by PMTs with non-optimized LCs was acceptable in all analyses as well. Thus, it is not necessary to minimize the amount of sensors at all costs either.
Therefore, the design goal of the LC shape for LENA was to minimize the detector inhomogeneities, which for a large cylindrical LSND limit the achievable energy resolution more severely than can be compensated by a maximized dpy and calibration.

For previous projects which used LCs (see next paragraph), the optimization of the homogeneity was not necessary, since the attenuation length of the liquid was large compared to the detector dimensions and/or the detector was spherical. This is very different for large LSNDs with cylindric shapes, however, as will be shown.

[15] MEgaton Mass PHYSics

Non-imaging light concentrators

Since large PMTs are point sensors and cannot detect the photon direction or incidence position, it is possible to employ non-imaging optics to construct light concentrators which maximize the light collection for PMTs .
For this reason, a great range of astroparticle physics experiments and projects have incorporated non-imaging concentrating collectors in their design, e.g. MACRO [731], SNO [732, 733], CTF [734], Borexino [80, 734], the LBNE WCD variant [185], MEMPHYS [729], FACT [735, 736], CTA [737–739] and KM3NeT [740, 741]. JUNO was discussing the use of LCs [742], and Hyper-Kamiokande is currently developing an LC as well [743].

Construction methods

There are several methods to construct a non-imaging LC, all of which use reflective or refractive surfaces, or both. [16]

Refractive materials, however, have substantial disadvantages for large LSNDs: Since not a surface but a solid is used to achieve the concentration, the mass and thus the amount of radioactive impurities is far greater. Furthermore, in liquids with already high refractive indices n (≈ 1.484 for LAB [81]), materials with extremely high values of n are required, which are either non-existent or possibly expensive, and whose compatibility with LAB is unknown. In addition, there is chromatic aberration. Finally, reflective surfaces achieve higher concentration ratios.
On the other hand, in dielectric materials the total internal reflection can be used to eliminate reflection losses. Refractive LCs also are shorter.
Since the disadvantages far outweigh the advantages, only reflective surfaces should by employed in LSNDs, and in fact all abovementioned detectors except FACT [17] have used reflective LCs.

Another important criterion is whether a construction method can produce an ideal concentrator, i.e. one whose concentration ratio is identical to the maximum allowed by the conservation of the étendue.

The string method (or edge-ray principle) is most relevant here. This construction method is based on the projection of all rays from the source [18] brink onto the sensor brink (extreme rays) with a single reflection on one surface. If this is the case, also all rays in between the extreme rays hit the sensor and an ideal LC results.
This technique produces ideal 2D concentrators, but for a LC that is suited for PMTs, a 3D shape has to be obtained by rotating [19] the calculated 2D profile (so-called compound concentrators). This introduces non-ideal behavior for non-meridional rays, i.e. trajectories which are not in the cross-section plane through the rotational axis. For some rays, now more than one reflection is necessary to reach the receiver [20],

[16] Content based on [560].
[17] Which is an imaging atmospheric Cherenkov telescope.
[18] The target volume of the detector.
[19] The 2D shape also can be translated perpendicular to the cross-section, which results in a trough concentrator with ideal properties (except for material imperfections). This, however, concentrates light only in one dimension, and has a shape which is unsuited for large hemispherical PMTs or HAPDs.
[20] The photosensor.

which leads to higher reflection losses. Other rays no longer reach the exit aperture at all but are reflected back towards the source, which lowers the concentration factor. There are also two subtypes of string method concentrators which use a refractive part in addition to the reflecting mirror and thus are not considered here.

The flow-line method treats the receiver as a *light source* and places reflective surfaces where the flux from both sides is identical (condition of detailed balance of rays). With this method, ideal 3D LCs *can* be constructed – but only for very special cases, which do not apply here.

The ideal 3D concentrators which are obtainable with the Poisson bracket method [21] are limited to special cases, gradient index refractive materials (which in addition to the above problems cannot be manufactured yet), or designs with infinitely many dielectric layers. This makes them unpractical for LSNDs as well.

Finally, the simultaneous multiple surfaces method uses two or more reflective/refractive surfaces which do not touch the receiver.
This, however, would result in large LC diameters for which the target OC most likely is not reachable, since the fraction of area that is used by the OM brink increases significantly with rising PMT diameter. Alternatively, the PMT would have to be mounted backwards, which would contradict one of the main reasons to use LCs: to move the PMT further away from the FV.

With no method ideal 3D LCs can be obtained for LSNDs so far, but the string method produces results for rotated 3D concentrators which are close to ideal concentration ratios. To date, there is no technique which promises noticeably better performance without downsides [22]. Therefore, the string method was used to calculate the LCs for LENA. The choice of this method is substantiated by the fact that it was employed by all abovementioned projects, except for one of three designs for CTA [23] and the LCs of MACRO and CTA, which did not use any of the above methods (simple geometric layouts).

Deficiencies

In addition to losses that are introduced by the rotation of a 2D design, the string method does neither consider reflection losses nor absorption and scattering in the buffer liquid inside the LC. Furthermore, it assumes a homogeneous light source, which does not account for absorption and scattering in the scintillator, which strongly influence the overall behavior (see below).
The theory is thus incomplete and the highest concentration factor and maximum mean dpy is not necessarily obtained with the LC which is constructed for the exact source and receiver geometry.
Moreover, the string method is designed to maximize the dpy but not to minimize its variation over the source, which actually is the target here.
Finally, for a cylindrical detector, it is not clear which source shape should be used to calculate the LC shape, since the detector cross sections that are visible from the OM range from a circle to a rectangle. Because the light is increasingly attenuated

[21] An alternative formulation of the flow-line method which uses Hamiltonian equations.
[22] String method LCs can also be derived with the flow-line and Poisson bracket method.
[23] Two reflective simultaneous multiple surfaces.

with the traversed distance in the scintillator, it is most reasonable to give more weight to the photons which are produced near the OM, however. Thus, the LC design should be based on the circular cross-section, which was used here.

Shape optimization

For these reasons, the optimal LC shape cannot be simply calculated based on the detector layout, but rather an optimization process is required. The LC shape must be varied to find the best design, which necessitates the introduction of one or more optimization variables.
Here two parameters are employed: On the one hand, the assumed shapes of source and sensor are varied (curved or flat, finite or infinite distance between both), which results in different subtypes of string method LCs. In addition, the field of view is modified by varying the assumed radius of the source (virtual source radius r_v).
These variables are optimized for the best detector homogeneity, and for combinations with comparable values, the one which maximizes the mean dpy is chosen.

String method concentrator types

The shape of rotated (compound) string method concentrators depends on the form of source and absorber and the distance between them.

The compound parabolic concentrator (CPC or Winston cone) [24] is designed for parallel incident light (i.e. a finite source in infinite distance, e.g. the sun) and a flat absorber [560].

The compound elliptic concentrator (CEC) [25] is constructed for a flat source in finite distance and a flat absorber and was used e.g. for the LBNE WCD [185] and MEMPHYS [729].
To construct the LC, here the circular source was reduced to its cross-section with a line that is perpendicular to the LC mirror axis. The LC shape then depends on the distance of this line to the OM. This can be chosen such that e.g. the solid angle which is spanned by the source is maximal (CECMA [26]) or that the extension of the source is maximal (CECME [27]; here the line goes through the circle origin). Since the viewing angle of the CECMA contains the whole source, whereas the CECME does not, the former is expected to show a better performance.
To introduce more degrees of freedom for the variation, alternatively the source could be defined as a line with arbitrary distance and length.

Furthermore, a concentrator which is based on the true source shape and distance and a flat absorber can be constructed. This variant is called circular–flat concentrator (CFC) here.

When the true geometry of both source (circular FV) and receiver (PMT cathode; can be approximated with a circular arc) is used, the resulting LC is simply called compound string concentrator (CSC) here. This type was used for example in

[24] This LC can be constructed by rotating a tilted parabola.
[25] Since this LC images one point (the brink of the source) onto another (the brink of the absorber), its shape is an ellipse section, where both boundary points are used as the focal points.
[26] Compound Elliptic Concentrator for Maximum incidence Angle
[27] Compound Elliptic Concentrator for Maximum Extension of light source

Borexino. Since the CSC uses the full detector geometry for its construction, it is expected to perform best.

The CPC, CECMA, and CFC have very similar shapes (CFC and CECMA are virtually identical). Compared to them, the CSC is wider and slightly longer, and the CECME has a very different shape (far longer than all others, wider than all except CSC).

To narrow down the selection, in a preliminary Monte Carlo (MC) simulation campaign, all five types (CPC, CECMA, CECME, CFC and CSC) were studied for three virtual source radii (9–13.5 m). As expected, the CECME showed inferior performance, while the resulting values for CECMA and CFC were identical within statistical errors. Therefore, the CECME and CFC were eliminated from the study.

Truncation

For LC types that are constructed for sources in finite distances, it can occur for small values of r_v that with growing distance from the PMT the LC shape first becomes wider and then narrower again.
In this case, the LC can be truncated at its maximum diameter to simplify the fabrication and save material. Simulations showed that this has a negligible influence on the performance or can even slightly improve it.

Simulations

Shape calculation

The LC shapes were calculated numerically with MuPAD/MATLAB [744] and were verified through a geometric construction with Geogebra [745] and by analytic calculations where possible.
The sensitive area of the PMT photocathode could in good approximation be described by a circular arc, since the cathodes are designed as spherical segments to minimize the cathode time spread (see p. 179).

Monte Carlo model

The existing detector MC of LENA is based on GEANT4 [746, 747] and ROOT [463, 464] and incorporates a full physical and optical model [77, 81, 748–754]. This framework was extended for LC simulations for this work.
On the detector walls, the OMs are distributed in a hexagonal grid. The distances between OMs in the same ring and between rings can be varied independently to achieve a closer match with the target OC, which results in a slightly distorted grid. On the end caps, the OMs are arranged in concentric rings. Since the number of the rings and the OMs per ring are integral number, the OC cannot be adjusted with arbitrary accuracy. Therefore, it was necessary to develop a complex algorithm to find the combination which most closely matches the target OC without introducing overlaps of adjacent OMs or too large distortions of the grid.
For the scintillator, LAB with 3 g/l PPO [28] and 20 mg/l Bis-MSB [29] was used. A light yield of 10^4/MeV and conservative values for the absorption length (20 m),

[28] 2,5-diphenyloxazole
[29] 1,4-Bis(2-methylstyryl)benzene

Rayleigh scattering length (40 m) and isotropic scattering length (67 m) were chosen. This results in a conservative attenuation length of 11.1 m. The JUNO collaboration achieved significantly higher values [39, 755].

Transmission

First, the transmission properties of the LCs were simulated.
This was performed by generating photons at the entry aperture and checking if they hit the cathode, are reflected or scattered back into the detector, or are absorbed.

The results agreed with theory and the data of other experiments, which served as a cross-check of the shape calculation and the MC implementation.

In a CPC with a flat cathode, for the meridional rays the distribution of the transmission vs. the incident angle (transmission curve) showed an abrupt drop from 1 to 0 at the acceptance angle θ_c, for which the CPC is designed. If rays from the complete entry aperture were used, the drop was blurred (some rays below θ_c are rejected, some above are accepted) due to non-ideal behavior.
When a CPC is used with a curved cathode, the drop in the transmission curve was shifted to larger angles, since the protruding cathode catches rays which otherwise would be rejected.
LCs that are designed for sources in finite distances showed a broader drop than CPCs, since here the source extends over an angular range. For these concentrators, only limited conclusions regarding the LC ideality can be drawn from the transmission curves.
If in addition to a finite distance the LC shape is also based on curved cathodes (CSC), the transition region became even wider.
It was found that a truncation of the LC could improve the transmission properties.
In LENA the source geometry is similar for the CPC and CEC. This led to comparable shapes and results of the transmission and homogeneity simulations (see below) for these LC types.
The losses which result from a finite reflectance (using the values measured in 5.3.2.2) amounted to 6–7%, while the attenuation (absorption and scattering) in the buffer liquid inside the LC removed about 2% of the rays.

Homogeneity optimization

In order to compare LC shapes, a figure of merit that measures the detector homogeneity is required.
The standard deviation σ of the dpy would be a logical choice. However, this quantity scales with the mean dpy μ, which depends on the OCs on the walls and caps, and the latter are never completely identical for different LCs. This would introduce additional uncertainties.
The coefficient of variation c_v, on the other hand, is a dimensionless quantity and is independent of the size of μ. It is defined as [756; p. 47]

$$c_v = \frac{\sigma}{\mu} \cdot 100\% \tag{5.1}$$

This variable thus simply expresses the standard deviation in percent of the mean dpy.

To obtain the systematic variation c_h of the dpy (called "homogeneity" here) the statistical fluctuations c_s have to be subtracted from c_v:

$$c_h = \sqrt{c_v^2 - c_s^2} = 100\% \cdot \sqrt{\frac{\sigma^2}{\mu^2} - \frac{1}{\mu \cdot E}} \tag{5.2}$$

where E is the event energy. This quantity is used to measure the detector homogeneity in the following.

Besides the well-known statistic errors, also systematic errors from variations in the OM arrangement can contribute to the total uncertainty and thus were estimated by means of simulations.
For two distinct OM arrangements (number of rings on the walls and caps, number of OMs per wall ring) with identical OC, the resulting dpy can still be different, because the average photon flux varies over the detector surface. This can introduce an error when comparing different LCs. For redistributions of the OMs on the walls, only upper limits could be found for this effect, while significant errors resulted when the number of rings on the caps was changed.
The resulting mean dpy varies between LC shapes due to different values of m and the obtained OC. This requires to scale the resulting dpy to a common value to be able to compare the homogeneity. The error which is introduced by this scaling was also determined by simulations.
Since the determination of these systematic errors is numerically costly, this was carried out only for a single combination of LC type, PMT and virtual source radius. The effect of a variation of these three parameters on the systematic error was estimated by scaling the error with fixed factors. The resulting total systematic errors are thus conservative estimates and their real values should be considerably smaller.
In addition, systematic errors which are established through simulations generally will be overestimated, since random statistical fluctuations are falsely attributed to systematics.

The homogeneity simulations were carried out for 1 MeV electrons which are randomly distributed over the detector volume, since this resembles the topology of low energy neutrino events.

Results

The resulting spatial distribution of the dpy in the detector for R11780-HQE PMTs is shown in figure 5.2 without LCs, with LCs optimized for homogeneity, and with an additional improved distribution of the OMs on the detector surface.

Due to the radial and mirror symmetries it is sufficient to examine the top right quadrant of a vertical cross-section through the detector center, and the simulated events in the whole detector were projected into this 2D section to increase statistics. To obtain the full cross-section, the shown distribution can be mirrored at the radial and height axes.
Since the event density in a cylindric shell is proportional to the radius, empty bins occur near the rotational axis (0 m radius) and the statistical fluctuations decrease with growing radius.

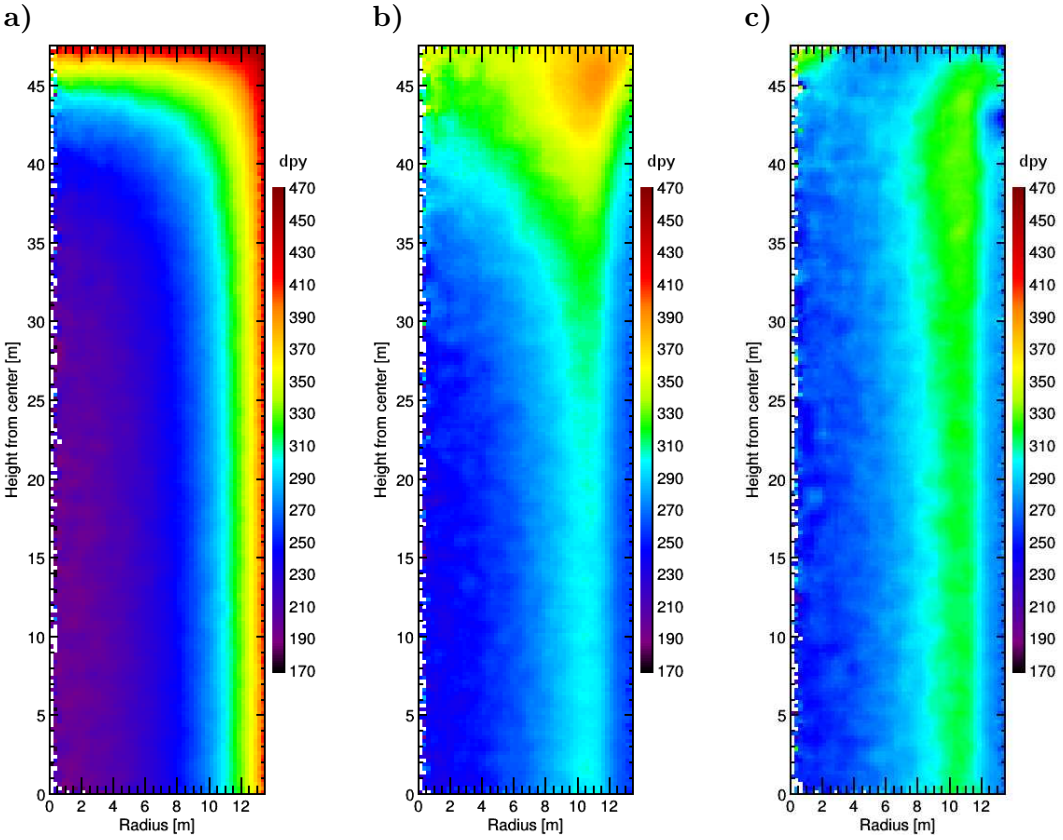


Figure 5.2: Spatial distribution of the detected photon yield (dpy) in the LENA detector using Hamamatsu R11780-HQE PMTs.
a) Without light concentrators (LCs).
b) With LCs optimized for detector homogeneity (compound string concentrator (CSC) with 12 m virtual source radius).
c) As b) with truncated LCs and an optimized OM distribution (omitted rings on walls and caps near the top right corner).
Note the successive decrease of the range of occurring dpy values (colors) by using optimized LCs and OM distributions. Discussion in text.
Due to symmetries only the top right quadrant of the detector cross-section is shown (compare fig. 1.1). Simulations for 1 MeV electrons which were randomly distributed in the detector volume. Number of simulated events: a) $1.2 \cdot 10^5$, b) $8 \cdot 10^5$, c) $6 \cdot 10^4$. Binning $0.25 \times 0.25\,\mathrm{m}^2$, empty bins colored white. The distributions were smoothed with a 2D Savitzky-Golay filter [757] with a 2^{nd} order polynomial and a 7×7 window (reduced near surfaces) to improve the visibility of fine structures.

Without light concentrators

Without LCs (fig. 5.2a), the dpy is lowest in the detector center, since scattering and absorption in the scintillator induce losses and the photons which are produced near the center have the longest paths to an OM. The number of detected photons increases strongly towards the surfaces and is highest in the corners, where both the OMs on the walls and the caps can detect photons and thus the solid angle that is spanned by LC apertures is largest.
The dpy ranges from about 170 to ca. 500/MeV with a mean value μ of $294.89^{+0.33}_{-0.24}$ (OC 30.16%). The homogeneity c_h amounts to $20.91 \pm 0.15\%$, which is significantly larger than the statistic error c_s of 5.82% for such events (1 MeV electrons).
Clearly, this arrangement is far from optimal, especially when considering that for analyses which are sensitive to background, the radius of the FV has to be lowered and the dpy thus is minimal where the remaining events occur. In other words, in this case the detector is optimized for the detection of background and not physical events.

With light concentrators

If LCs are added, two effects occur: First, the effective magnification m increases the amount of detected photons for events close to the center. Second, the limited field of view suppresses the collection of photons for events close to the surfaces, because only the LCs directly in front of the event can detect them [30]. This leads to an abrupt drop of the dpy close to the virtual source radius.
These effects can be seen by comparing figs. 5.2a and fig. 5.2b, where the latter shows the dpy distribution for the best LC shape regarding homogeneity, a CSC for $r_v = 12.0\,\text{m}$.
The combination of both effects decreases the asymmetry between detector center and walls, which now have nearly identical values, and introduces a high-dpy "ridge" (cylinder shell) at radii slightly below r_v. For very small r_v values, the dpy is maximal at the detector center. With rising virtual radius, a ridge is formed and moves outwards. And for large r_v, the maximum is close to the surface – in this case the LCs are very short and only slightly alter the behavior compared to without LCs. The ridge radius which is optimal for homogeneity is closer to the surface than to the center, because the volume of the cylindrical shell grows with radius and for large ridge radii the dpy drop to the walls and center is comparatively small.
The dpy still rises towards the end caps, since here a larger solid angle is covered by LCs, but far less than without LCs. As for the walls, the LCs also shift the region with high dpy inward near the caps. The absolute maximum of the dpy now occurs where the ridges from walls and caps meet.
The distributions in fig. 5.2 do not show the whole detector volume, which actually extends up to a radius of 14 m and a height of 48 m radius. In the outermost about 0.5 m strong local inhomogeneities occur, since events either lie directly in front of a single LC or not. In the latter case the light can only be detected by OMs on the opposing walls. Such large radii, however, show severe rates of background events and are only usable for physics analyses at higher energies.

In sum, the homogeneity can be notably improved by using LC shapes which are optimized for this. Interestingly, the achievable minimal values of c_h are nearly

[30] And those on the other side of the detector, but only after severe attenuation.

identical for the different LC variants which were studied here (CPC, CEC, CSC), although the r_v for which they are reached varies between types.
With a CPC, a homogeneity of $10.80^{+0.13}_{-0.09}\%$ is obtained for an r_v of 11.1 m, with a CEC $10.84^{+0.13}_{-0.09}\%$ at 11.0 m (the shapes of CPC and CEC are very similar), and with a CSC $10.80^{+0.10}_{-0.07}\%$ at 11.9 m. The values of μ, the effective OC[31], and c_s are nearly identical to those without LCs, since the number of OMs and their distribution was adjusted accordingly to improve comparability.

The magnification factor m rises with falling r_v and drops again at very low values. The maximum is reached for lower virtual radii than the optimal homogeneity, namely around 7 m for the CPC ($m \approx 1.21$) and CEC (about 1.205) and near 9 m for the CSC (ca. 1.245).
However, these LCs would be too large to reach an effective OC of 30%, because the LC diameter grows with falling r_v and the rising m cannot compensate the falling fill factor that is introduced by the insensitive OM brink. The lowest virtual radius which can reach 30% OC is 11.0 m for the CPC, 10.9 m for the CEC and 11.8 m for the CSC. From this, it can be seen that the variants with the best homogeneity barely fit. With homogeneity-optimized LCs the dpy thus cannot be increased much beyond 293/MeV, which corresponds to an effective OC of 30%.
For the variants with the lowest c_h, the obtained m was highest for the CSC (1.211, with truncation 1.213) and lower for the CPC (1.148) and CEC (1.146, with truncation 1.152).

The homogeneity which is obtainable with the three LC types is virtually identical; however, the CSC has a far larger magnification factor, which allows to minimize the number of PMTs. Therefore, the CSC for 12 m virtual source radius was chosen as the LC for LENA.
Considering that, unlike CPC and CEC, the CSC uses the full geometry of detector and PMT to calculate the LC shape, it is not surprising that this LC type shows the best overall performance. It came as a surprise, however, that all LC types can achieve similar homogeneities.
As discussed above, the lowest spread of the dpy is obtained for a *reduced* viewing window compared to the FV for which the LC is actually optimized (here 13.5 m radius and 47.5 m height).

These studies were also carried out for the ETE[32] D784 PMT (11" diameter), which has a more bulbous shape than the R11780 (12"). This showed qualitatively similar results.
Without LCs, here a mean dpy of 301.47 at 29.95% OC and a c_h of $21.95^{+0.16}_{-0.15}\%$ was found.
Again, the lowest reachable homogeneity was similar for all types, while the CSC had the highest magnification factor: The CPC with an r_v of 10.9 m showed a c_h of $10.77^{+0.13}_{-0.09}\%$ and an m of 1.160, the CEC with 10.6 m resulted in $10.76^{+0.13}_{-0.10}\%$ and 1.162, and the CSC with 12.1 m provided $10.70^{+0.13}_{-0.09}\%$ and 1.267.
However, the listed CPC and CEC variants are too large to reach 30% OC. The designs with the lowest r_v that can reach the target OC have higher homogeneities c_h and lower magnifications m. This clearly favors the CSC for this PMT series.

[31] OC without LCs that would be required to reach the same mean dpy.
[32] Electron Tubes Enterprises Ltd.

It was found that truncation of the optimized LCs [33] increases their magnification factor (by 0.002–0.006) while raising c_h minimally (about 0.05%). For the R11780 CSC, c_h was virtually unaltered, though. Thus, truncation is possible and favorable.

With light concentrators and optimized distribution

While the homogeneity is significantly improved by the addition of optimized LCs, there is still a notable increase of the dpy towards the caps. This can be improved by optimizing the arrangement of the OMs.
The vertical ridge, on the other hand, is caused by the interplay of the attenuation in the scintillator and the magnification of the LCs on the wall and cannot be smoothed by a redistribution of the OMs. It constitutes a physical limitation that is imposed by the scintillator properties and can only be improved by altering them. A change of the absorption and scattering lengths has a strong influence on the dpy distribution, as was verified in simulations, and for large attenuation lengths, the region near the center can be made far more homogeneous. However, then a different LC shape would be optimal and the c_h minimization campaign would have to be carried out anew.

To further reduce the variation of the dpy near the caps for the selected LC (truncated R11780 CSC for 12 m, see fig. 5.5b), first, the number of rings on the caps was reduced. The lowest c_h was found for 21 instead of 29 rings by interpolating a fit function; the lowest simulated value was $6.85^{+0.08}_{-0.34}\%$ for 20 rings.
In a second step, the distribution of the cap rings was optimized and several wall rings were omitted. The best results were obtained by using the original layout with 29 cap rings of which 8 were left empty, where the empty rings were distributed over medium and large radii (rings 8, 11, 14, 17, 20, 23, 26 and 29). In addition, two of 212 wall rings were omitted (ring 11 and 201, only the former is within the shown detector quadrant). This arrangement could reduce c_h slightly further to $6.57^{+0.08}_{-0.24}\%$.
The homogeneity can likely be lowered a little more by adjusting the distances between rows (here identical for all) and testing different ring omission schemes, which provides many parameters for an optimization. However, the vertical ridge imposes a hard limit, which will not allow to obtain much lower c_h values.

In fig. 5.2c the result of these optimizations is shown. The inhomogeneity from the end caps is almost eliminated, which explains the significantly lower value of c_h compared to optimized LCs without an adapted OM arrangement. The heightened dpy near 0 m radius and 47 m height was kept to reduce the drop that occurs between this region and the vertical ridge. One can see where a row of OMs was omitted on the wall at 43 m height and large radii. The dpy in the ridge is higher than in fig. 5.2b, since here the LC is truncated and a higher OC was chosen for the walls to reduce the influence of the caps (total OC 31.25%).

As can be seen, the homogeneity is greatly improved by optimizing the distribution of the OMs. The effect (reduction of c_h by a factor of 1.64 vs. optimal LCs only) is almost as large as from using an optimized LC (reduction by a factor of 1.94 vs. no LCs). When the systematic spread of the dpy in a detector is to be minimized, thus, both methods should be combined. In sum, a substantial improvement of the

[33] This is not possible for the CPC, whose shape at the entry aperture is parallel to the mirror axis per design.

homogeneity to values that are comparable to the statistic error is possible, which might be further improvable still. This is expected to notably lower the obtainable energy resolution of the detector.

Maximum detected photon yield

If the goal is to maximize the mean dpy instead, the optimal LC shape is quite different.
To reach a maximum dpy, it is beneficial to pack the OMs as close as possible and to maximize the OC. The reachable dpy is then limited by the diameter of the LC which determines the dimensions of the surrounding OM.
While the magnification factor rises towards smaller r_v, it turns out that the fill factor from the dead area of the OM brink (flange, distance of LC to housing) dominates, since it grows quadratically with the radius of LC and PMT. Therefore, the maximum reachable dpy (and thus the effective OC) increases with rising virtual source radii and the maximum μ is about 380/MeV (OC 39%) for a CSC with an r_v of 14 m (maximum).
If PMTs without LCs are used, the OM radius decreases further and an even closer packing of the PMTs is possible, which results in a mean dpy of 395/MeV (OC 40%). Since this is higher than reachable with LCs, in this case it is best to omit the LCs altogether.

JUNO (see 1.3.2) is an example of an LSND where the ultimate objective is the maximization of the dpy. Here, the use of LCs was also discussed for a time [742] but eventually was dismissed. While the studies of the present work were carried out for a cylindrical LSND and lower attenuation lengths and thus are not completely comparable, the results discussed above nevertheless indicate that the omission of LCs in JUNO was reasonable.

5.3.2 Materials

While an optimized light concentrator shape is required to obtain the best possible detector performance (this was addressed in 5.3.1), the layout and material of an LC are no less important. Their selection will be addressed now.

Requirements

The following requirements were identified:

- **Reflectance:** A light concentrator should have a high reflectance in the spectral range where the emission spectrum of the scintillator with WLSs after passage through the liquid overlaps with the quantum efficiency (QE) spectrum of the photosensor.
  In case of LENA, the emission spectrum extends from about 380 to 550 nm [81, 446], which falls into the sensitive range of a PMT with an (high QE) bialkali photocathode (see fig. 2.8a). The emission will be shifted to longer wavelengths after passage through several meters of scintillator and is weak above ca. 500 nm, so the range between 400 and 500 nm was considered to be most important.

- **Radiopurity and mass:** Furthermore, the used materials should contain only extremely low amounts of radioactive isotopes. To assess their compliance with requirements, all materials will have to be screened with a germanium detector. For the same reason, the mass of the LCs should be minimal, but their walls should not be too thin to avoid a deformation by gravity.
- **Lifetime:** In addition, chemical and physical compatibility with the buffer liquid (LAB) must be fulfilled over the course of the foreseen detector lifetime of 30 a.
- **Costs:** Finally, the materials and the production of the LC should be affordable.

Concentrator structure

For the LC structure, three layouts are possible: Either a massive LC made of a single reflective material is employed, a thin reflective coating is applied on a substrate, or a reflecting foil is glued or attached onto a substrate.

Massive LCs are made of metal and are heavier than acrylic glass substrates even if Al is used. The LC can be fabricated by deep-drawing and subsequent polishing of the inner surface. The material would have to be checked for radiopurity. For Ag as reflective material, the price for solid LCs would be prohibitively high.
Borexino used solid Al concentrators for the ID, which had a cost of about €100 per piece compared to a cost per PMT and electronics channel of ca. €2000 [734].

Coated LCs would weigh less and the radiopurity of the coatings is likely negligible due to their low thickness. Furthermore, higher reflectances than with massive LCs can be reached, since Ag can be used as coating. Acrylic glass substrates can achieve low backgrounds and can be fabricated by thermal deep-drawing or casting. The coating must be applied as an additional production step.
The CTF employed an LC made from UV transparent glass that was coated with Ag and a thin Cu layer for protection against water. [734] This was incompatible with organic liquids (cracks), however. The cost of an LC was about €230.
Other substrate materials which have been used are PETG [34], which is stronger and less brittle than acrylic glass [185], or electroformed nickel [758].

Gluing a highly reflective foil onto a substrate allows to reach a very high reflectance in the target wavelength range and is possibly cheaper than a coating, since no evaporation is needed. However, either the glue has to be compatible with LAB and be sufficiently radiopure or a mechanical fixation is required. Furthermore, it is more difficult to obtain a flat surface (no air pockets or folding) than with the other methods.

Studied structure and materials

Coated reflective surfaces were successfully realized for similar experiments (use in liquids, high radiopurity) as LENA. Compared to massive LCs, they allow to reach a higher reflectance, a lower weight and possibly lower backgrounds, whereas for reflective foils potentially several new technological challenges might have to be solved.
Therefore, a coated LC layout was selected.

[34] Polyethylene terephthalate glycol-modified; $n = 1.57$.

In a next step, a material study was conducted to test the reflectance and compatibility with LAB for multiple coatings.

As substrate material, acrylic glass was chosen, because its radiopurity was already verified by Borexino and CTF.
For the samples, thermoelastic poured acrylic glass (GS) from Röhm (Darmstadt) was used, which was cut to strips of $15 \times 2.5\,\mathrm{cm}^2$ size and 3 mm thickness.

Regarding the coating, Al and Ag show a high reflectivity in the relevant wavelength range [489; p. 172] [759] and were thus chosen as materials.
The coatings were applied to the test strips by Creavac Creative Vakuumbeschichtung GmbH (Dresden) by physical vapor deposition and in one case by chemical vapor deposition (CVD). After consultation with the manufacturer, also variants with additional protective layers (SiO_2, MgF_2, oxide layer) were produced for the tests.
For silver, the following coatings were studied: 100 nm Ag, 100 nm Ag + 500 nm SiO_2, 100 nm Ag + 150 nm MgF_2, and 100 nm Ag + 150 nm MgF_2 + 50 nm protective oxide layer [35].
The tested aluminum coatings comprised: 200 nm Al, 100 nm Al + 100 nm SiO_2, 100 nm Al + 20 nm SiO_2 (CVD), 100 nm adhesion agent (oxide layer) + 100 nm Al + 250 nm SiO_2.
Eight strips per coating type were produced.

5.3.2.1 Compatibility with buffer liquid

Accelerated aging

For a combination of substrate and coating to be eligible, its chemical and physical compatibility with the buffer liquid has to be ascertained for timescales of a similar magnitude than the foreseen operating time in the detector.

This was tested by immersing samples of each coating type in LAB and subjecting them to accelerated aging though elevated temperatures of up to 92°C. This method was also used for the LCs of the CTF and Borexino [734] and the LBNE WCD [185].
The simulated time of exposure t_s can be obtained in good approximation through the dependence of the chemical reaction rate on the temperature, which is described by the Arrhenius equation: [734] [185; pp. 6-206–6-207]

$$t_s = t \cdot 2^{\frac{T_e - T_o}{10}} \tag{5.3}$$

where t is the elapsed laboratory time, T_e is the elevated temperature, and T_o is the operating temperature in the detector.
Both physical and chemical reactions can occur between the liquid and the samples. Physical reactions include the solubility of the surface layer(s) and cracks, erosions and flaking from the capillary action [36]. Chemical reactions with LAB can take place for both the coatings and the substrate.

[35] Produced with a proprietary method.
[36] Seepage from the (unprotected) sample brinks and faults in the coatings.

Test methods

The compatibility with LAB was assessed through two methods.

First, the spectral reflectance of a sample strip was determined before immersion in LAB and after the maximum simulated exposure time. This is discussed in the subsequent section, 5.3.2.2.

Second, the absorbance of the liquid in which the samples were placed was also measured before, during and after aging. This method will be discussed in the following. The basic idea is that if dissolved particles are created by chemical of physical reactions, this raises the absorbance, which allows to detect the deterioration. In addition, in this way the time behavior of the transparency of the buffer liquid, which is important for the mean dpy, can be determined. The same method was applied by [435] to verify the compatibility of FINEMET with mineral oil[37].
Measurements were performed before contact with the test strips and after one day of exposure at room temperature to check whether an increase occurred due to dust or impurities on the surface or from an extreme reaction. Afterwards, the temperature was raised and additional measurements were carried out after effective elapsed times of about 300 d of immersion and after 6.2–6.7 years. For all materials, the same surface (eight test strips) was exposed to 0.9 l LAB. The absorbance spectra were measured with a Perkin Elmer Lambda 850 spectrometer (sensitive range 200–800 nm).

Results of the absorbance measurements

The results of the absorbance measurements are shown in fig. 5.3 exemplarily for the unprotected 100 nm Ag coating, which showed the best reflectance both before and after accelerated aging.

After 1 d at room temperature no notable change was observed for any sample, so subsequent results were compared to this baseline measurement.

After exposures at elevated temperatures that were equivalent to about one year and 6–7 years, all samples showed an absorbance increase at short wavelengths up to 450 nm with nearly the same amplitude and spectral shape (compare fig. 5.3b).
For an unprotected Ag layer after 6.7 a aging, the absorbance increase displayed a quasi-exponential behavior between 430 and 800 nm with larger values for shorter wavelengths. From 430 to 360 nm, the change of absorbance increased faster with falling wavelengths, which was followed by an even stronger linear growth down to ca. 315 nm. Below ca. 320 nm large absorbance values occurred in all measurements, which produced strong fluctuations in the difference between two measurements due to statistical and systematic errors. Therefore no reliable statement about the wavelength dependence can be made in this region, although a large average increase was found here as well.
Between 400 and 450 nm, for all samples the absorbance increases are mostly compatible with zero within errors, since the absorbance of LAB is small in this region and the systematic errors are comparatively large for a differential measurement. Since the dimensions of the systematic errors were estimated very conservatively, the true measurement uncertainties will likely be notably smaller, though. This is

[37] No notable increase between 300 and 650 nm was found after 28 d at 40°C.

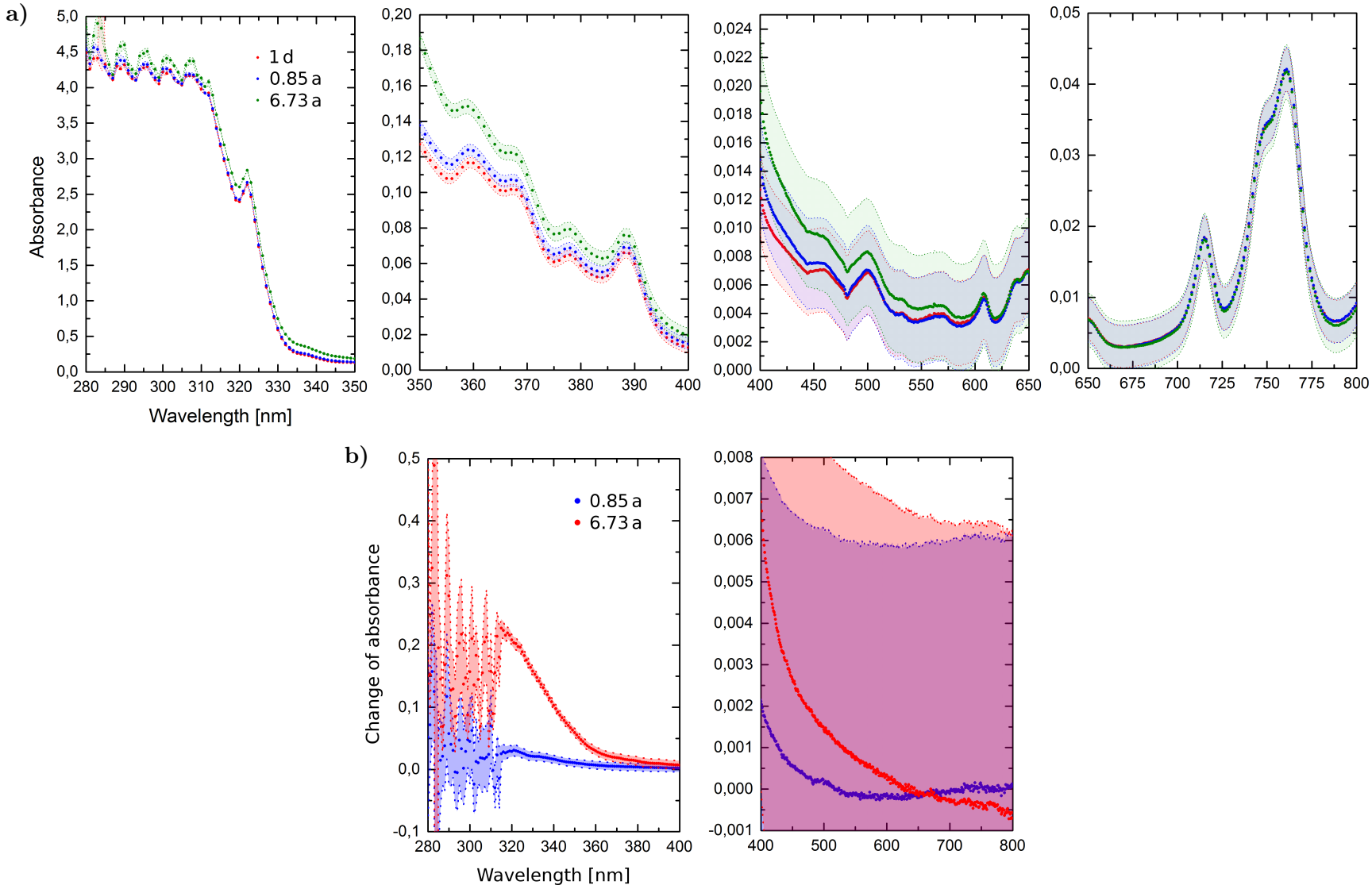


Figure 5.3: Material compatibility tests: Absorbance increase of LAB (linear alkylbenzene) after immersion of acrylic glass samples coated with 100 nm Ag and accelerated aging through heating. Eight sample strips of 15×2.5 cm² and 3 mm thickness were immersed in 0.9 l LAB. Discussion in text. Figures adapted from [725].
a) Absorbance after 1 d at room temperature (values virtually identical to before immersion) and after 0.85 a and 6.73 a using accelerated aging.
b) Absorbance changes compared to the initial measurement.

corroborated by the observance of a systematic behavior within the error bars which is in line with physical expectations (discussed below).
When linearly scaling the measurement for 0.85 a to an elapsed time of 6.73 a for 100 nm Ag, the absorbance increases are similar (slightly higher) to the longer measurement below 500 nm, which supports a linear increase with time. Above this wavelength, the values are compatible with no increase but are difficult to compare due to the small absolute values compared to the errors.

The uniform rise of the absorbance in all samples suggests a solution or reaction of the acrylic glass substrate.
Unfortunately, the acrylic glass type of the samples is unknown. However, measurements of the transmittance of bulk acrylic glass sheets in the literature show a similar spectral dependence for the UV transparent clear acrylic glass variants PLEXIGLAS GS 218, 238 or 2458 [760–762]. This points to a solution of the substrate.
In order to verify this, the accelerated aging measurements would have to be repeated for bare substrate strips. At the same time, multiple acrylic glass types from different manufacturers should be tested to find compatible variants in case the substrate is confirmed as origin of the absorbance increase.
Alternatively, the higher absorbance might be caused by a thermal degradation of the LAB, and this should be studied by subjecting LAB without sample strips to accelerated aging.

In order to qualitatively analyze the origins of the observed absorbance changes of individual samples, the reflectance and absorptance of thin bulk films of the used coating materials were calculated using the Reflectance Calculator of Filmetrics [763].

For the Ag-based samples, an additional absorbance increase between about 400 and 700 nm was found in the measurements, which would match the calculated spectral absorptance of a thin Ag film very well. The measured increase was suppressed reliably with a protective coating of SiO_2 or MgF_2 + oxide, while MgF_2 alone either was not sufficient or dissolved itself.

In the Al-based samples, a slight increase of the measured absorbance was found above 550 nm, which increased with wavelength. This would match the calculated absorptance of bulk Al. With a protective layer of SiO_2, the increase was slightly reduced, with SiO_2 (CVD) or oxide + Al + SiO_2 it was significantly better, and with a MgF_2 layer no increase was observed. This suggests the partial solution of the Al coating.

Regarding the benefit of the protective coatings, MgF_2 appears to protect the Al layer (no increase at large wavelengths) but not the Ag layer. For the latter, the absorbance showed an even higher increase than for unprotected Ag, where the spectral shape matches that which is expected from Ag of MgF_2. Thus, either the protective MgF_2 coating is dissolved or the solution of Ag is enhanced by its presence; the former seems more likely. With two protective layers (Ag + MgF_2 + oxide) the protection of Ag was improved and above 450 nm the absorbance was lower than for a bare layer, although for one of two samples an additional absorbance peak emerged around 800 nm.
An SiO_2 layer also appears to improve the resistance of Ag. For Al, on the other hand, the absorbance was significantly higher with an extra SiO_2 layer up to 500 nm,

which could be caused by the solution of SiO_2. When the SiO_2 layer is applied to Al with CVD evaporation instead, this additional increase and the solution of Al were reduced or eliminated (depending on sample), while an oxide underlayer prevented the solution of Al but did not reduce the solution of SiO_2.

In summary, unprotected Ag or Al layers showed some signs of partial solution after 6–7 a. However, these indications are well within the (conservative) systematic errors and are also compatible with the absence of an effect. A suitable protective layer (SiO_2 for Ag, MgF_2 for Al) could eliminate this erosion, but was found to notably reduce the reflectance (see 5.3.2.2).
Since an unprotected Ag film of 100 nm thickness showed the highest weighted spectral reflectance, while its absorbance increase in the crucial wavelength range between 400 and 500 nm was small, this material was chosen as coating for the LENA LCs.

Estimated transmission drop

Based on these results, the decrease of the transmission in the OM due to the solution or reaction of LC materials can be estimated under the worst case assumption that the behavior cannot be improved further.

From the Beer-Lambert law follows for the intensity I after a traversed distance x [764; p. 595]

$$I(x) = I_0 \cdot e^{-x/\lambda_a} \tag{5.4}$$

where I_0 is the original intensity and λ_a is the attenuation length in the liquid.
When assuming a mean photon path length of 0.5 m and $\lambda_a = 10$ m in LAB, the transmittance at the time of commissioning is 95.1%. When the measured absorbance at 400 nm [38] for 100 nm Ag is scaled to 1900 cm^2 LC surface and a buffer volume of 56 l (see 5.4), the transmittance after 6.7 a amounts to $94.4 \pm 0.8\%$. When a linear increase of the absorbance with time is assumed, a transmittance of $91.4 \pm 3.6\%$ remains after 30 a, which corresponds to a decrease of 3.7% due to aging.
This reduction of the mean dpy is secondary compared to the PMT failure rate (see 2.7.3) but sizeable enough to warrant further improvements. The large uncertainties should be lowered by eliminating some currently used approximations through additional measurements, e.g. the spectral refractive index of the employed cuvettes and liquids (see [725] for details).

Conclusion

In conclusion, reactions of the LC materials with the buffer liquid were found to occur, but they did not strongly affect the transmission around 420 nm. The observed amount of solution was therefore deemed acceptable if the reflectance of the respective materials is still high after a projected age of 30 a (studied in the next section). However, the decreasing transmittance of the OM has to be considered in analyses of the physics potential. Further studies of the thermal aging of pure LAB and the compatibility of different uncoated acrylic glass substrates are advisable to establish the source of the absorbance increases and to possibly reduce them.

[38] At this wavelength the highest absorbance value in the critical spectral range was found.

5.3.2.2 Reflectance

The reflectance measures the amount of reflection from a material with finite thickness, for which e.g. the interference with rays that are reflected on the backside play a role. The material here can consist of multiple layers of different substances.
The reflectivity, on the other hand, measures the reflection for a bulk material with infinite thickness, and thus constitutes an intrinsic property of the material itself.

In addition to the absorbance measurement, the spectral reflectance of all coating types was determined.
This was done not only before but also after the accelerated aging as an additional method to determine the material compatibility and the rate of degradation. This is important, since some effects like flaking might not affect the absorbance; in this case large flakes could sink to the bottom of the bottle or cuvette instead.
The reflectance was measured with a Perkin Elmer Lambda 900 spectrometer with a module for spectral reflectivity measurements. The studied wavelength range extended from 300 to 550 nm, and an incident angle of 6° and a spot size of ca. $4.4 \times 11.7\,\mathrm{mm}^2$ (maximal size to average over the largest possible area) were used.
The weighted mean reflectance was calculated using the QE spectrum of the R11780-HQE from [765] and the emission spectrum of LAB + 3 g/l PPO + 20 mg/l Bis-MSB after passing through 0.5 cm of scintillator (excited at 190 nm) [81, 446]. This results in a conservative estimate of the reflectance for Ag-based samples, since after long distances in the scintillator the spectrum will be shifted to larger wavelengths, where the reflectance of Ag is higher.

A reduced sample reflectance after aging can be due to a reduced surface layer thickness, cracks and surface roughness from erosion (diffuse reflection is not detected in the setup), and chemical reactions of the surface layer. It can, however, also originate from scintillator residues.

For an unbiased measurement of the reflectance, therefore, the test strips have to be cleansed from the thin LAB layer that wets the surface, since this alters the reflectance or might smooth out surface irregularities.
However, the cleaning agents can attack the surface with the same mechanisms as LAB, which makes it difficult to assess the influence from aging in LAB alone. Therefore, only upper limits for the reflectance losses due to the exposure to LAB can be given.
In order to exclude effects that result from reactions of the cleaning agent with the samples, two different solvents were used for cleaning: ethanol and isopropyl.
Some samples were indeed visibly attacked by the solvent: For $\mathrm{Al} + \mathrm{MgF_2}$ and oxide + $\mathrm{Al} + \mathrm{SiO_2}$, no damages were visible after aging while the test strips were in the bottle or when they were still wetted with LAB; however, after cleaning in ethanol fine cracks developed.
In general, ethanol was found to be more aggressive and led to a stronger reduction of the reflectance than isopropyl for all samples where notable changes were found. Other solvents and cleaning methods should be tested, as well, to see if they can provide a better compatibility.

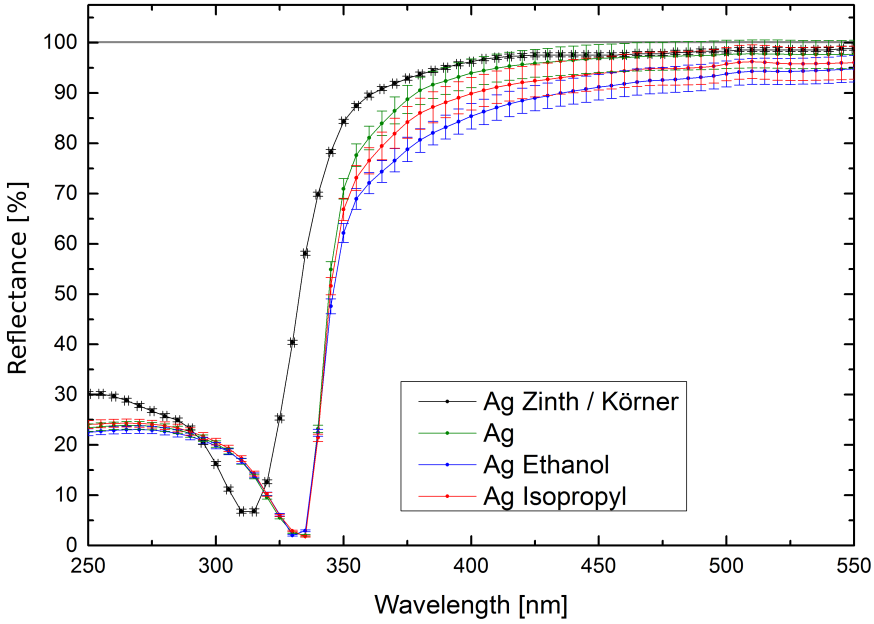


Figure 5.4: Spectral reflectance of uncoated silver: Measured values for 100 nm Ag coating on 3 mm acrylic glass. Green before aging, blue and red after 6.7 a accelerated aging and cleaning with ethanol and isopropyl, respectively. Black curve for bulk Ag (reflectivity), values from [759]. Discussion and comparison in text. Figure adapted from [725].

Results

For Ag, the highest reflectance before and after aging was found with both cleaning solvents for the samples that were coated with 100 nm Ag without protective layers (see fig. 5.4).
Unprotected Ag was less affected by isopropyl than ethanol, and the resulting reflectance difference between both solvents is as strong as the total effect from aging and cleaning with isopropyl. The ethanol-cleaned sample showed no changes in a visual inspection except for a small dull area inside the beam region. According to estimates, however, this might already be sufficient to partially or even completely explain the observed reflectance decrease.
With isopropyl, a systematic decrease of the reflectance by a few percent is observed between about 340 and 550 nm. The isopropyl-cleaned sample showed scratches, which might stem from dust in the shrink-wrapped foils, in which the samples were stored before immersion. The measured reflectance decrease could be explained partially to completely by the surface fraction which was covered with scratches. In addition, due to the surprisingly large systematic errors of any reflectance measurement [39], the observed reduction is still compatible with zero within the error bars and thus only constitutes an upper limit.
The weighted mean reflectance of 100 nm Ag amounts to $95.5 \pm 3.2\%$ before immersion and $\geq 92.1 \pm 3.6\%$ after 6.7 a aging and cleaning with isopropyl. A linear extrapolation to 30 a results in a residual reflectance of $\gtrapprox 80.3\%$, but the large error bars prevent a reliable prediction.

[39] Originally, a different reflectance characterization setup was used, which was improved several times but could not achieve the targeted uncertainties.

For use in a detector, the Ag layer thickness should be higher to achieve an improved resistance to cleaning and a better durability, since unprotected Ag showed signs of solution in the absorbance measurements. An exposure to cleaning agents should be less critical in a real LC, however, since here no open side faces which can allow seepage are present as in the samples. Furthermore, scratches and defects from dust likely can be reduced in the manufacturing as well. Although these measures should allow to increase the reflectance after aging, in the LC simulations in 5.3.1, the measured reflectance after 6.7 a aging was used as a conservative estimate.
For unprotected Ag, the reflectance drop in the UV range was found at higher wavelengths than expected from the bulk material according to literature [759] [489; p. 172]; however, the position of the minimum also varies between both listed sources. This shift cannot be reproduced by a finite thickness of the material using the reflectance calculator, and likely is due to a surface oxide layer of varying thickness.

The measured reflectances were generally smaller than in literature for short wavelengths for all samples, and for uncoated Al the reflectance started to drop below 350 nm, which is not expected. This likely is a systematic effect of the setup and the reflectance standard sample that was used to calibrate the reflectance. For Al the effect might also be partially explainable by a surface oxide layer.

For unprotected Al, the weighted mean reflectance showed virtually no changes after aging when cleansed with isopropyl (88.0% → 87.6%), except for a slight decrease below 320 nm which is irrelevant here. This material would be suitable for the LCs as well, but has a lower mean reflectance than uncoated Ag.

With protective coatings, the reflectance decrease for silver-based materials was similar or stronger than with bare Ag, except for Ag + MgF_2, which showed a smaller reduction between 400 and 450 nm. For all combinations of Ag with protective layers, the mean reflectance was lower than with Ag alone both before and after aging. Since for Ag in all cases the additional layers mostly provided no notable protection and decreased the reflectance at the relevant wavelengths, it is better to use an uncovered silver layer.
For Al, a protective layer also leads to larger reflectance losses, except for SiO_2 (CVD), and is not needed either.

Conclusion

The reflectance and material compatibility of Al- and Ag-based coatings on acrylic glass substrates with optional protective layers were studied.
An unprotected Ag layer of 100 nm thickness showed the highest weighted reflectance before and after 6.7 a of accelerated aging with no significant deterioration of the reflectance or the transmission of the buffer liquid, and thus was selected as reflective coating for the LENA LCs. For the production of the LCs, a larger layer thickness should be chosen, however.
As next steps, different cleaning methods or solvents should be tried and the tested exposure time should be increased to the full foreseen operating time of 30 a. Furthermore, the radiopurity of substrate and coating should be studied, and a full LC prototype that is adapted to the selected photosensor should be tested.

5.4 Pressure housing

Since in LENA peak pressures of up to 1.3 MPa can occur during filling and even recently developed pressure-optimized PMT series cannot withstand such loads (see table 3.7) [40], it is imperative to encapsulate the sensors in a pressure housing.
The peak pressure is lower in other detectors which are more compact. Here, it can be sufficient to house only the most sensitive stem region of the PMT envelope together with the adjoining voltage divider and to surround the PMT with a shockwave shielding to prevent a chain reaction in case of an implosion.

The pressure housing for the LENA ID additionally serves as an OM which contains the LC, buffer liquid, magnetic shielding, voltage divider, and possibly additional electronics. The OM also protects its components from water during the cleaning phase.

Regarding the requirements of a pressure encapsulation, besides the obvious capability to absorb pressure, the housing should be as light as possible to minimize the amount of radioactive isotopes in its materials which is introduced into the detector.
It should also be compact, since this allows to reach a higher OC and reduces the mass of the buffer liquid inside the OM. Before the detector is filled, the weight of the OMs is not reduced by buoyancy and the scaffolding that holds the OMs has to be designed to carry the full mass of the OMs. For a higher OMs weight, thus, the scaffolding has to made sturdier and thus heavier, which again introduces radioactive impurities.

Method

In order to design and optimize the pressure housing, static pressure simulations were carried out for multiple prototypes. For this, the software SolidWorks [766] was used, since it allowed to both construct the models with CAD and perform FEA simulations on them.
These studies were started in a previous work [726] (at that time still for OMs without buffer liquid and LC) and were continued for the new OM layout in [725] and afterwards.
In the final simulations, the maximum stress was assumed to be 22.5 N/mm^2 for acrylic glass and 200 N/mm^2 for stainless steel. According to DIN IEC60404-8-6, a factor of safety of 1.5 would likely be sufficient, but here a more conservative value of 2 was used to account for the effect of aging.
The thickness of the acrylic window and steel housing that is required for a load of 1.3 MPa was determined through FEA simulations (see fig. 5.6). The layout of the different designs (see below) was iteratively optimized to reduce the mass. Here, similar considerations as for the pressure optimization of the PMT envelopes apply (see 2.7.2). Areas with small curvature radii lead to large bending strains and shear stresses, so abrupt kinks and sharp curvatures should be avoided, and the transitions between straight sections should be as smooth as possible. If needed, the curve regions can be reinforced.

[40] For the Hamamatsu R11780-HQE, the more conservative pressure capability of 0.9 MPa stated by the manufacturer seems more realistic, since short term pressure tests reach higher values than achievable in long-term operation (see p. 439).

Care should be exercised when analyzing the stress distributions that result from FEA studies (fig. 5.6), since the method tends to transfer forces to the edges, where unphysical singularities can occur. The resulting values can exceed the compressive strength of the material, although a real housing would be stable. This was observed for example for the screw holes in the flanges (maximum value in fig. 5.6), and experience is needed to distinguish critical from uncritical excesses.

Pressure housing layouts

The pressure housing itself consists of two parts: a transparent acrylic glass window and a stainless steel housing.
Both materials can be obtained with high radiopurity and are chemically sufficiently compatible with LAB. Acrylic glass also has a low density but is more brittle. This requires a larger thickness, which partially offsets the density advantage. It is planned to fabricate the window and steel part with deep-drawing.
Gaskets are used to tightly seal the contact surface between the window and the metallic rear.
Due to the incompressibility of liquids, also a small gas bubble should be included in the OM to compensate a slight reduction of the internal volume by ambient pressure or temperature changes. This might also improve the shock resistance.

The shape of the pressure housing is already determined roughly by the PMT and LC. In order to account for production variations, a minimum safety distance of 5 mm between LC and PMT and to the housing was required.
The design of the housings was based on the Hamamatsu R11780 PMT (12”) and its homogeneity-optimized LC (CSC with 12 m virtual source radius). In an earlier stage, a CSC for 11.25 m was used (the best LC shape at the time).
Since the shapes of the housing, the magnetic shielding and the internal holders depend on each other, the design of all three structures was an iterative process.
The assembly process was also considered in the layout of the OM.

In [726] a study was performed for the old OM layout without buffer liquid and LC with the target to find the best housing shape and the respective minimum wall thicknesses. Here, three different PMT series were used, the Hamamatsu R6594 (5”), R5912 (8”) and R7081 (10”). For each PMT, up to four different OM shapes were studied where adequate, all of which were based on a hemispherical window. This included spherical (window and rear hemispherical), cylindric (cylinder-shaped rear), elliptical (rear with flat bottom and otherwise elliptic), and conical OMs (rear first cylindric, then conic near the window).
Although these studies were carried out for a different OM concept and other PMTs, some general results can be used for the updated design.
The spherical layout has a significantly different shape than the combination of PMT and LC and thus would be very heavy and too bulbous to reach the target OC.
For a cylindrical design, the internal volume is also very large and the edges or the bottom plate would have to be reinforced or rounded.
The elliptical and conical shape approximate the OM components far better and should result in lighter housings, which is why they were included in the final selection and were optimized for the new OM concept (see fig. 5.5).

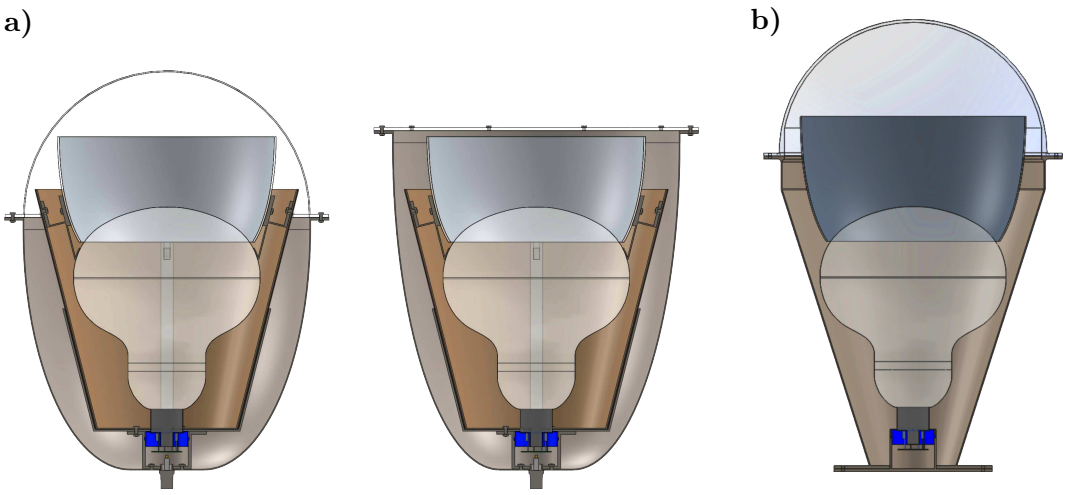


Figure 5.5: Final pressure housing layouts.
a) Elliptic variants with spherical and flat windows. Adapted from a figure created by German Beischler during his master thesis [725].
b) Conic variant.
The elliptic housings were still constructed for a preliminary LC shape, while the conic encapsulation uses the final, longer LC and already includes some pressure optimizations (reinforcements near the top flange). In the elliptic designs, all OM components including the holders were integrated and the assembly was considered; the conic layout is far simpler, so holders were omitted.

From the old simulations and additional tests it was estimated that a thickness of about 2 mm should suffice for the metal part and an acrylic window of 3 mm (hemispherical) or 6 mm (flat) thickness should be able to withstand the pressure.

Elliptic housing

For the elliptic pressure housing (fig. 5.5a) the steel part has a flat bottom to allow an easier assembly (see 5.2). This flat section is connected to a rotated ellipse which ends perpendicular to the flange. Since no kinks or edges are present and the shape still resembles a sphere, a good stability is expected, which could translate to a lower mass of the steel part.

Two variants were developed for this design: One with a hemispherical window and one with a flat window and a longer metal housing. As this was done at an earlier time, a preliminary LC shape (CSC for 11.25 m) was used, which is shorter than the final LC (compare figs. 5.5a and 5.5b).

Due to time constraints, no FEA simulations could be carried out in [725] to determine the required thicknesses. However, based on previous studies, the resulting weight could be estimated empirically using the density of semi- to high-impact acrylic glass and stainless steel 1.4404 in SolidWorks.
With a spherical window, the OM has a radius of 26.5 cm, a height of 68.3 cm, contains 66 l LAB, and has a total weight of about 74 kg, of which 11.1 kg are stainless steel (2 mm thickness) and 1.1 kg are acrylic glass (3 mm thick). A flat window results in a radius of 26.6 cm, 59.1 cm height, 56 l LAB and a total weight of ca. 69 kg, including 13.8 kg steel (2 mm) and 1.3 kg acrylic (6 mm).

The hemispherical variant places a larger buffer volume before the PMT, which is beneficial for background reduction.

With a flat window, the OM is slightly lighter due to the smaller buffer volume, and the fabrication is easier and possibly cheaper. However, a flat window is unfavorable with regard to stability and thus a significantly larger thickness is necessary. This reduces the transmission of the window minimally and the acrylic glass weighs a bit more (mostly compensated by the smaller surface vs. the hemisphere).

Since the used buffer liquid is extremely clean, the OM weight matters only for handling purposes and the needed stability of the OM scaffolding before filling.
For the spherical type, 2.7 kg more steel are required, but the total weight is slightly lower, so the scaffolding might be somewhat lighter.
The mass of the acrylic glass is similar in both variants, and since this material can be obtained with low backgrounds, this should be no crucial issue.
In sum, it is difficult to estimate the total background from the OMs and their support structure without engineering the scaffoldings for both variants.
A buoyancy of the OM due to the PMT vacuum might also be an issue when the detector is filled with water or scintillator, which could require to anchor the scaffolding. However, the mass of the solid OM components compensates the buoyancy in both designs.

As far as can be assessed with these estimates, the hemispherical layout should lead to lower background rates because of a larger amount of buffer liquid in front of the PMTs and thus is favored. However, further simulations would be required to obtain the final thicknesses and determine the resulting background rates.

Conic housing

A conic encapsulation has a shape which closely follows the outline of PMT and LC if the metal part starts as a cone and ends in a cylindric part and a hemispherical window is used (fig. 5.5b) [41].
The studies of this design could not be completed due to time constraints, but the shape should result in a smaller internal volume than for the elliptic design and lower masses of the acrylic glass, steel and OM. In the figure, the final prototype which uses the optimized LC design is shown with 5 mm wall thickness for the acrylic glass and 2 mm for steel. The OM is longer but narrower than the elliptic variants, which allows to achieve higher OCs.
In this design no complicated holder is required for the magnetic shielding and LC. The former can simply be placed on the walls and attached with glue or clips, and the LC can be attached to small holders on the steel walls. This greatly simplifies the design.
For the base plate, an additional flange is used, which allows to preassemble the VD cylinder and the base plate. The assembled structure is then attached with gaskets to a flange which was welded to the conic part and has a slot for the VD cylinder. This further simplifies the construction.
In order to improve the stability, a small inner metal rim was added at the top flange and the thickness of the acrylic was increased in this region.

[41] Alternatively, the cylindric section could be extended to allow to use a flat window, but this would reduce the amount of buffer liquid in front of the PMT.

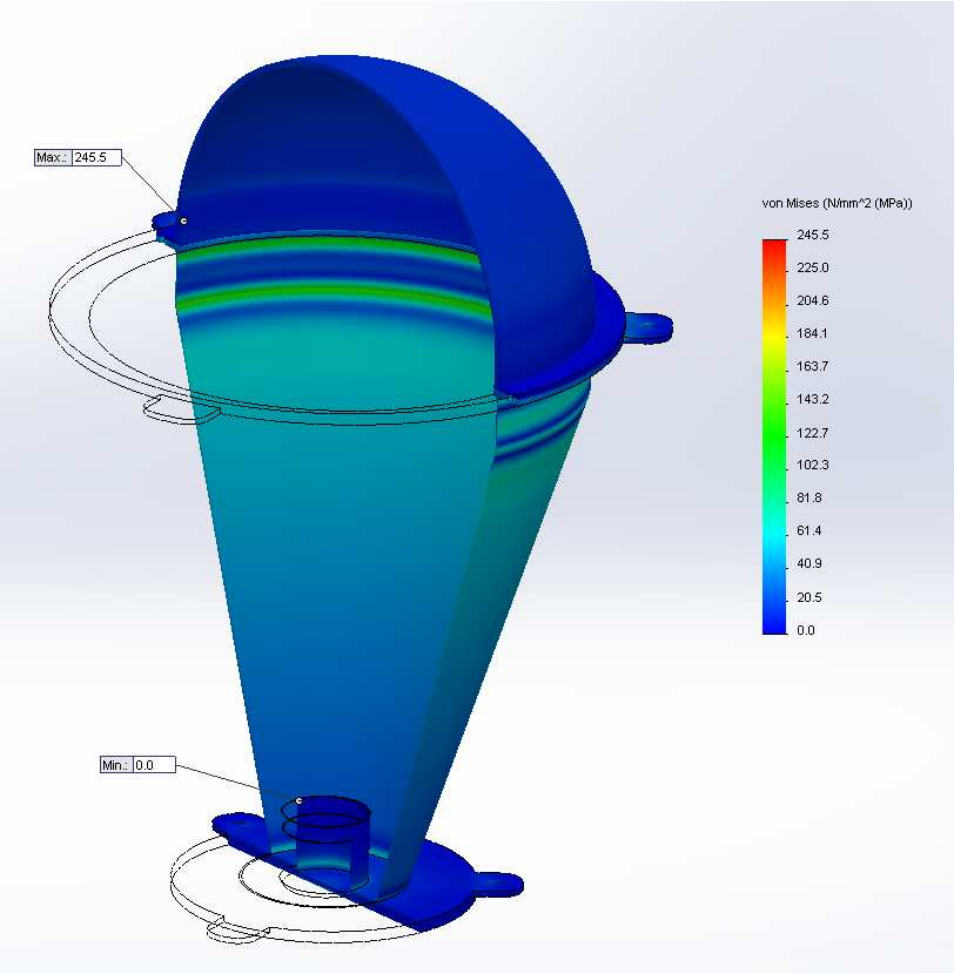


Figure 5.6: Stress distribution in a pressurized conic OM, obtained with finite element analysis in SolidWorks. The maximum values occur in the screw holes of the top flange and are artifacts of the simulation method. The actual maximum stresses are within material tolerances for both acrylic glass and stainless steel.

In fig. 5.6 the stress distribution which was obtained with the static pressure simulation for the final prototype is shown. Here a thickness of 5 mm was used for the acrylic glass and 2 mm for the steel.
When ignoring artifacts of the FEA, for the window no region exceeds the compressive strength of acrylic glass, and for the steel housing the most critical point is at the inner side of the kink between the conic and cylindric part (148 N/mm^2). This is within the allowed range for stainless steel, but the kink should be rounded nevertheless to further reduce the maximum stresses and thus possibly allow to use thinner walls.
In addition, one could add ribs to the acrylic window or the conic part for added stability. This should not affect the optical properties much, because the refractive index is virtually identical for LAB and acrylic glass.

Conclusion

Due to the expectable lower masses of steel, acrylic glass and OM and the far simpler layout and assembly, the conic OM is favored over the elliptic layouts for the R11780. Based on FEA simulations, a method to design pressure housings and to optimize them for minimum material need was established.

In a next step, this procedure could be carried out for the new favored PMT series and photosensor types (e.g. R12860, R12850, R12112; see 3.3 and 4.5.1). At the same time, possible improvements such as larger curvature radii around kinks and the addition of ribs should be tried to potentially further reduce the mass of the housing.
In principle, one could also employ different pressure housings for different depths in the detector, which would allow to use thinner hulls for the upper OMs and in this way reduce the total radioactive load.

5.5 Summary

The example of LENA shows that a cutting edge neutrino detector with a broad physics agenda can be realized with existing sensors if assistive equipment in form of optical modules is used. These here consist of a pressure housing, a light concentrator, a magnetic shielding and buffer liquid. Combining this with new photosensor developments might allow to further improve the detector performance and extend the discovery potential.
Since THEIA is a cylindrical scintillator detector as well, a similar OM layout could be used. The procedures which were employed here to construct an LC that is optimized for homogeneity, to test LC materials, and to design a light-weight pressure housing would be directly applicable.

Conclusions

Liquid scintillator neutrino detectors (LSNDs) have proven to be one of the most promising detector technologies for neutrino physics. In these and similar detectors, the photosensors are possibly the most crucial component and their performance has a strong impact on the discovery potential for new physics.
Therefore, this work had a two-fold objective: First, to determine which existing and future photosensors are best suited for large LSNDs such as LENA [1], JUNO [2] and THEIA. And second, to pave the way for improvements of photosensor performance through a) an enhanced understanding of the effects causing irregular behavior, b) a systematic analysis and classification of all relevant photosensor types and their properties, and c) the development of auxiliary photosensor equipment.
Due to its mature detector design and a physics case which is similar to many other planned large LSNDs, the LENA project was used as a benchmark for LSND photodetection requirements throughout the work.
Photomultipliers (PMTs) continue to be the sensor of choice together with emerging new photosensors whose characteristics are in large part similar, since they are based on phototubes as well. Therefore, a focus was put on the effects causing transit time shifts, early afterpulses (EAPs, delay 0 to ≈ 100 ns) and light emission in PMTs – all of which can have a detrimental influence on detector performance and could limit the discovery potential.

Chronology

In order to find the sensors best suited for LSNDs, first the properties of promising PMT series were studied through measurements at the Laboratori Nazionali del Gran Sasso. This revealed that important parameters – the transit time distribution, EAPs and light emission from PMTs – were not yet fully understood. Therefore, in the following a focus was put on analyzing the influence of these effects. For this purpose, a shielded photosensor characterization setup optimized for timing and waveform acquisition (SPAX [3]) and a new method of PMT waveform analysis based on pulse fitting (PEST [4]) were developed. These efforts culminated in the formulation of a novel theory (time shift theory, TST) of the effects which induce EAPs and cause the strong variance of transit time, charge and pulse shape. This model allowed to completely explain the charge, time, and probability of both primary pulses (PPs) and EAPs in all measured distributions and extended the understanding of PMT behavior. The development of all three – TST, SPAX and PEST – occurred in

[1] Low Energy Neutrino Astronomy
[2] Jiangmen Underground Neutrino Observatory
[3] Shielded Photosensor Analysis eXperiment
[4] Photosensor Evaluation Software Toolkit

parallel, where progress in one method led to improvements and adaptations of the others.
At the same time, an optical module was developed based on early stage PMT series candidates which includes a 12" diameter Hamamatsu [5] PMT, a light concentrator optimized for detector homogeneity, and shielding buffer liquid.
In parallel to these endeavors, alternative photosensor types were analyzed through an extensive literature research. To structure the results, a classification of photosensor types based on the particle interactions involved in signal formation was devised, which can be used to systematize future photosensor developments.

The findings of these various projects will now be presented shortly, following the structure of the work.

Chapter 1: Photodetection requirements

To set the goals for photosensor performance, in ch. 1, first the advantages of LSNDs compared to other neutrino detector types were listed, before the detector requirements for LSNDs were discussed. This showed the central importance of photosensor performance. To serve as a baseline for the studies, the photodetection requirements of LENA were detailed and updated using new results.

Chapter 2: Photomultiplier theory and properties

TST is based on the fundamental interactions of particles with structures inside the photosensor and the transport processes in between. Deviations from the processes involved in regular pulse formation are labeled time shift effects. In order to derive this theory, a thorough theoretic understanding of all processes in the sensor was necessary, which required to compile the state of knowledge on PMT properties. Therefore ch. 2 also acts as a comprehensive review of PMT theory – with additions and corrections resulting from TST and the associated measurements – as well as of practical PMT properties and how to measure them.

Statistics and analysis methods

Independent from TST, throughout the chapter the statistical and analysis methods which are used for photosensor characterizations were summarized and – where necessary – extended. New contributions were made regarding the statistical treatment of the influence on measurement accuracy from occupancy, multi-hits, dark count rate, afterpulses (APs), the subtraction of baseline and dark count rate, and statistics. The connections between some properties of the charge distribution were deduced. Also, the advantages of waveform smoothing and deconvolution of the light source emission time distribution were shown. In addition, the minimum pulse height threshold

[5] Hamamatsu Photonics K. K.

resulting from electronics noise was derived, the correction procedure of unshadowing was extended to PPs, and a generalized statistical treatment of AP rates and how they are influenced by correlations was introduced.

Photodetection efficiency

Section 2.1 addresses the photodetection efficiency and its components. Here, the definition of the photodetection efficiency had to be extended in response to TST results. The contributions to the multiplication efficiency were analyzed and classified into effects from cascade termination or threshold losses. The value of the multiplication efficiency was estimated and possible ways to separately measure collection efficiency and multiplication efficiency were discussed. Findings from the fields of scanning electron microscopy, transmission electron microscopy and Auger electron spectroscopy were introduced to gain a better understanding of the possible particle interactions of electrons in a PMT. This allowed to compile a complete list of possibly created particles and quasiparticles and occurring scattering types.

Charge

In 2.2.2, the charge response of PMTs was treated. The contributions to the charge distribution (single electron response) were analyzed. New fit functions (compound and scaled Poisson distributions), which provided excellent agreement with the results, were introduced and compared to existing models. Possible sources of underamplified pulses and their influence on the charge distribution shape were identified. Furthermore, the existence of overamplified pulses which result from light feedback was predicted and their contribution to the mean single photon charge was estimated.

Pulse height

The pulse height was covered in 2.2.3. The differences between pulse height distribution and charge distribution due to statistics, noise and time shift effects were summarized. Also, methods to estimate the height from the charge (or vice versa) were discussed.

Linearity

In 2.2.5 the allegedly better pulsed linearity of BeCu over Cs_3Sb dynodes was analyzed and found to originate mostly from higher applied voltages for BeCu.

Time response

The time response of PMTs was treated extensively in 2.3. Here, all factors contributing to the transit time of regular pulses[6] and its fluctuation (transit time spread)

[6] Pulses resulting from regular amplification in the PMT without time shift effects.

were collected and compared. Furthermore, the dependence of the transit time on the PMT diameter was analyzed and found to be in good agreement with a linear function. It was described how to accurately determine and eliminate the influence of the setup on a time response measurement. In addition, three definitions – depending on the purpose – for the rates of early pulses and late pulses were provided.

Time shift theory

The transit time and delay distributions of PPs and APs[7] show a multitude of effects producing deviations from the transit times and charges of regular pulses. A thorough literature study revealed that only a few of these effects were known and had been described before, which is why this was made a focus of this work. To explain the origins of the observed pulse clusters, thus TST was developed. This theory was summarized in 2.3.5 and is based on possible particle interactions on PMT structures connected by transport processes. This allows to predict the probability of time shift effects and the time, charge and height of all created pulses.

A time shift effect is based on an irregular interaction of a particle in the photosensor. Here, an interaction is defined as a combination of a particle variety, interaction type and location.
For this reason, the first step towards understanding PMT behavior was the identification of all occurring particles and their subtypes, feasible interactions for these particles, and the locations where each interaction can take place. Possible combinations of particles, interaction types and locations were identified and where physically possible multiple interactions were stacked to produce higher-order time shift effects. This results in the postulation of many new effects (a *selection* of over 200 time shift effects was studied here), mostly resulting from the systematic combination of variables. This includes many previously unknown effect types: intradynode scattering; the extension of known effects onto the second and later dynodes; interactions of Auger electrons and secondary electrons with above-average energies; circular effects; internally produced photons inducing photoemission on structures other than the cathode; secondary electron production by cathode photoelectrons not on the first dynode; and multiple interactions[8].
Several notations were introduced to label these effects, represent their topology, and facilitate their calculation. Furthermore, the terms used in literature were unified and made unambiguous.

For all effects the occurrence probabilities and the properties of the PPs and, where applicable, the APs were calculated analytically. The variables required for these computations were taken from literature, calculated, or estimated and only in a few cases extracted from measurements. All variables were fixed ab initio except for two which were tuned to improve the match with results for photonic afterpulses. The effect probability was determined by the product of all factors contributing to cascade generation. If more than one particle of the electron cascade can undergo an interaction, the number of trials was taken into account. The influence of the detection threshold and the effect of charge variation on the detectable fraction were considered as well.

[7] For PPs the transit time distribution and charge–time distribution, for EAPs the delay distribution and charge–delay distribution.
[8] Apart from two effects which had been postulated before.

To obtain the charge, the electron impact energy was calculated based on preceding interactions. In combination with the secondary emission yield from literature, this allows to determine the number of secondary electrons produced in an interaction.
The interstage transit time was calculated analytically using approximated electric fields. For this it was necessary to derive formulas for the reverse motion against the field for linear and quadratic potentials and depending on the initial energy.
In addition, the dependence of the charge and charge fraction on the time/delay for PPs and APs was determined qualitatively for all effects. These functions were also calculated exemplarily for one effect, which showed very good agreement with observations considering the used approximations.
The combination of these calculation methods allows to predict the transit time distribution (early and late pulses), delay distribution (EAPs) and the charge distribution (underamplified and overamplified pulses) of a sensor.
The theory was also extended to ionic afterpulses (delay ca. 100 ns–20 µs) as far as possible based on available data.

An extensive literature research showed that of the more than 200 time shift effects anticipated by TST, a subset of 15 had already been described and in part observed before this work, which supports the theory.

To develop and test TST through precise measurements of the timing behavior and very small pulses, an electromagnetically shielded setup with a fast light source was designed and constructed (SPAX, described in 3.2.1). Through the shielding a hardware threshold of only 4.5% pe[9] was reached.
For this setup, a novel photosensor measurement and analysis method was developed (PEST). This combines waveform acquisition[10] with pulse shape analysis to find pileup by fitting the pulses with an average pulse shape. The devised algorithm reliably recognized all individual pulses with a mean reduced χ^2 of about 1.2–1.3, although only one pulse shape model was used to fit all pulses independent of the effect causing them. Thus, further improvements are certainly feasible. In addition, through pulse fitting it was possible to further reduce the already low threshold to merely 1.4% pe when the waveform contains a pulse exceeding the hardware threshold[11].

A thorough test of TST was carried out for different PMT series at varying voltages with both SPAX and an additional setup (reported in 2.3.6, 2.5.2.2 and 3.2.2) using all available data from a pulse shape analysis with PEST (eight distinct distribution types). This experiment agreed completely with the theory within the error margins expected from the used approximations and the limits of the method (no electron optics simulations was used).
The results were also compared with independent data in form of measured distributions and values from literature. This showed exactly the same effects as here, including the 15 previously reported ones, which proved that the measurements were not flawed.
As a consequence, the theory was assumed to be accurate.

[9] 1 pe denotes the mean pulse charge or height.
[10] Using a flash analog-to-digital converter.
[11] Otherwise, no acquisition is triggered.

For the first time, it is possible to explain the properties of all observed effects in all measured distributions within expected uncertainties. Furthermore, no effects were predicted which did not occur.
Through TST, SPAX and PEST it was shown that EAP production is intrinsically connected to time shift effects. Furthermore, it was possible for the first time to observe and identify effects due to photoemission from the deck, intradynode scattering, secondary electron emission from the deck, and multiple interactions.
For the Hamamatsu R5912 PMT series a total photonic afterpulse rate of 19.7% was measured, which is in agreement with predictions from TST. For this PMT it was also found that the right flank of the main peak in the transit time distribution stems mostly from cathode time spread and incident photons which are reflected back onto the photocathode from internal structures.

TST is directly applicable to all sensors based on vacuum phototubes [12] and could be extended to encompass all photosensors. The implementation of field simulations would allow to further improve the accuracy of predictions and, if necessary, adapt the theory.

Furthermore, the event reconstruction in detectors (e.g. energy, time and spatial resolutions) could be improved by the analysis of photosensor pulse shapes. A fit-based pulse search likely would allow to improve the transit time spread, photodetection efficiency and charge resolution through the recognition of early pulses, late pulses and pileup. The extent of possible improvements still has to be quantified.

Pulse shape

Following the sections on timing and TST, the pulse shape was covered in 2.4 and 3.2.2.
Using PEST, the average pulse shape was found to be strongly affected by pileup from EAPs.
About 63–68% of pulses contain fluctuating pileup (low occurrence rate per pulse), which explains the strong variation of observed PMT pulse shapes. It was described how to measure the average pulse shape without distortions from this pileup source.
In addition, the falling flank of regular pulses is to a large extent composed of invariant pileup (rate $\gg 1$/pulse) from late pulse effects occurring on the middle dynodes, which explains why the shape differs from the rising flank.
For the R5912 the gain would only be 0.62 times as high without fluctuating and invariant pileup, and the fall time amounts to 11.2 ns with and only 4.3 ns without fluctuating pileup. The remaining asymmetry between fall and rise time (3.5 ns with, 2.7 ns without pileup) is due to the higher probability and larger transit time differences of high-rate late pulses vs. early pulses. The measured pulse shapes matched predictions by TST which further corroborates the theory.

Afterpulses

In 2.5.2 all effects causing afterpulses in PMTs are covered. Three groups with different physical origins and delay scales were identified.

[12] For instance MCP-PMTs, HAPDs, HPDs, ABALONE, QUASAR-370, IEBCMOS, and TiPC; see ch. 4 for a description of the respective photosensors and a complete list.

Early afterpulses

EAPs (sec. 2.5.2.2 and 3.2.2.5) occur with delays smaller than about 100 ns and can be subdivided into electronic afterpulses (branching APs) and photonic afterpulses.

With SPAX and PEST for the first time the charge–delay distribution could be measured down to 0.8 ns delay and 1.4% pe charge. Through the extension of both the minimum resolvable delay and the minimum charge, a plethora of previously unknown effects was uncovered. This was made possible through the extreme measures taken to detect small pulses with short delays.
An exceedingly high EAP rate of 109% (R5912; 95–111% for all series and voltages) was observed due to the extraordinarily small threshold. The rate predicted by TST (586%) is even higher. The difference between measurement and prediction is caused by two effects: On the one hand, a large fraction of pulses is not resolvable due to the limited delay resolution (0.7–0.8 ns) of the fitting algorithm. On the other hand, many effects have such high mean occurrence rates per pulse that they must be considered a statistic contribution to the average pulse shape and the gain process.
In the charge–delay distribution of EAPs five regions can be distinguished:

- Electronic afterpulses dominate up to 11 ns.
- This is followed by undersized photonic afterpulses occurring on the first dynode, which prevail up to ca 30 ns.
- After this, regular sized photonic afterpulses on the cathode are most prevalent up to ca. 50 ns.
- Beyond these delays, secondary and tertiary APs (APs from APs) of the former types dominate.
- Around 100 ns the first interdynode ionic afterpulses appear, and in the transition region both these and higher-order EAPs occur.

The highest contribution to the EAP rate comes from very short delays below about 11 ns (electronic afterpulses). Photonic afterpulses constitute 16.4% (TST predicted 16.9%), where most pulses are underamplified and only a few percent show regular charges (around 1 pe).
The large majority of EAP effects was detected unambiguously in this work for the first time. This includes the first observation of the effect types intradynode scattering, photonic afterpulses on the first dynode, and scattering effects from secondary electrons on later dynodes. Many more effects were detected which are difficult to disentangle and would require further analyses for certain verification.

In total, early and ionic afterpulses were found to contribute an additional mean charge of 59–79% pe after each pulse. This is not very far from a continuous discharge (at 100% pe), since afterpulses also induce afterpulses. When combined with the voltage dependence of the AP rates this explains why breakdown occurs at high voltages.

In addition, it was possible to derive the recommended hardware threshold which minimizes EAPs while preserving PPs from measured data.

Ionic afterpulses

In 2.5.2.3 and 3.2.2.5 ionic afterpulses, which have delays from ca. 100 ns to 15–30 µs, were treated.
To disambiguate effects, first a nomenclature including ion type, origin region, impact stage, target element and additional interactions was introduced.
Analytic delay calculations with TST were carried out, whereas the available data was insufficient to also predict charge and probability. Possible new ion scattering effects were analyzed, and formulas for the reverse ion movement in various fields were derived.
By comparison of calculations and measurements, all observed event clusters in each distribution could be explained, although with unresolved ambiguities. It was found that lighter ions produce shorter delays and larger AP charges and also create a higher amount of secondary APs which are mostly undersized.
In the R5912 the peaks for H^+, He^+, Cs^+ and Hg^+ were identified with certainty, N^+, $N_2{}^+$, H_2O^+ and Ar^+ are present with high probability, and $H_2{}^+$ also occurred, but only very rarely. Light ions have a higher probability to leave the dynode stack and create ionic afterpulses. Interestingly, ionic afterpulses which are produced on the cathode and whose ions originate from the interdynode space between the second and third dynode had shorter delays than those from between the cathode and the first dynode. Calculations showed that this is due to their higher initial energy when traversing the large distance from the first dynode to the cathode. Ion backscattering was not unambiguously observed but might occur for light ions.

Late afterpulses

Late afterpulses (2.5.2.4 and 3.2.2.5) have delays even larger than those of ionic afterpulses and only occur very rarely.
To study this elusive effect class, the few available reports were compiled, dedicated measurements were performed and possible origins were analyzed. The excitation of metastable cathode levels appears to be the most likely explanation.

Light emission

Another focus of this work was the study of light emission from PMTs (sec. 2.6.1).

Possible mechanisms and their locations were identified and the predicted cathodoluminescence emission wavelengths for Cs_3Sb dynodes and the bremsstrahlung spectra at the dynodes were calculated.

The light emission of several PMT series was investigated regarding location, spectrum, spectrally resolved location, time, temporally resolved location, and self-feedback. This led to the development of improved light shields by Hamamatsu.
Light emission was observed a) directly from the dynode stack through the gaps between dynodes if these were unshielded and b) from internal reflections in the dynode chain through the cathode.
One type of dynode mount contained ruby which resulted in extremely strong cathodoluminescence emission lines around 700 nm. The probability of inducing photonic afterpulses, however, is greatly suppressed by the low quantum efficiency in this range.

Bremsstrahlung was identified as the dominant light-feedback effect, closely followed by ruby cathodoluminescence where present. Cathodoluminescence of Cs_3Sb was observed within the expected wavelength uncertainties as a secondary effect. Gas excitation/ionization and photoluminescence could also contribute to weaker features in the spectrum.
A video of the light emission was recorded which could be understood with TST calculations.
It was confirmed that photonic afterpulses are created both via internal (reflections in dynode chain) and external photon paths (reflections on glass bulb) from the late dynodes to the cathode.

Failure rate

The failure rate of PMTs was covered in 2.7.3, where three failure modes were identified.
The observed failure rate from Borexino was analyzed and was found to be higher than expected, which likely is due to environmental factors. The rate of failures was significantly lower for the PMTs encapsulated in optical modules, however, which is a strong argument for the use of housings.

Chapter 3: Photomultiplier measurements and selection

After establishing the background knowledge necessary to assess PMT properties, the compliance of available PMT series with LENA requirements was studied in ch. 3.
First, existing series were evaluated based on an in-depth literature research to preselect candidate series in 3.1.
Several requirements were newly identified for LENA or had not been studied in detail for the selected PMT series before, especially EAPs, light emission from PMTs, and time shift effects. For this reason, samples of these series were measured first at an existing setup from the Borexino PMT characterization and then with a specialized improved setup (SPAX) developed at the Technische Universität München for this work (see 3.2). The setups and data analyses were described in 3.2.1 and the obtained results were presented in 3.2.2.

SPAX

SPAX was created by upgrading an existing dark box with a Faraday cage, an optical table and antireflective measures. This allowed to reach a noise level of as low as 240 µV root mean square.
As light sources first a ps diode laser and later several self-built fast LED pulsers were used.
These illuminated the photosensor via a reflective diffuser (Spectralon). The homogeneity of illumination was measured and found to be sufficient. Alternative methods for diffuse illumination in form of a transmissive Spectralon plate and strongly diffusing lens types were also studied shortly.

A PMT holder and a passive magnetic shielding for a 12” PMT were fabricated.
To minimize the time jitter of the setup, an electronics was designed which uses an analog sum to preserve the light source trigger pulse and as few components as possible. The resulting setup jitter and transit time offset were determined.
For ionic and late afterpulse measurements an additional triple coincidence with delay range selection was constructed.

PEST

As measurement evaluation software, the generic waveform analysis software PEST was developed. This provides waveform and timing corrections, noise reduction, and pulse shape analysis with pulse fitting and noise recognition. The algorithm can discern pulses with delays down to 0.7 ns, a minimum pulse height of 260 µV (limited by baseline noise) and charges differing by a factor of 40. The mean reduced χ^2 of the pulse fits ranged from 1.2 to 1.3 for 99.2% up to $> 99.9\%$ of pulses.

Measurement results

These measurements provided important missing data for the comparison of PMT series with requirements. Two series (Hamamatsu R5912 and R11780-HQE) could also be characterized with SPAX for varying voltages which provided additional information regarding PMT theory.

First of all, it was found that an analysis which uses pulse fitting can detect substantially more undersized PPs and EAPs with short delays than a classic analysis, in which most time shift effects are invisible. Furthermore, pulse fitting allows to discriminate pulses from one and two photons with high efficiency and might allow to improve the time and charge resolution through effect identification.
Interestingly, the early and late pulse rates decreased with rising voltage, which was found to comply with theoretical expectations from TST. Based on data, the recommended optimum threshold for early pulse exclusion and PP preservation was determined. This ranged from 14 to 30% pe depending on PMT series and voltage, since most early pulses are undersized.
Pileup from EAPs was clearly visible in the distributions of the pulse shape parameters. The rise time either increased or decreased with rising voltage, which could also be understood with TST.
For EAPs the majority of pulses is underamplified. With growing voltage the delays decrease and the rate rises while the charge rate rises or falls due to a decreasing mean AP charge. The difference between both rate definitions likely stems from the behavior of photonic afterpulses on the cathode. The rate of this EAP type rises only slowly to moderately with voltage, since the photons stem mostly from the middle dynodes.
The pileup probability caused by EAPs amounts to 63–68% for the lowest reachable threshold and rises slowly with growing voltage. At a typical threshold of 0.25 pe pileup still occurs for 19–27% of pulses. The recommended threshold to discriminate EAPs while retaining PPs was found to be 0.25–0.38 pe.

In most measurements no late afterpulses were observed. There was one exception, however: For the R5912 at large voltages a pulse cluster was found around 35 µs. The delay is not as large as for the late afterpulses reported in literature but can only be explained by similar effects, namely the excitation of metastable levels in the cathode, which might have a shorter time scale in this PMT.

Selection

Based on these measurements and a thorough literature research, the compliance of all considered series with the LENA requirements was discussed in 3.3.
The early pulse rate, dark count rate and ionic afterpulse rate proved to be the most critical requirements. For a strict application of the criteria only the ETE[13] 9354, Hamamatsu R7081 and Hamamatsu R12860 pass[14], although it is doubtful for the latter if a sufficiently low dark count rate can be reached. However, for a reliable statement regarding the acceptance or exclusion of a series the measurement of additional samples will be required. For this reason, the Hamamatsu R6594, R5912, R7081-HQE, and R11780-HQE are preliminary candidates as well.
For some requirements values are still missing and potentially no single PMT series fulfills all demands. Therefore, it is recommended to carry out a physics Monte Carlo with the obtained values to determine which series allows to maximize the physics case. Alternatively, a mixed configurations of two or more sensor series/types as in JUNO and THEIA could be used.

Chapter 4: Alternative photosensors

In chapter 4 other photosensor types than PMTs were examined for suitability for LSNDs and similar detectors.

To enable a better comparison, a photosensor classification based on the physical interaction steps during the detection process was developed in 4.1. The detection mechanisms of photosensors as well as the amplification methods and readout techniques employed in them were summarized. A unified hybrid photosensor classification was put forth. Furthermore, direction sensitivity and the possibility of a directionality upgrade through segmented light guides were covered, and wavelength sensitive sensors and their possible drawbacks were discussed.

In 4.2–4.4 the properties of 84 visible/UV-sensitive sensor types and 77 subtypes and series were collected and discussed by means of a comprehensive literature research. To the author's knowledge such an extensive compilation and overview has not been provided so far.

Following this, in 4.5 the photosensors best suited for LSNDs were selected from this extensive list.

[13] Electron Tubes Enterprises Ltd.
[14] The measurements reported here indicate that a higher gain might improve some properties (early pulse rate and transit time spread) but also raises the dark count rate and ionic afterpulse rate and thus likely is not an option.

Application of the LENA requirements (see 4.5.1) excludes the vast majority of sensors. Critical exclusion criteria were discussed (photodetection efficiency, dark count rate and transit time spread eliminated the most types) and sensor classes which are generally well-suited were identified. 15 sensors (including series) were included in the shortlist and their properties were compared to each other and LSND requirements. Of these sensors currently only HAPDs [15] by Hamamatsu – the R12112, 5" HAPD and R12850 [16] – and the IEBCMOS [17] are true candidates to replace PMTs. However, many critical properties are still unknown, especially for the IEBCMOS, and further tests are required to arrive at a conclusive statement.
This confirms that PMTs are in fact very good photosensors. Intriguingly, only sensors based on vacuum phototubes entered the shortlist, since this sensor class combines favorable properties. It should also be noted that some requirements are quite severe (e.g. dark count rate) and their relaxation would enable greater flexibility in the choice of sensors and series.

In addition to the shortlist, a list of 11 sensors and series which show promising developments was compiled in 4.5.2 (watch list). In the near or medium term, these photosensors might fulfill all requirements with very favorable properties. Seven of them could even prove to become game-changers by combining a low position resolution with a small transit time spread, high gain, large sensitive area and a low dark count rate (e.g. the LAPPD [18] variants).

Finally, promising approaches for possible future photosensor developments were discussed in 4.5.3.
Auspicious existing sensor classes and types were identified and suggestions for their further development were made.
Possible new classes and types were analyzed based on combinations of sensor structures or hybrid sensors which are missing in the photosensor classification so far. Most were found to be absent for good reasons, but three novel sensor classes were identified which might warrant further studies.
In addition, two novel amplification structures were proposed which could be able to provide a transit time spread in the single-digit picosecond range: the tynode channel plate and the cynode.
These deliberations were summarized in form of a list of future photosensor types, classes and structures which are recommended for development (wish list).

Chapter 5: Auxiliary equipment for photosensors

In addition to this extensive photosensor research, a complete optical module was developed for LENA to improve PMT properties and enable operation in the detector. These works were summarized in ch. 5.
The developed optical module incorporates a PMT, light concentrator (LC), magnetic shielding, voltage divider and buffer liquid in a pressure housing designed for 13 bar.

[15] Hybrid avalanche photodiodes
[16] Assuming an improved preamplifier.
[17] Intensified electron bombarded CMOS
[18] Large Area Picosecond PhotoDetector

The design was based on a candidate PMT series from an early stage of selection (Hamamatsu R11780-HQE, 12"). For this series an LC optimizing the detector homogeneity and a pressure encapsulations with minimum weight were devised.

Section 5.3.1 treats the development of the LC shape.
Existing methods to calculate the shape of a reflective non-imaging LC based on the geometries of detector and photosensor were reviewed. No method is able to produce ideal 3D LCs given the geometry, properties and constraints of an LSND, be it for the maximization of the detected photon yield [19] or the minimization of its variance. The string method was selected as the simplest and at the same time most promising approach.
To achieve the best detector homogeneity, thus a shape optimization was conducted by varying the assumed shapes of source and receiver and the virtual radius of the source. To this end, the existing detector Monte Carlo of LENA was extended, cross-checked for correctness, and an extensive simulation campaign was carried out.
It was found that the use of an LC with a reduced viewing window (lower virtual radius) compared to the dimensions of the fiducial volume could significantly improve the detector homogeneity (20.9% → 10.8%). On the contrary, different virtual shapes of source and receiver did not notably affect the reachable homogeneity. For the LC shape constructed using the full geometrical information, however, the concentration factor was highest. Therefore, the LC shape of this type with the optimal virtual radius was chosen. This behavior was also observed for the ETE D784 (11"), another early PMT candidate series. Truncation of the concentrator, which reduces manufacturing costs, was found to also slightly improve LC properties.
For the truncated optimal LC of the R11780 also the arrangement of optical modules on the detector walls was optimized. This further improved the homogeneity considerably (10.8% → 6.6%) – the effect was almost as large as using the optimal LC vs. no LC. Through the combination of optimized LC and arrangement, the spatial variation of the detected photon yield could almost be eliminated. For the systematic contribution to the energy resolution values could be reached which were comparable to the statistic contribution at 1 MeV; this might be further improvable still. Combined with a calibration with radioactive sources this should notably lower the reachable energy resolution of the detector.
If the target is to maximize the detected photon yield instead of minimizing its variation, it was found that it is best to omit LCs altogether and maximize the optical coverage. While these studies were performed for a cylindrical detector, this nevertheless indicates that the omission of LCs in JUNO was reasonable.

In 5.3.2 the results of material studies conducted for the LCs were summarized.
As reflective material an acrylic glass substrate coated with a reflective layer and optionally additional protective layers was chosen.
Substrates covered with Ag or Al (bare or with protective layers) were studied regarding their reflectance and chemical compatibility with the buffer liquid. This was carried out by immersing coated test strips in buffer liquid and subjecting them to accelerated aging. The attenuation length of the liquid as well as the reflectance of the strips was measured before and after aging.
Unprotected layers showed signs of a commencing solution after 6–7 a. The proper choice of protective coating eliminated this effect but significantly reduced the

[19] The number of detected photons per MeV energy deposited in the scintillator.

reflectance, which was not notably lowered by aging without the protective layers. An unprotected Ag film had the highest weighted reflectance before and after aging, showed little evidence for solution, and thus was chosen as LC material. The acrylic glass appears to react with the buffer liquid, and thus further tests should be conducted to eliminate other explanations and assess acrylic glass samples from multiple manufacturers. The choice of cleaning agent was also found to affect the samples.

For the pressure housing multiple layouts consisting of a transparent acrylic window and a stainless steel body were studied with finite element analysis in 5.4. A housing with a hemispherical window and a conic stainless steel section proved to be most favorable. The development procedure was established and can now be carried out for other PMTs or photosensors, which might be of relevance to THEIA.

Outlook

Liquid scintillator neutrino detectors continue to provide indispensable contributions to the fields of particle, astroparticle, and geophysics.
With the novel photosensor theory and measurement method, the photodetector classification and research suggestions, and the sensor improvements through equipment which were provided by this work, it might be able to further improve the performance and physics case of LSNDs and similar detectors.
In addition, the comprehensive photosensor research which was presented here might also be relevant beyond the field of physics, since photosensors are ubiquitous in the modern world.

List of Abbreviations

In the following list, letter-by-letter sorting is used. Spaces, hyphens, slashes and colons are ignored for sorting purposes. Symbols precede numbers, which are followed by lower case and then upper case letters. Greek symbols are ordered by the Greek alphabet.
Abbreviated photosensor and gaseous detector types are listed, while the reader is referred to ch. 4 for the meaning of sensor labels which can be considered proper names.
Any abbreviation not found in this list likely is the label of a time shift effect. The corresponding nomenclature is introduced in 2.3.5.1.

γAP	photonic afterpulse
γn1	γAP where photon from dn hits d1
γnk	γAP where photon from dn hits k
δn	gain of dn
μ-PIC	Micro Pixel Chamber
σ_T	TTS using standard deviation of main peak
$0\nu\beta\beta$	neutrinoless double beta decay
a	anode
ADC	analog-to-digital converter
ALD	atomic layer deposition
AP	afterpulse
AP-CTD	afterpulse charge–time distribution
APD	avalanche photodiode
APP	astroparticle physics
APS	average pulse shape; active pixel sensor
AP-TTD	afterpulse transit time distribution
AR	antireflective
ASIC	application-specific integrated circuit
B+G	box and grid
B+L	box and line
BAP	branching afterpulse
Bis-MSB	1,4-Bis(2-methyl-styryl)benzene
BLVST	Baksan large volume scintillation telescope
BNO	Baksan Neutrino Observatory

BSE backscattered electron

BSL backscattering losses

C+L circular and linear focused

CB conduction band

CC charged current; circular cage

CCD charge coupled device

CDD charge–delay distribution

CDF cumulative distribution function

CE collection efficiency

CEνNS coherent elastic neutrino-nucleus scattering

CEC compound elliptic concentrator

CECMA Compound Elliptic Concentrator for Maximum incidence Angle

CECME Compound Elliptic Concentrator for Maximum Extension of light source

CEM channel electron multiplier

CEY combined electron yield

CF constant fraction; circular focused

CFC circular–flat concentrator

CFD constant fraction discriminator

ch channel

CIS CMOS image sensor

CJPL China Jinping Underground Laboratory

CL confidence level; cathodoluminescence

CM coarse mesh

CMOS complementary metal–oxide–semiconductor

CMT cynode multiplier tube

CP charge–parity; capillary plate

CPC compound parabolic concentrator

CPIHMB cyclopentadienyl-iron-hexamethyl-benzene

C/S Cherenkov/ scintillation light

CSC compound string concentrator

CTA Cherenkov Telescope Array

CTD charge–time distribution

CTS cathode time spread

CVD chemical vapor deposition

DC direct current

DCR dark count rate

DD delay distribution

DGPM dual gaseous photomultiplier

DM dark matter

DMP dimethylphtalate

dn n-th dynode

DN dark noise

DPS digital pixel sensor

dpy detected photon yield

dSiPM digital SiPM

DSNB diffuse supernova neutrino background

DS-QUASAR directional-sensitive QUASAR

EAP early afterpulse

EB electron bombardment; electron bombarded

EBCCD electron bombarded CCD

EBCMOS electron bombarded CMOS

EF ethylferrocene

EFR early failure rate

EGADS Evaluating Gadolinium's Action on Detector Systems

e–h pair electron–hole pair

EMCCD electron-multiplying CCD

EMCMOS electron-multiplying CMOS

ENC equivalent noise charge

ENF excess noise factor

EP early pulse

EPE external photoelectric effect

EPP early primary pulse

EPP-CTD early primary pulse charge–time distribution

EPP-TTD early primary pulse transit time distribution

ET Electron Tubes Ltd.

ETD light source emission time distribution

ETE Electron Tubes Enterprises Ltd.

ETS emission time spread

f focusing grid

F excess noise factor

FADC flash analog-to-digital converter

FDD charge fraction–delay distribution

FEA finite element analysis

F_F Fano factor

FF fill factor

FFT fast Fourier transform

FIFO fan-in fan-out

FM fine mesh

FOM figure of merit

FPGA field programmable gate array

fps frames per second

F-R-MHSP flipped reversed MHSP

FV fiducial volume

FWHM full width at half maximum

FWTM full width at tenth of maximum

g multiplier gain

g aluminized glass

G-APD Geiger-mode APD

GD gaseous detector

GEM gas electron multiplier

GID gaseous ionization detector

GMF geomagnetic field

GPGD gaseous photocathode GD

GS gigasamples

Hamamatsu Hamamatsu Photonics K. K.

HAPD hybrid APD

HCAL hadron calorimeter

h_f pulse fitting threshold

HGPM hybrid gaseous photomultiplier

HGS-PD hybrid gas scintillation photodiode

HGS-PGD hybrid gas scintillation PGD

HHPD Hamamatsu hybrid photodetector

HM-PD hybrid MCP photodiode

HMS-PD hybrid MCP scintillator photodiode

HPD hybrid photodiode; hybrid photodetector

HPS hybrid photosensor

HPSAPD hybrid position sensitive avalanche photodiode

HQE high quantum efficiency

h_s instrumental (setup) threshold

HSiPM hybrid SiPM

HS-PD hybrid scintillator photodiode

HS-PGD hybrid scintillator PGD

HS-PMT hybrid scintillator PMT

HS-SiPM hybrid scintillator SiPM

HTGD hole-type gaseous detector

HT-PD hybrid tynode photodiode

HV high voltage

Hyper-K Hyper-Kamiokande

HZC Hainan Zhanchuang Photonics Technology Co., Ltd.

IACT imaging atmospheric Cherenkov telescope

iAP ionic afterpulse

IBD inverse beta decay

IBF ion back-flow

ICCD intensified CCD

ICMOS intensified CMOS

ICS ionization cross-section

ID inner detector

IEBCCD intensified EBCCD

IEBCMOS intensified EBCMOS

IH inverted hierarchy

IHEP Institute of High Energy Physics

iiAP interdynode ionic afterpulse

ik regular iAP: ion created between k and d1 hits k

ikRx predynode iAP (x = k or f): ion created between k and d1 hits k or f

IMB Irvine-Michigan-Brookhaven detector

i:n interdynode iAP where ion created between dn and dn+1 hits dn

INFN Istituto Nazionale di Fisica Nucleare

inRk interdynode iAP where ion created between dn and dn+1 hits k

inRx interdynode iAP (x = dn or stage before it): ion created between dn and dn+1 hits x

IO inverted ordering

IPE internal photoelectric effect

IR infrared

ISEY ion-induced secondary electron yield

ISPA-tube imaging silicon pixel array tube

ixRk cathode iAP (x = k or n): ion created between k and d1 or between dn and dn+1 hits k

ixRn dynode iAP (x = dn or stage after it): ion created between dx and dx+1 hits dn

JNE Jinping Neutrino Experiment

JUNO Jiangmen Underground Neutrino Observatory

k backscatter coefficient

k photocathode

KamLAND Kamioka Liquid-scintillator Antineutrino Detector

KamLAND-Zen KamLAND Zero-Neutrino Double-Beta Decay

KM3NeT Cubic Kilometre Neutrino Telescope

KNO Korean Neutrino Observatory

LAB linear alkylbenzene

LAP late afterpulse

LAPPD Large Area Picosecond PhotoDetector

LAr liquid argon

LAr-TPC liquid argon time projection chamber

LBL long baseline

LBNE-WCD Long-Baseline Neutrino Experiment Water Cherenkov Detector

LC light concentrator

L/D length–diameter ratio

LED light emitting diode; leading edge discriminator

LEM large electron multiplier

LENA Low Energy Neutrino Astronomy

LF linear focused

LIDAR light detection and ranging

LLL low light level

LNGS Laboratori Nazionali del Gran Sasso

LP late pulse

LPGD liquid photocathode GD

LPP late primary pulse

LPP-CTD late primary pulse charge–time distribution

LPP-TTD late primary pulse transit time distribution

LSND liquid scintillator neutrino detector

LSO Lu_2SiO_5:Ce

LXe liquid xenon

ly light yield

LYSO $Lu_{2(1-x)}Y_{2x}SiO_5$

m dynode mount or deck

MAGIC Major Atmospheric Gamma Imaging Cherenkov Telescopes

MaPMT multi-anode PMT

MBE molecular-beam epitaxy

MB-GP micro-blasting glass plate

MC Monte Carlo; metal channel

MCA multichannel analyzer

MCP micro-channel plate

MCP-PMT micro-channel plate PMT

ME multiplication efficiency

MEMPHYS MEgaton Mass PHYSics

MEMS micro-electro-mechanical systems

MGC micro gap chamber

MHSP micro-hole and strip plate

Micromegas micro mesh gaseous structure

MIS metal–insulator–semiconductor

MO mass ordering

MOS	metal–oxide–semiconductor
MPGD	micro pattern gas detector
MPP	Max-Planck-Institut für Physik
MSGC	micro strip gas chamber
MSM	metal–semiconductor–metal
MSW	Mikheyev–Smirnov–Wolfenstein effect
MTDC	multi-hit time-to-digital converter
MWPC	multi-wire proportional chamber
NC	neutral current
NDF	neutral-density filter
NEA	negative electron affinity
NH	normal hierarchy
NIM	Nuclear Instrumentation Module
NIR	near-infrared
NNVT	North Night Vision Technology Co., Ltd.
NO	normal ordering
NP	neopentane
npe	n photoelectrons/ photons
o	occupancy
OC	optical coverage
OD	outer detector; optical density
OM	optical module
p	probability
PC	pseudocumene (1,2,4-trimethylbenzene); primary cluster
PCB	printed circuit board
PD	photodiode
PDA	photodiode array
PDE	photodetection efficiency
PDF	probability density function
pe	photoelectron
PE	pulse efficiency
PEST	Photosensor Evaluation Software Toolkit
PET	positron emission tomography
pey	photoelectron yield
PGD	photosensitive gaseous detector
PHD	pulse height distribution
PI	photoionization
PL	photoluminescence
PMNS matrix	Pontecorvo–Maki–Nakagawa–Sakata matrix
PMT	photomultiplier

PP primary pulse

PPAC parallel plate avalanche chamber

PPO 2,5-diphenyloxazole

PPS passive pixel sensor

PP-TTD primary pulse transit time distribution

PSAPD position sensitive APD

PSD pulse shape discrimination; position sensitive diode

PTFE polytetrafluoro-ethylene (Teflon)

P/V peak-to-valley ratio

QD quasidegenerate hierarchy

QDC charge-to-digital converter

QE quantum efficiency

QIS quanta image sensor

QUPID QUartz Photon Intensifying Detector

R charge resolution; resistance; reliability

R_σ charge resolution (standard deviation)

RETGEM resistive electrode TGEM

REY reflection electron yield

R_F charge resolution (FWHM)

RICH ring-imaging Cherenkov

RMS root mean square

RP regular pulse

RPC resistive plate chamber

RSEY reflection secondary electron yield

S1 prompt scintillation light from the liquid phase

S2 delayed scintillation light from the gas phase

SBA super bialkali

SBL short baseline

sCMOS scientific CMOS

SE secondary electron

SEM scanning electron microscopy

SER single electron response

SEY secondary emission yield

SiEM silicon electron multiplier

SiGHT Silicon Geiger Hybrid Tube

SiPM silicon photomultiplier

SiPMT Silicon Photomultiplier Tube

SMD surface-mount device

sn SE from dn

SN supernova

S/N signal-to-noise ratio

SNO Sudbury Neutrino Observatory

SPAD single photon avalanche diode

SPADA SPAD array

SPAX Shielded Photosensor Analysis eXperiment

spe single photoelectron/ photon

SPGD solid photocathode GD

SQE standard quantum efficiency

S-RETGEM strip RETGEM

Super-K Super-Kamiokande

SWC single wire counter

T transit time; temperature

T_0 transit time of regular pulses (γk)

TAC time-to-amplitude converter

TCP tynode channel plate

TDC time-to-digital converter

TE threshold efficiency

TEA triethylamine

TEY transmission electron yield

T_{FWHM} TTS using FWHM of main peak

TGEM Thick GEM

THCOBRA Thick COBRA

THGEM Thick GEM

TiPC Timed Photon Counter

TMA trimethylamine

TMAE tetrakis (dimethylamino) ethylene

TMP tetramethyl pentane

TMPD tetramethyl-p-phenylenediamine

TMS tetramethyl silane

TPC time projection chamber

T_r relative transit time, $T_r = T - T_0$

TSE time shift effect

TSEY transmission secondary electron yield

TST time shift theory

TTD transit time distribution

TTS transit time spread

TUM Technische Universität München

UBA ultra bialkali

UC Davis University of California, Davis

UCLA University of California, Los Angeles

UHE ultra high energy

UV ultraviolet

VB	valence band; venetian blind
VD	voltage divider
VLPC	visible light photon counter
VPGD	visible light PGD
VPT	vacuum phototriode
VSiPMT	Vacuum Silicon PhotoMultiplier Tube
VUV	vacuum ultraviolet
WbLS	water-based liquid scintillator
WCD	water Cherenkov detector
WLS	wavelength-shifter
XDL	cross delay line
XS	cross strip
YSO	Y_2SiO_5:Ce
Z_{eff}	effective atomic number

List of Figures

List of Tables

Bibliography

[1] P. A. **Zyla**, **et al.**, *Review of particle physics*, Prog. Theor. Exp. Phys., vol. 2020, no. 8, pp. 1–2093, **2020**

[2] L. **Oberauer** and J. **Oberauer**, *Neutrinophysik [Neutrino Physics, in German]*, Springer, Berlin, Heidelberg, ISBN 978-3-662-59334-9, **2019**

[3] W. **Pauli**, *Writings on Physics and Philosophy*, Springer, Berlin, Heidelberg, ISBN 978-3-642-08163-7, **1994**

[4] C. L. **Cowan**, **et al.**, *Detection of the Free Neutrino: a Confirmation*, Sci., vol. 124, no. 3212, pp. 103–104, **1956**

[5] G. **Danby**, **et al.**, *Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos*, Phys. Rev. Lett., vol. 9, no. 1, pp. 36–44, **1962**

[6] K. **Kodama**, **et al.**, *Observation of tau neutrino interactions*, Phys. Lett. B, vol. 504, no. 3, pp. 218–224, **2001**

[7] S. **Schael**, **et al.**, *Precision electroweak measurements on the Z resonance*, Phys. Rep., vol. 427, no. 5-6, pp. 257–454, arXiv:hep-ex/0509008, **2006**

[8] Y. **Fukuda**, **et al.**, *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett., vol. 81, no. 8, pp. 1562–1567, **1998**

[9] Q. R. **Ahmad**, **et al.**, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett., vol. 89, no. 1, p. 011301, arXiv:nucl-ex/0204008, **2002**

[10] I. **Esteban**, **et al.**, *The fate of hints: updated global analysis of three-flavor neutrino oscillations*, J. High Energy Phys., vol. 2020, no. 9, p. 178, arXiv:2007.14792, **2020**

[11] I. **Esteban**, **et al.**, *NuFIT 5.0 [website]*, **2020**

[12] R. **Davis**, *A review of the homestake solar neutrino experiment*, Prog. Part. Nucl. Phys., vol. 32, pp. 13–32, **1994**

[13] Y. **Fukuda**, **et al.**, *Solar Neutrino Data Covering Solar Cycle 22*, Phys. Rev. Lett., vol. 77, no. 9, pp. 1683–1686, **1996**

[14] M. **Agostini**, **et al.**, *Improved measurement of ^{8}B solar neutrinos with 1.5 kt · y of Borexino exposure*, Phys. Rev. D, vol. 101, no. 6, p. 062001, arXiv:1709.00756, **2020**

[15] M. **Agostini**, **et al.**, *Simultaneous precision spectroscopy of pp, ^{7}Be, and pep solar neutrinos with Borexino Phase-II*, Phys. Rev. D, vol. 100, no. 8, p. 082004, **2019**

[16] M. **Agostini**, **et al.**, *Experimental evidence of neutrinos produced in the CNO fusion cycle in the Sun*, Nat., vol. 587, no. 7835, pp. 577–582, **2020**

[17] K. **Hirata**, **et al.**, *Observation of a neutrino burst from the supernova SN1987A*, Phys. Rev. Lett., vol. 58, no. 14, pp. 1490–1493, **1987**

[18] R. M. **Bionta**, **et al.**, *Observation of a neutrino burst in coincidence with supernova 1987A in the Large Magellanic Cloud*, Phys. Rev. Lett., vol. 58, no. 14, pp. 1494–1496, **1987**

[19] E. **Alexeyev**, **et al.**, *Detection of the neutrino signal from SN 1987A in the LMC using the INR Baksan underground scintillation telescope*, Phys. Lett. B, vol. 205, no. 2-3, pp. 209–214, **1988**

[20] M. G. **Aartsen**, **et al.**, *Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector*, Sci., vol. 342, no. 6161, p. 1242856, arXiv:1311.5238, **2013**

[21] M. **Aartsen**, **et al.**, *Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A*, Sci., vol. 361, no. 6398, p. eaat1378, arXiv:1807.08816, **2018**

[22] A. **Gando**, **et al.**, *Reactor on-off antineutrino measurement with KamLAND*, Phys. Rev. D, vol. 88, no. 3, p. 033001, arXiv:1303.4667, **2013**

[23] M. **Agostini**, **et al.**, *Spectroscopy of geoneutrinos from 2056 days of Borexino data*, Phys. Rev. D, vol. 92, no. 3, p. 031101, arXiv:1506.04610, **2015**

[24] A. **Aghamousa**, **et al.**, *The DESI Experiment Part I: Science,Targeting, and Survey Design*, arXiv e-prints, arXiv:1611.00036, **2016**

[25] R. **Laureijs**, **et al.**, *Euclid Definition Study Report*, arXiv e-prints, arXiv:1110.3193, **2011**

[26] P. A. **Abell**, **et al.**, *LSST Science Book, Version 2.0*, arXiv e-prints, arXiv:0912.0201, **2009**

[27] **SKA Organisation**, *SKA Phase 1 Construction Proposal [from website]*, **2020**

[28] K. N. **Abazajian**, **et al.**, *CMB-S4 Science Book, First Edition*, arXiv e-prints, arXiv:1610.02743, **2016**

[29] A. **Kogut**, **et al.**, *The Primordial Inflation Explorer (PIXIE): a nulling polarimeter for cosmic microwave background observations*, J. Cosmol. Astropart. Phys., vol. 2011, no. 07, p. 025, arXiv:1105.2044, **2011**

[30] D. **Baumann**, **et al.**, *CMBPol Mission Concept Study: Probing Inflation with CMB Polarization*, AIP Conf. Proc., pp. 10–120, arXiv:0811.3919, **2008**

[31] J. **Delabrouille**, **et al.**, *Exploring cosmic origins with CORE: Survey requirements and mission design*, J. Cosmol. Astropart. Phys., vol. 2018, no. 04, p. 014, arXiv:1706.04516, **2018**

[32] M. **Aker**, **et al.**, *Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN*, Phys. Rev. Lett., vol. 123, no. 22, p. 221802, arXiv:1909.06048, **2019**

[33] A. A. **Esfahani**, **et al.**, *Determining the neutrino mass with cyclotron radiation emission spectroscopy—Project 8*, J. Phys. G: Nucl. Part. Phys., vol. 44, no. 5, p. 054004, **2017**

[34] L. **Gastaldo**, **et al.**, *The Electron Capture ^{163}Ho Experiment ECHo*, J. Low Temp. Phys., vol. 176, no. 5-6, pp. 876–884, arXiv:1309.5214, **2014**

[35] B. **Alpert**, **et al.**, *HOLMES – The Electron Capture Decay of ^{163}Ho to Measure the Electron Neutrino Mass with sub-eV sensitivity*, Eur. Phys. J. C, vol. 75, no. 3, p. 112, arXiv:1412.5060, **2015**

[36] M. P. **Croce**, **et al.**, *Development of Holmium-163 Electron-Capture Spectroscopy with Transition-Edge Sensors*, J. Low Temp. Phys., vol. 184, no. 3-4, pp. 958–968, arXiv:1510.03874, **2016**

[37] M. **Betti**, **et al.**, *Neutrino physics with the PTOLEMY project: active neutrino properties and the light sterile case*, J. Cosmol. Astropart. Phys., vol. 2019, no. 07, p. 047, arXiv:1902.05508, **2019**

[38] K. J. **Kelly**, **et al.**, *Neutrino mass ordering in light of recent data*, Phys. Rev. D, vol. 103, no. 1, p. 013004, arXiv:2007.08526, **2021**

[39] T. **Adam**, **et al.**, *JUNO Conceptual Design Report*, arXiv e-prints, arXiv:1508.07166, **2015**

[40] B. **Abi**, **et al.**, *Volume I. Introduction to DUNE*, J. Instrum., vol. 15, no. 08, p. T08008, **2020**

[41] K. **Abe**, **et al.**, *Hyper-Kamiokande Design Report*, arXiv e-prints, arXiv:1805.04163, **2018**

[42] K. **Abe**, **et al.**, *Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations*, Nat., vol. 580, no. 7803, pp. 339–344, arXiv:1910.03887, **2020**

[43] M. A. **Acero**, **et al.**, *First measurement of neutrino oscillation parameters using neutrinos and antineutrinos by NOvA*, Phys. Rev. Lett., vol. 123, no. 15, p. 151803, arXiv:1906.04907, **2019**

[44] S. **Adrián-Martínez**, **et al.**, *Letter of intent for KM3NeT 2.0*, J. Phys. G: Nucl. Part. Phys., vol. 43, no. 8, p. 084001, **2016**

[45] M. G. **Aartsen**, **et al.**, *IceCube-Gen2: the window to the extreme Universe*, J. Phys. G: Nucl. Part. Phys., vol. 48, no. 6, p. 060501, arXiv:2008.04323, **2021**

[46] M. **Agostini**, **et al.**, *Probing Majorana neutrinos with double-β decay*, Sci., vol. 365, no. 6460, pp. 1445–1448, arXiv:1909.02726, **2019**

[47] S. I. **Alvis**, **et al.**, *Search for neutrinoless double-β decay in ^{76}Ge with 26 kg yr of exposure from the Majorana Demonstrator*, Phys. Rev. C, vol. 100, no. 2, p. 025501, arXiv:1902.02299, **2019**

[48] A. **Gando**, **et al.**, *Search for Majorana Neutrinos Near the Inverted Mass Hierarchy Region with KamLAND-Zen*, Phys. Rev. Lett., vol. 117, no. 8, p. 082503, arXiv:1605.02889, **2016**

[49] S. **Andringa**, **et al.**, *Current Status and Future Prospects of the SNO+ Experiment*, Adv. High Energy Phys., vol. 2016, pp. 1–21, arXiv:1508.05759, **2016**

[50] G. **Anton**, **et al.**, *Search for Neutrinoless Double-Beta Decay with the Complete EXO-200 Dataset*, Phys. Rev. Lett., vol. 123, no. 16, p. 161802, arXiv:1906.02723, **2019**

[51] J. J. **Gomez-Cadenas**, *Status and prospects of the NEXT experiment for neutrinoless double beta decay searches*, in E. **Auge**, **et al.** (editors), *Proceedings, 54rd Rencontres de Moriond on Electroweak Interactions and Unified Theories (Moriond EW 2019)*, pp. 201–206, ISBN 9791096879120, arXiv:1906.01743, **2019**

[52] C. **Alduino**, **et al.**, *First Results from CUORE: A Search for Lepton Number Violation via $0\nu\beta\beta$ Decay of ^{130}Te*, Phys. Rev. Lett., vol. 120, no. 13, p. 132501, arXiv:1710.07988, **2018**

[53] O. **Azzolini**, **et al.**, *Final Result of CUPID-0 Phase-I in the Search for the ^{82}Se Neutrinoless Double-β Decay*, Phys. Rev. Lett., vol. 123, no. 3, p. 032501, arXiv:1906.05001, **2019**

[54] V. **Alenkov**, **et al.**, *First results from the AMoRE-Pilot neutrinoless double beta decay experiment*, Eur. Phys. J. C, vol. 79, no. 9, p. 791, arXiv:1903.09483, **2019**

[55] R. **Arnold**, **et al.**, *Results of the search for neutrinoless double-β decay in ^{100}Mo with the NEMO-3 experiment*, Phys. Rev. D, vol. 92, no. 7, p. 072011, arXiv:1506.05825, **2015**

[56] M. **Agostini**, **et al.**, *Final Results of GERDA on the Search for Neutrinoless Double-β Decay*, Phys. Rev. Lett., vol. 125, no. 25, p. 252502, **2020**

[57] A. **Aguilar**, **et al.**, *Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam*, Phys. Rev. D, vol. 64, no. 11, p. 112007, arXiv:hep-ex/0104049, **2001**

[58] A. A. **Aguilar-Arevalo**, **et al.**, *Significant Excess of Electronlike Events in the MiniBooNE Short-Baseline Neutrino Experiment*, Phys. Rev. Lett., vol. 121, no. 22, p. 221801, arXiv:1805.12028, **2018**

[59] C. **Giunti** and M. **Laveder**, *Statistical significance of the gallium anomaly*, Phys. Rev. C, vol. 83, no. 6, p. 065504, arXiv:1006.3244, **2011**

[60] T. A. **Mueller**, **et al.**, *Improved predictions of reactor antineutrino spectra*, Phys. Rev. C, vol. 83, no. 5, p. 054615, arXiv:1101.2663, **2011**

[61] P. **Huber**, *Determination of antineutrino spectra from nuclear reactors*, Phys. Rev. C, vol. 84, no. 2, p. 024617, arXiv:1106.0687, **2011**

[62] V. I. **Kopeikin**, **et al.**, *Measurement of the Ratio of Cumulative Spectra of Beta Particles from ^{235}U and ^{239}Pu Fission Products for Solving Problems of Reactor-Antineutrino Physics*, Phys. At. Nucl., vol. 84, no. 1, pp. 1–10, **2021**

[63] V. **Kopeikin**, **et al.**, *Reevaluating reactor antineutrino spectra with new measurements of the ratio between ^{235}U and ^{239}Pu β spectra*, arXiv e-prints, pp. 239–242, arXiv:2103.01684, **2021**

[64] P. A. **Machado**, **et al.**, *The Short-Baseline Neutrino Program at Fermilab*, Annu. Rev. Nucl. Part. Sci., vol. 69, no. 1, pp. 363–387, arXiv:1903.04608, **2019**

[65] S. **Ajimura**, **et al.**, *Technical Design Report (TDR): Searching for a Sterile Neutrino at J-PARC MLF (E56, JSNS2)*, arXiv e-prints, arXiv:1705.08629, **2017**

[66] Y. J. **Ko**, **et al.**, *Sterile Neutrino Search at the NEOS Experiment*, Phys. Rev. Lett., vol. 118, no. 12, p. 121802, arXiv:1610.05134, **2017**

[67] I. **Alekseev**, **et al.**, *Search for sterile neutrinos at the DANSS experiment*, Phys. Lett. B, vol. 787, pp. 56–63, arXiv:1804.04046, **2018**

[68] H. **Almazán**, **et al.**, *Improved sterile neutrino constraints from the STEREO experiment with 179 days of reactor-on data*, Phys. Rev. D, vol. 102, no. 5, p. 052002, **2020**

[69] J. **Ashenfelter**, **et al.**, *The PROSPECT reactor antineutrino experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 922, pp. 287–309, arXiv:1808.00097v1, **2019**

[70] A. P. **Serebrov**, **et al.**, *First Observation of the Oscillation Effect in the Neutrino-4 Experiment on the Search for the Sterile Neutrino*, JETP Lett., vol. 109, no. 4, pp. 213–221, arXiv:1809.10561, **2019**

[71] Y. **Abreu**, **et al.**, *A novel segmented-scintillator antineutrino detector*, J. Instrum., vol. 12, no. 04, p. P04024, arXiv:1703.01683, **2017**

[72] V. N. **Gavrin**, **et al.**, *Neutrino-Oscillation Searches in the Short-Baseline Gallium Experiment BEST-2 with a ^{65}Zn Source*, Phys. At. Nucl., vol. 82, no. 1, pp. 70–76, **2019**

[73] A. **de Gouvêa**, **et al.**, *Fundamental physics with the diffuse supernova background neutrinos*, Phys. Rev. D, vol. 102, no. 12, p. 123012, arXiv:2007.13748, **2020**

[74] J. F. **Beacom** and M. R. **Vagins**, *Antineutrino Spectroscopy with Large Water Čerenkov Detectors*, Phys. Rev. Lett., vol. 93, no. 17, p. 171101, arXiv:hep-ph/0309300v1, **2004**

[75] M. **Askins**, **et al.**, *Theia: an advanced optical neutrino detector*, Eur. Phys. J. C, vol. 80, no. 5, p. 416, arXiv:1911.03501, **2020**

[76] M. **Agostini**, **et al.**, *Comprehensive measurement of pp-chain solar neutrinos*, Nat., vol. 562, no. 7728, pp. 505–510, **2018**

[77] R. **Möllenberg**, *Monte Carlo Study of Solar ^{8}B Neutrinos and the Diffuse Supernova Neutrino Background in LENA*, Ph.D. thesis, Technische Universität München, **2013**

[78] E. **Baracchini**, **et al.**, *PTOLEMY: A Proposal for Thermal Relic Detection of Massive Neutrinos and Directional Detection of MeV Dark Matter*, arXiv e-prints, arXiv:1808.01892, **2018**

[79] J. F. **Beacom**, **et al.**, *Letter of Intent: Jinping Neutrino Experiment*, arXiv e-prints, arXiv:1602.01733, **2016**

[80] G. **Alimonti**, **et al.**, *The Borexino detector at the Laboratori Nazionali del Gran Sasso*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 600, no. 3, pp. 568–593, arXiv:0806.2400, **2009**

[81] D. **Hellgartner**, *Advanced Event Reconstruction in LENA and Precision Attenuation-Length Measurements in Liquid Scintillators*, Ph.D. thesis, Technische Universität München, **2015**

[82] K. **Loo**, *Extending Physics Potential of Large Liquid Scintillator Neutrino Detectors*, Ph.D. thesis, University of Jyväskylä, **2016**

[83] B. **Wonsak**, **et al.**, *Topological track reconstruction in unsegmented, large-volume liquid scintillator detectors*, J. Instrum., vol. 13, no. 07, p. P07005, arXiv:1803.08802, **2018**

[84] M. **Agostini**, **et al.**, *The Pacific Ocean Neutrino Experiment*, Nat. Astron., vol. 4, no. 10, pp. 913–915, arXiv:2005.09493, **2020**

[85] C. E. **Aalseth**, **et al.**, *DarkSide-20k: A 20 tonne two-phase LAr TPC for direct dark matter detection at LNGS*, Eur. Phys. J. Plus, vol. 133, no. 3, p. 131, arXiv:1707.08145v1, **2018**

[86] B. **Abi**, **et al.**, *Volume IV. The DUNE far detector single-phase technology*, J. Instrum., vol. 15, no. 08, p. T08010, **2020**

[87] B. **Abi**, **et al.**, *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics*, arXiv e-prints, arXiv:2002.03005, **2020**

[88] A. **Zani**, *The WArP Experiment: A Double-Phase Argon Detector for Dark Matter Searches*, Adv. High Energy Phys., vol. 2014, p. 205107, **2014**

[89] J. **Calvo**, **et al.**, *Commissioning of the ArDM experiment at the Canfranc underground laboratory: first steps towards a tonne-scale liquid argon time projection chamber for Dark Matter searches*, J. Cosmol. Astropart. Phys., vol. 2017, no. 03, p. 003, arXiv:1612.06375, **2017**

[90] A. **Marchionni**, **et al.**, *ArDM: a ton-scale LAr detector for direct Dark Matter searches*, J. Physics: Conf. Ser., vol. 308, no. 1, p. 012006, arXiv:1012.5967, **2011**

[91] B. **Bottino**, **et al.**, *The DarkSide experiment*, Nuovo Cim. C, vol. 40, no. 1, p. 52, **2017**

[92] D. **Akimov**, **et al.**, *WIMP-nucleon cross-section results from the second science run of ZEPLIN-III*, Phys. Lett. B, vol. 709, no. 1-2, pp. 14–20, **2012**

[93] D. **Akerib**, **et al.**, *The Large Underground Xenon (LUX) experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 704, pp. 111–126, **2013**

[94] B. J. **Mount**, **et al.**, *LUX-ZEPLIN (LZ) Technical Design Report*, arXiv e-prints, arXiv:1703.09144, **2017**

[95] E. **Aprile**, **et al.**, *The XENON1T dark matter experiment*, Eur. Phys. J. C, vol. 77, no. 12, p. 881, arXiv:1708.07051, **2017**

[96] E. **Aprile**, **et al.**, *Projected WIMP sensitivity of the XENONnT dark matter experiment*, J. Cosmol. Astropart. Phys., vol. 2020, no. 11, p. 031, arXiv:2007.08796, **2020**

[97] J. **Aalbers**, **et al.**, *DARWIN: towards the ultimate dark matter detector*, J. Cosmol. Astropart. Phys., vol. 2016, no. 11, p. 017, arXiv:1606.07001, **2016**

[98] M. **Auger**, **et al.**, *The EXO-200 detector, part I: detector design and construction*, J. Instrum., vol. 7, no. 05, p. P05010, arXiv:1202.2192, **2012**

[99] J. B. **Albert**, **et al.**, *Sensitivity and discovery potential of the proposed nEXO experiment to neutrinoless double-β decay*, Phys. Rev. C, vol. 97, no. 6, p. 065503, arXiv:1710.05075, **2018**

[100] X. **Chen**, **et al.**, *PandaX-III: Searching for neutrinoless double beta decay with high pressure ^{136}Xe gas time projection chambers*, Sci. China Physics, Mech. & Astron., vol. 60, no. 6, p. 061011, **2017**

[101] M. **Spurio**, *Probes of Multimessenger Astrophysics*, Astronomy and Astrophysics Library, Springer International Publishing, Cham, ISBN 978-3-319-96853-7, **2018**

[102] I. **Kravchenko**, *Status of RICE and preparations for the next generation radio neutrino experiment in Antarctica*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 692, pp. 233–235, **2012**

[103] T. **Winchen**, **et al.**, *Status of the Lunar Detection Mode for Cosmic Particles of LOFAR*, J. Physics: Conf. Ser., vol. 1181, p. 012077, arXiv:1903.08472, **2019**

[104] C. W. **James**, **et al.**, *Overview of lunar detection of ultra-high energy particles and new plans for the SKA*, EPJ Web Conf., vol. 135, p. 04001, arXiv:1704.05336, **2017**

[105] J. **Álvarez-Muñiz**, **et al.**, *The Giant Radio Array for Neutrino Detection (GRAND): Science and design*, Sci. China Physics, Mech. & Astron., vol. 63, no. 1, p. 219501, arXiv:1810.09994, **2020**

[106] R. **Lahmann**, *Acoustic detection of high energy neutrinos in sea water: status and prospects*, EPJ Web Conf., vol. 135, p. 06001, arXiv:1811.11871, **2017**

[107] R. **Lahmann**, *History of acoustic neutrino detection*, EPJ Web Conf., vol. 216, p. 01001, **2019**

[108] R. **Lahmann**, **et al.**, *Thermo-acoustic sound generation in the interaction of pulsed proton and laser beams with a water target*, Astropart. Phys., vol. 65, pp. 69–79, arXiv:1501.01494, **2015**

[109] W. H. **Trzaska**, **et al.**, *Acoustic detection of neutrinos in bedrock*, EPJ Web Conf., vol. 216, p. 04009, arXiv:1909.00417, **2019**

[110] R. **Lahmann**, *Acoustic Detection of Neutrinos: Review and Future Potential*, Nucl. Part. Phys. Proc., vol. 273-275, pp. 406–413, **2016**

[111] A. **Avrorin**, **et al.**, *Status of the early construction phase of Baikal-GVD*, Nucl. Part. Phys. Proc., vol. 273-275, pp. 314–320, **2016**

[112] J. **Rothe**, **et al.**, *NUCLEUS: Exploring Coherent Neutrino-Nucleus Scattering with Cryogenic Detectors*, J. Low Temp. Phys., vol. 199, no. 1-2, pp. 433–440, **2020**

[113] W. **Maneschg**, *The Status of CONUS – talk at the XXVIII International Conference on Neutrino Physics and Astrophysics (Neutrino 2018), 2018-06-07*, **2018**

[114] V. **Belov**, **et al.**, *The νGeN experiment at the Kalinin Nuclear Power Plant*, J. Instrum., vol. 10, no. 12, p. P12011, **2015**

[115] M. K. **Singh**, **et al.**, *Background rejection of TEXONO experiment to explore the sub-keV energy region with HPGe detector*, Indian J. Phys., vol. 91, no. 10, pp. 1277–1291, **2017**

[116] G. **Agnolet**, **et al.**, *Background studies for the MINER Coherent Neutrino Scattering reactor experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 853, pp. 53–60, **2017**

[117] J. **Billard**, **et al.**, *Coherent neutrino scattering with low temperature bolometers at Chooz reactor complex*, J. Phys. G: Nucl. Part. Phys., vol. 44, no. 10, p. 105101, arXiv:1612.09035, **2017**

[118] D. Y. **Akimov**, **et al.**, *The RED-100 two-phase emission detector*, Instrum. Exp. Tech., vol. 60, no. 2, pp. 175–181, **2017**

[119] A. **Aguilar-Arevalo**, **et al.**, *The CONNIE experiment*, J. Physics: Conf. Ser., vol. 761, no. 1, p. 012057, arXiv:1608.01565, **2016**

[120] R. **Strauss**, **et al.**, *The ν-cleus experiment: a gram-scale fiducial-volume cryogenic detector for the first detection of coherent neutrino–nucleus scattering*, Eur. Phys. J. C, vol. 77, no. 8, p. 506, **2017**

[121] V. B. **Petkov**, **et al.**, *Baksan large volume scintillation telescope: a current status*, J. Physics: Conf. Ser., vol. 1468, p. 012244, **2020**

[122] L. **Oberauer**, **et al.**, *LAGUNA LBNO – Liquid Scintillator Specification Document, version 0.30 [unpublished draft]*, Tech. Rep., Technische Universität München, **2012**

[123] M. **Wurm**, **et al.**, *The next-generation liquid-scintillator neutrino observatory LENA*, Astropart. Phys., vol. 35, no. 11, pp. 685–732, arXiv:1104.5620, **2012**

[124] S. **Lorenz**, *Topological Track Reconstruction in Liquid Scintillator and LENA as a Far-Detector in an LBNO Experiment*, Ph.D. thesis, Universität Hamburg, **2016**

[125] M. **Wurm**, **et al.**, *Low-energy Neutrino Astronomy in LENA*, Phys. Procedia, vol. 61, pp. 376–383, **2015**

[126] A. **Rubbia**, *Final Report Summary - LAGUNA (Design of a pan-European Infrastructure for Large Apparatus studying Grand Unification and Neutrino Astrophysics) [website]*, **2014**

[127] A. **Rubbia**, *Final Report Summary - LAGUNA-LBNO (Design of a pan-European Infrastructure for Large Apparatus studying Grand Unification, Neutrino Astrophysics and Long Baseline Neutrino Oscillations) [website]*, **2015**

[128] **LAGUNA-LBNO Consortium**, *LAGUNA/LAGUNA-LBNO - Design of a pan-European infrastructure for Large Apparatus for Grand Unification, Neutrino Astrophysics, and Long Baseline Neutrino Oscillations [website]*, **2015**

[129] M. **He**, *Double calorimetry system in JUNO*, Radiat. Detect. Technol. Methods, vol. 1, no. 2, p. 21, **2017**

[130] F. **An**, **et al.**, *Neutrino physics with JUNO*, J. Phys. G: Nucl. Part. Phys., vol. 43, no. 3, p. 030401, arXiv:1507.05613, **2016**

[131] A. F. **Tietzsch**, *Development , Installation and Operation of a Container-based Mass Testing System for 20-inch Photomultiplier Tubes for JUNO*, Ph.D. thesis, Eberhard Karls Universität Tübingen, **2020**

[132] H. **Lu** and B. **Yu**, *Status of the JUNO experiment – talk at the Asian Forum for Accelerators and Detectors 2021 (AFAD-2021), 2021-03-16*, **2021**

[133] H. T. J. **Steiger**, *JUNO – Detector Design and Status – talk at the XIX International Workshop on Neutrino Telescopes, 2021-02-22*, **2021**

[134] C. **Cao**, **et al.**, *Mass production and characterization of 3-inch PMTs for the JUNO experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 1005, p. 165347, arXiv:2102.11538, **2021**

[135] A. **Abusleme**, **et al.**, *TAO Conceptual Design Report: A Precision Measurement of the Reactor Antineutrino Spectrum with Sub-percent Energy Resolution*, arXiv e-prints, arXiv:2005.08745, **2020**

[136] D. J. **Dörflinger**, *Relative Light Yield Measurement of Liquid Scintillators for the Future Neutrino Experiments JUNO and THEIA*, Bachelor thesis, Technische Universität München, **2020**

[137] B. **Wonsak**, *Neutrino Physics with THEIA – talk at the RAL Seminar, Universität Hamburg, 2020-11-04*, **2020**

[138] J. **Sawatzki**, **et al.**, *Detecting the diffuse supernova neutrino background in the future water-based liquid scintillator detector Theia*, Phys. Rev. D, vol. 103, no. 2, p. 023021, arXiv:2007.14705, **2021**

[139] J. R. **Alonso**, **et al.**, *Advanced Scintillator Detector Concept (ASDC): A Concept Paper on the Physics Potential of Water-Based Liquid Scintillator*, arXiv e-prints, arXiv:1409.5864, **2014**

[140] J. **Caravaca**, **et al.**, *Cherenkov and scintillation light separation in organic liquid scintillators*, Eur. Phys. J. C, vol. 77, no. 12, p. 811, arXiv:1610.02011, **2017**

[141] J. **Caravaca**, **et al.**, *Characterization of water-based liquid scintillator for Cherenkov and scintillation separation*, Eur. Phys. J. C, vol. 80, no. 9, p. 867, **2020**

[142] M. **Yeh**, **et al.**, *A new water-based liquid scintillator and potential applications*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 660, no. 1, pp. 51–56, **2011**

[143] D. R. **Onken**, **et al.**, *Time response of water-based liquid scintillator from X-ray excitation*, Mater. Adv., vol. 1, pp. 71–76, arXiv:2003.10491, **2020**

[144] J. **Dalmasson**, **et al.**, *Distributed imaging for liquid scintillation detectors*, Phys. Rev. D, vol. 97, no. 5, p. 052006, arXiv:1711.09851, **2018**

[145] J. F. **Beacom**, **et al.**, *Physics prospects of the Jinping neutrino experiment*, Chin. Phys. C, vol. 41, no. 2, p. 023002, **2017**

[146] L. **Wan**, *Jinping Neutrino Experiment*, J. Physics: Conf. Ser., vol. 888, no. 1, p. 012058, **2017**

[147] H. **Wei**, **et al.**, *Discovery potential for supernova relic neutrinos with slow liquid scintillator detectors*, Phys. Lett. B, vol. 769, pp. 255–261, **2017**

[148] B. **Xu**, *Jinping Neutrino Experiment: a Status Report*, J. Physics: Conf. Ser., vol. 1468, p. 012212, **2020**

[149] Z.-y. **Guo**, **et al.**, *Muon flux measurement at China Jinping Underground Laboratory*, Chin. Phys. C, vol. 45, no. 2, p. 025001, **2021**

[150] Z. **Wang**, **et al.**, *Design and analysis of a 1-ton prototype of the Jinping Neutrino Experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 855, pp. 81–87, arXiv:1703.01478, **2017**

[151] I. R. **Barabanov**, **et al.**, *Large-volume detector at the Baksan Neutrino Observatory for studies of natural neutrino fluxes for purposes of geo- and astrophysics*, Phys. At. Nucl., vol. 80, no. 3, pp. 446–454, **2017**

[152] A. F. **Yanin**, **et al.**, *Development of the Scintillation Track Detector Based on Multichannel Photodetectors*, Phys. Part. Nucl., vol. 49, no. 4, pp. 804–812, **2018**

[153] G. F. **Knoll**, *Radiation Detection and Measurement*, John Wiley & Sons, 3rd ed., ISBN 978-0471073383, **2000**

[154] J. Q. **Meindl**, *Reconstruction and Measurement of Cosmogenic Signals in the Neutrino Experiment Borexino*, Ph.D. thesis, Technische Universität München, **2013**

[155] P. **Mosteiro**, *First measurement of pp neutrinos in real time in the Borexino detector*, Ph.D. thesis, Princeton University, **2014**

[156] C. **Arpesella**, **et al.**, *Measurements of extremely low radioactivity levels in BOREXINO*, Astropart. Phys., vol. 18, no. 1, pp. 1–25, **2002**

[157] M. **Goodman** and **Thierry Lasserre**, *Double Chooz: A Search for the neutrino mixing angle* θ_{13}, arXiv e-prints, arXiv:hep-ex/0606025, **2006**

[158] Y. **Abe**, **et al.**, *Indication of Reactor Electron Antineutrino Disappearance in the Double Chooz Experiment*, Phys. Rev. Lett., vol. 108, no. 13, p. 131801, arXiv:1112.6353, **2012**

[159] C. **Aberle**, **et al.**, *Large scale Gd-beta-diketonate based organic liquid scintillator production for antineutrino detection*, J. Instrum., vol. 7, no. 06, p. P06008, arXiv:1112.5941, **2012**

[160] C. **Bauer**, **et al.**, *Qualification tests of 474 photomultiplier tubes for the inner detector of the Double Chooz experiment*, J. Instrum., vol. 6, no. 6, p. P06008, **2011**

[161] T. **Matsubara**, **et al.**, *Evaluation of 400 low background 10-in. photo-multiplier tubes for the Double Chooz experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 661, no. 1, pp. 16–25, arXiv:1104.0786, **2012**

[162] J. **Boger**, **et al.**, *The Sudbury Neutrino Observatory*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 449, no. 1, pp. 172–207, **2000**

[163] E. **Caden**, *Status of the SNO+ Experiment*, arXiv e-prints, arXiv:1711.11094, **2017**

[164] N. **Li**, **et al.**, *Characterization of 3-inch photomultiplier tubes for the JUNO central detector*, Radiat. Detect. Technol. Methods, vol. 3, no. 1, p. 6, **2019**

[165] Y. **Nishimura**, *Large Aperture Photodetectors for Water Cherenkov Detectors – talk at TMEX 2018 WCP: European Workshop on Water Cherenkov Precision Detectors for Neutrino and Nucleon Decay Physics, 2018-09-20*, **2018**

[166] E. **Thébault**, **et al.**, *International Geomagnetic Reference Field: the 12th generation*, Earth, Planets, Space, vol. 67, no. 1, p. 79, **2015**

[167] L. **Wen**, **et al.**, *A quantitative approach to select PMTs for large detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 947, p. 162766, arXiv:1903.12595, **2019**

[168] J. P. **Dakin** and R. G. W. **Brown** (editors), *Handbook of Optoelectronics: Concepts, Devices, and Techniques (Volume One)*, CRC Press, Boca Raton, 2nd ed., ISBN 9781315157009, **2017**

[169] **Hamamatsu Photonics K.K. Editorial Committee**, *Photomultiplier Tubes - Basics and Applications*, Hamamatsu Photonics K.K. Electron Tube Division, 4th ed., **2017**

[170] A. G. **Wright**, *The Photomultiplier Handbook*, Oxford University Press, ISBN 9780199565092, **2017**

[171] **ET Enterprises Limited**, *Understanding Photomultipliers*, ET Enterprises Limited, Uxbridge, UK, upmt/11 ed., **2011**

[172] S.-O. **Flyckt** and C. **Marmonier**, *Photomultiplier tubes – principles & applications*, Photonis, Brive, France, **2002**

[173] **Burle Industries Inc.**, *Photomultiplier Handbook*, Burle Industries Inc., **1980**

[174] R. **Mirzoyan**, **et al.**, *Very high quantum efficiency PMTs with bialkali photo-cathode*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 230–232, **2006**

[175] M. **Kapusta**, **et al.**, *Breakthrough in quantum efficiency of bi-alkali photocathodes PMTs*, in *2007 IEEE Nuclear Science Symposium Conference Record*, pp. 73–77, IEEE, ISBN 978-1-4244-0922-8, **2007**

[176] K. **Nakamura**, **et al.**, *Latest bialkali photocathode with ultra high sensitivity*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 623, no. 1, pp. 276–278, **2010**

[177] R. **Mirzoyan**, **et al.**, *Novel 1.5' Size PMTs of Outstanding Parameters for the CTA Project*, in *Proceedings of International Conference on New Photo-detectors — PoS(PhotoDet2015)*, p. 018, Sissa Medialab, Trieste, Italy, **2016**

[178] Y. **Hotta**, *Hamamatsu Photonics K.K. – talk at LIGHT17, 2017-10-19*, **2017**

[179] **Hamamatsu Photonics K.K. Electron Tube Division**, *Datasheet Photomultiplier Tubes R13478, R13449, R13408, R13089*, **2017**

[180] A. G. **Wright**, *Sources of noise in photomultipliers - Technical reprint RP/068*, Tech. Rep., ET Enterprises Ltd., **2011**

[181] Y. **Yoshizawa**, *Recent Progress in Vacuum Photon Detectors from Hamamatsu – talk at the International Baikal Astroparticle Physics Workshop (APP-2012), 2012-07-27*, **2012**

[182] T. **Toyama**, **et al.**, *Novel Photo Multiplier Tubes for the Cherenkov Telescope Array Project*, arXiv e-prints, arXiv:1307.5463, **2013**

[183] **Photonis**, *Photomultiplier tubes basics*, Photonis, **2002**

[184] **Hamamatsu Photonics K.K. Editorial Committee**, *Photomultiplier Tubes – Basics and Applications*, Hamamatsu Photonics K.K. Electron Tube Division, 3a ed., **2007**

[185] **The LBNE Collaboration**, *The Long Baseline Neutrino Experiment (LBNE) Water Cherenkov Detector (WCD) Conceptual Design Report (CDR)*, arXiv e-prints, arXiv:1204.2295, **2012**

[186] A. **Cormack**, *Photomultiplier Developments at ET Enterprises and ADIT Electron Tubes – talk at the International Baikal Astroparticle Physics Workshop (APP-2012), 2012-07-27*, **2012**

[187] J. A. **Nikkel**, **et al.**, *Demonstration of photomultiplier tube operation at 29 K*, J. Instrum., vol. 2, no. 11, p. P11004, **2007**

[188] A. **Falcone**, **et al.**, *Comparison between large area photo-multiplier tubes at cryogenic temperature for neutrino and rare event physics experiments*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 787, pp. 55–58, **2015**

[189] **Hamamatsu Photonics K.K. - Electron Tube Division**, *Photomultiplier Tubes – Photomultiplier Tubes and Related Products [catalog]*, **2016**

[190] J. **Brack**, **et al.**, *Characterization of the Hamamatsu R11780 12 in. photomultiplier tube*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 712, pp. 162–173, arXiv:1210.2765, **2013**

[191] N. **Dinu**, **et al.**, *Photodetection - Principles, Performance and Limitations – talk at Excellence in Detectors and Instrumentation Technologies 2011 (EDIT 2011), 2011-02-01*, **2011**

[192] C. **Faham**, **et al.**, *Measurements of wavelength-dependent double photoelectron emission from single photons in VUV-sensitive photomultiplier tubes*, J. Instrum., vol. 10, no. 09, p. P09010, **2015**

[193] B. **López Paredes**, **et al.**, *Response of photomultiplier tubes to xenon scintillation light*, Astropart. Phys., vol. 102, pp. 56–66, arXiv:1801.01597, **2018**

[194] C. **Grupen** and I. **Buvat** (editors), *Handbook of Particle Detection and Imaging*, Springer, Berlin, Heidelberg, ISBN 978-3-642-13270-4, **2012**

[195] W. R. **Leo**, *Techniques for Nuclear and Particle Physics Experiments*, Springer, Berlin, Heidelberg, 2nd ed., ISBN 978-3-540-57280-0, **1994**

[196] T. **Francke** and V. **Peskov**, *Position-Sensitive Gaseous Photomultipliers: Research and Applications*, Advances in Chemical and Materials Engineering, IGI Global, ISBN 9781522502425, **2016**

[197] C. **Fontaine** and **Photonis**, *Significant increase of QE is opening new avenues for standard photodetectors – talk at the LIGHT07, 2007-09-23–28*, **2007**

[198] X.-D. **Li**, **et al.**, *The QE numerical simulation of PEA semiconductor photocathode*, Chin. Phys. C, vol. 36, no. 6, pp. 531–537, **2012**

[199] D. **Motta** and S. **Schönert**, *Optical properties of bialkali photocathodes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 539, no. 1-2, pp. 217–235, arXiv:physics/0408075, **2005**

[200] M. E. **Moorhead** and N. W. **Tanner**, *Optical properties of an EMI K2CsSb bialkali photocathode*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 378, no. 1-2, pp. 162–170, **1996**

[201] **ET Enterprises Limited**, *Photomultipliers [catalog]*, **2011**

[202] C. **Ghosh**, *Photoemissive Materials*, in *Physics of Thin Films: Advances in Research and Development*, vol. 12, pp. 53–166, Academic Press, Inc., ISBN 0-12-533012-X, **1982**

[203] P. D. **Townsend**, *Photocathodes - Past performance and future potential*, Contemp. Phys., vol. 44, no. 1, pp. 17–34, **2003**

[204] C. **Starr**, *Biology: Concepts and Applications*, Thomson Brooks/Cole, 6th ed., ISBN 978-0534462260, **2005**

[205] W. **Becker** and A. **Bergmann**, *Detectors for high-speed photon counting*, Tech. Rep., Becker & Hickl GmbH, **2018**

[206] R. **Mirzoyan**, **et al.**, *Enhanced quantum efficiency bialkali photo multiplier tubes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 572, no. 1, pp. 449–453, **2007**

[207] W. M. **Haynes**, **et al.** (editors), *CRC Handbook of Chemistry and Physics*, CRC Press, 97th ed., ISBN 978-1-4987-5429-3, **2017**

[208] A. **Ebina** and T. **Takahashi**, *Transmittance Spectra and Optical Constants of Alkali-Antimony Compounds K_3Sb, Na_3Sb, and Na_2KSb*, Phys. Rev. B, vol. 7, no. 10, pp. 4712–4719, **1973**

[209] R. **Nathan** and C. H. B. **Mee**, *Photoelectric and Related Properties of the Potassium—Antimony—Caesium Photocathode*, Int. J. Electron., vol. 23, no. 4, pp. 349–354, **1967**

[210] T. **Rao** and D. H. **Dowell** (editors), *An Engineering Guide To Photoinjectors*, CreateSpace Independent Publishing Platform, ISBN 978-1481943222, arXiv:1403.7539, **2014**

[211] A. **Ettema** and R. **de Groot**, *Electronic structure of Cs_2KSb and K_2CsSb*, Phys. Rev. B, vol. 66, no. 11, p. 115102, **2002**

[212] **Hamamatsu Photonics K.K. - Electron Tube Division**, *Large Photocathode Area Photomultiplier Tubes [catalog]*, **2015**

[213] **Hamamatsu Photonics K.K. - Electron Tube Division**, *Photomultiplier Tubes and Assemblies for Scintillation Counting & High Energy Physics [catalog]*, **2017**

[214] F. **Wooten**, **et al.**, *Photoemission and electron scattering in Cs_3Bi and Cs_3Sb*, J. Appl. Phys., vol. 44, no. 3, pp. 1112–1117, **1973**

[215] **Hamamatsu Photonics K.K. - Electron Tube Division**, *Photomultiplier Tubes – Photomultiplier Tubes and Related Products [catalog]*, **2020**

[216] J. **Grames**, **et al.**, *Charge and fluence lifetime measurements of a dc high voltage GaAs photogun at high average current*, Phys. Rev. Special Top. - Accel. Beams, vol. 14, no. 4, p. 043501, **2011**

[217] R. **Orito**, **et al.**, *Development of HPD Clusters for MAGIC-II*, arXiv e-prints, arXiv:0907.0865, **2009**

[218] T. **Saito**, **et al.**, *Very high QE HPDs with a GaAsP photocathode for the MAGIC telescope project*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 610, no. 1, pp. 258–261, **2009**

[219] J. **Liu**, *Review of High Quantum Efficiency Large Area Photomultiplier Tubes – talk at the Workshop for Next-Generation Solar Neutrino Detection at JinPing, 2014-06-10*, **2014**

[220] E. **Chevallay**, **et al.**, *Photocathodes tested in the dc gun of the CERN photoemission laboratory*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 340, no. 1, pp. 146–156, **1994**

[221] S. **Hallensleben**, **et al.**, *Limitations on the enhancement of photomultiplier quantum efficiency through multiple total internal reflection*, J. Phys. D: Appl. Phys., vol. 32, no. 5, pp. 623–628, **1999**

[222] S. W. **Lee**, **et al.**, *Optimization of Transmission Mode Metallic (Aluminum) Photocathodes*, Phys. Procedia, vol. 37, pp. 757–764, **2012**

[223] A. **Cormack** and R. **Leclercq**, *The potassium bialkali photocathode : sources and methods of manufacture for pmts – talk at the Second Workshop on Photocathodes: 300-500nm, University of Chicago, 2012-06-29*, **2012**

[224] A. F. **McDonie** and W. K. **Miller**, *Method of making potassium, cesium, rubidium, antimony photocathode – US Patent 4339469*, **1982**

[225] J. **Cao**, *PMT at JUNO – talk at the 7th Open Meeting for the Hyper-Kamiokande Project, Queen Mary University of London, 2016-07-11*, **2016**

[226] J. **Xia**, **et al.**, *A performance evaluation system for photomultiplier tubes*, J. Instrum., vol. 10, no. 03, p. P03023, **2015**

[227] W. **Wang**, **et al.**, *Performance of the 8-in. R5912 photomultiplier tube with super bialkali photocathode*, J. Instrum., vol. 10, no. 08, p. T08001, **2015**

[228] Y. **Nishimura**, *Photodetector Development for Hyper-Kamiokande – talk at the Advances in Neutrino Technology 2014 conference (ANT 2014), 2014-09-22*, **2014**

[229] W. **Spicer** and F. **Wooten**, *Photoemission and photomultipliers*, Proc. IEEE, vol. 51, no. 8, pp. 1119–1126, **1963**

[230] H. **Xie**, **et al.**, *Experimental measurements and theoretical model of the cryogenic performance of bialkali photocathode and characterization with Monte Carlo simulation*, Phys. Rev. Accel. Beams, vol. 19, no. 10, p. 103401, **2016**

[231] D. E. **Persyk**, **et al.**, *Photomultipliers for High-Temperature Applications*, IEEE Trans. Nucl. Sci., vol. 23, no. 1, pp. 185–188, **1976**

[232] B. H. **Candy**, *Photomultiplier characteristics and practice relevant to photon counting*, Rev. Sci. Instrum., vol. 56, no. 2, pp. 183–193, **1985**

[233] Y. **Yoshizawa** and **Hamamatsu Photonics**, *Recent Progress of Photosensor – talk at KEK International SCHOOL, Hamamatsu Photonics Electron Tube Center, 2003-11-17*, **2003**

[234] Y. **Yoshizawa** and **Hamamatsu Photonics**, *Recent Progress in PMTs – talk at the Workshop on Next Generation Nucleon Decay and Neutrino Detectors 2007 (NNN07), 2007-10-05*, **2007**

[235] R. **Mirzoyan**, *Very High QE bialkali PMTs – talk at the LIGHT07, 2007-09-23–28*, **2007**

[236] Y. **Nishimura**, *The Hyper-Kamiokande Photodetector System using new 50 cm Photomultiplier Tubes – talk at New and Enhanced Photosensor Technologies for Underground/underwater Neutrino Experiments (NEPTUNE), Naples, 2018-07-19*, **2018**

[237] K. **Arisaka**, *New trends in vacuum-based photon detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 442, no. 1-3, pp. 80–90, **2000**

[238] D. **Paneque**, **et al.**, *A method to enhance the sensitivity of photomultipliers for air Cherenkov telescopes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 504, no. 1-3, pp. 109–115, **2003**

[239] I. **Al Samarai**, **et al.**, *Scintillator-based hybrid photon detector development for the KM3NeT (km^3-scale) deep sea neutrino telescope*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 602, no. 1, pp. 197–200, **2009**

[240] A. **Braem**, **et al.**, *The X-HPD—conceptual study of a large spherical hybrid photodetector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 570, no. 3, pp. 467–474, **2007**

[241] S. **Yamashita**, **et al.**, *Photocathode, electron tube, and photomultiplier tube – US Patent 2010/0253218 A1*, **2010**

[242] B. **Combettes**, *PHOTONIS tubes for large scale experiments – talk at the International Workshop on Next Nucleon Decay and Neutrino Detectors 2008 (NNN08), 2008-09-12*, **2008**

[243] S. **Rossnagel**, **et al.** (editors), *PVD for Microelectronics: Sputter Desposition to Semiconductor Manufacturing, Volume 26*, Academic Press, ISBN 9780080542928, **1998**

[244] R. **Barbier**, *Tutorial Photodetection – talk at the 6th International Conference on New Developments In Photodetection (NDIP 2011), 2011-07-04*, **2011**

[245] D. **Dornic**, **et al.**, *Characterisation of improved photocathode in large hemispherical photomultiplier*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1 SPEC. ISS., pp. 27–30, **2006**

[246] R. **Abbasi**, **et al.**, *The design and performance of IceCube DeepCore*, Astropart. Phys., vol. 35, no. 10, pp. 615–624, **2012**

[247] R. **Mirzoyan**, *Highest QE's Measured So Far – talk at the Second Workshop on Photocathodes: 300-500nm, University of Chicago, 2012-06-29*, **2012**

[248] D. **Ferenc**, *Development of New Photosensors – talk at the Neutrino Super Beam, Detectors and Proton Decay Workshop, 2004-03-04*, **2004**

[249] B. **Smith**, **et al.**, *Development of hybrid photodetectors using single-crystal III–V photocathodes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 504, no. 1-3, pp. 182–187, **2003**

[250] Y. **Wang**, **et al.**, *A new design of large area MCP-PMT for the next generation neutrino experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 695, pp. 113–117, **2012**

[251] L. **Ren**, *Mass Production of MCP-PMT for JUNO and Development of 20-inch MCP-PMT with TTS Improved – talk at the 5th International Workshop on New Photon-Detectors (PD18), 2018-11-29*, **2018**

[252] F. **Gao**, **et al.**, *Time Performance of large area 20 inch MCP-PMTs*, in *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–5, IEEE, ISBN 978-1-5386-2282-7, **2017**

[253] H. **van der Graaf**, **et al.**, *The Tynode: A new vacuum electron multiplier*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 847, pp. 148–161, **2017**

[254] J. **Orloff** (editor), *Handbook of Charged Particle Optics*, CRC Press, Boca Raton, 2nd ed., ISBN 9781315219882, **2017**

[255] F. **Krumeich**, *Properties of Electrons, their Interactions with Matter and Applications in Electron Microscopy [website]*, **2015**

[256] M. **Willinger**, *Scanning methods in electron microscopy I – Lecture in Modern Methods in Heterogeneous Catalysis Research, Fritz-Haber-Institut, 2013-11-22*, **2013**

[257] T. **Staedler**, *Electron microscope techniques - Scanning Electron Microscope (SEM), Transmission Electron Microscope (TEM) – Lecture Moderne Methoden der Materialcharakterisierung (Modern Methods of Material Characterization), Universität Siegen*, **2016**

[258] S. **Tao**, **et al.**, *Secondary Electron Emission Materials for Transmission Dynodes in Novel Photomultipliers: A Review*, Mater., vol. 9, no. 12, p. 1017, **2016**

[259] D. J. **Morrissey**, *Chap. 9 – Photomultiplier Devices – Lecture Radiation Detection and Measurement, Michigan State University*, **2009**

[260] V. **Taillandier**, *High speed imaging detectors with diamond dynode materials*, Ph.D. thesis, University of Leicester, **2013**

[261] S. **Seibert**, **et al.**, *Characterization of Photomultiplier Tubes for the Long-Baseline Neutrino Experiment – talk at the Advances in Neutrino Technologies 2011 (ANT11), 2011-10-10*, **2011**

[262] **ET Enterprises Limited**, *280 mm (11") photomultiplier D784KFLB provisional data sheet*, **2012**

[263] **ET Enterprises Limited**, *200 mm (8") photomultiplier 9354KB series data sheet*, **2014**

[264] B. **Lubsandorzhiev**, **et al.**, *Photoelectron backscattering in vacuum phototubes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 12–16, **2006**

[265] R. E. **Simon** and B. F. **Williams**, *Secondary-Electron Emission*, IEEE Trans. Nucl. Sci., vol. 15, no. 3, pp. 167–170, **1968**

[266] A. **Shih**, **et al.**, *Secondary electron emission studies*, Appl. Surf. Sci., vol. 111, pp. 251–258, **1997**

[267] S. **Cova**, *Photodetectors PD4 - Photo Multiplier Tubes (PMT) – Lecture Signal Recovery, Politecnico di Milano*, **2014**

[268] L. **Reimer**, *Scanning Electron Microscopy – Physics of Image Formation and Microanalysis*, vol. 45 of *Springer Series in Optical Sciences*, Springer, Berlin, Heidelberg, 2nd ed., ISBN 978-3-540-38967-5, **1998**

[269] A. **Ul-Hamid**, *A Beginners' Guide to Scanning Electron Microscopy*, Springer, Cham, ISBN 978-3-319-98482-7, **2018**

[270] H. S. **Leipner** and P. **Werner**, *Wechselwirkung zwischen Elektronen und Materie – Vorlesung Einführung in die Elektronenmikroskopie [Interaction Between Electrons and Matter – Lecture Introduction to Electron Microscopy, in German], Martin-Luther-Universität Halle–Wittenberg*, **2009**

[271] H. **Aksela** and S. **Aksela**, *Auger-electron spectrum of atomic cesium in the 15-75-eV energy range*, Phys. Rev. A, vol. 28, no. 5, pp. 2851–2858, **1983**

[272] P. K. **Chu**, *Auger electron spectroscopy (AES) – Lecture Instrumental Methods of Analysis, City University of Hong Kong*, **2017**

[273] A. **Thompson**, **et al.**, *X-Ray Data Booklet*, Lawrence Berkeley National Laboratory, University of California, 3rd ed., **2009**

[274] E. **Lorenz**, *About measuring the PDE of PMTs – talk at the SiPM Advanced Workshop, Max-Planck-Institute for Physics, Munich, 2013-07-16*, **2013**

[275] O. J. **Smirnov**, **et al.**, *Precision Measurements of Time Characteristics of ETL9351 Photomultipliers*, Instrum. Exp. Tech., vol. 47, no. 1, pp. 69–80, **2004**

[276] P.-A. **Amaudruz**, **et al.**, *In-situ characterization of the Hamamatsu R5912-HQE photomultiplier tubes used in the DEAP-3600 experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 922, pp. 373–384, arXiv:1705.10183, **2019**

[277] A. G. **Wright**, *Method for the determination of photomultiplier collection efficiency, F*, Appl. Opt., vol. 49, no. 11, p. 2059, **2010**

[278] K. **Lung**, **et al.**, *Characterization of the Hamamatsu R11410-10 3-in. photomultiplier tube for liquid xenon dark matter direct detection experiments*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 696, pp. 32–39, arXiv:1202.2628, **2012**

[279] Y. **Nishimura**, *New Large Aperture Photodetectors for a Water Cherenkov Detector – poster at the 15th Vienna Conference on Instrumentation (VCI2019), 2019-02-18–22*, **2019**

[280] S. V. **Polyakov**, *Photomultiplier Tubes*, in *Single-Photon Generation and Detection: Physics and Applications*, vol. 45 of *Experimental Methods in the Physical Sciences*, pp. 69–82, Elsevier Inc., ISBN 9780123876959, **2013**

[281] A. **Braem**, **et al.**, *Design, fabrication and characterization of an 8-in. X-HPD*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 610, no. 1, pp. 61–64, **2009**

[282] D. **Ferenc**, **et al.**, *The Novel ABALONE Photosensor Technology: 4-Year Long Tests of Vacuum Integrity, Internal Pumping and Afterpulsing*, arXiv e-prints, arXiv:1703.04546, **2017**

[283] D. **Ferenc**, **et al.**, *ABALONE™ Photosensors for the IceCube experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 954, p. 161498, arXiv:1810.00280, **2020**

[284] T. **Kaptanoglu**, **et al.**, *Spectral photon sorting for large-scale Cherenkov and scintillation detectors*, Phys. Rev. D, vol. 101, no. 7, p. 072002, arXiv:1912.10333, **2020**

[285] K. **Arisaka**, *Vacuum based Photon Detectors – talk at the 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference (IEEE NSS/MIC 2012), 2012-10-28*, **2012**

[286] R. **Dossi**, **et al.**, *Methods for precise photoelectron counting with photomultipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 451, no. 3, pp. 623–637, **2000**

[287] P. R. **Bevington** and D. K. **Robinson**, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 3rd ed., ISBN 0-07-247227-8, **2003**

[288] A. **Butcher**, **et al.**, *A method for characterizing after-pulsing and dark noise of PMTs and SiPMs*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 875, pp. 87–91, arXiv:1703.06204, **2017**

[289] S. **Jetter**, **et al.**, *PMT waveform modeling at the Daya Bay experiment*, Chin. Phys. C, vol. 36, no. 8, pp. 733–741, **2012**

[290] R. **Abbasi**, **et al.**, *Calibration and characterization of the IceCube photomultiplier tube*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 618, no. 1-3, pp. 139–152, arXiv:1002.2442, **2010**

[291] S. **Aiello**, **et al.**, *Characterisation of the Hamamatsu photomultipliers for the KM3NeT Neutrino Telescope*, J. Instrum., vol. 13, no. 05, p. P05035, **2018**

[292] J. **Haser**, **et al.**, *Afterpulse measurements of R7081 photomultipliers for the Double Chooz experiment*, J. Instrum., vol. 8, no. 04, p. P04029, arXiv:1301.2508, **2013**

[293] S. **Gundacker** and A. **Heering**, *The silicon photomultiplier: fundamentals and applications of a modern solid-state photon detector*, Phys. Medicine & Biol., vol. 65, no. 17, p. 17TR01, **2020**

[294] Z. **Liu**, **et al.**, *Estimating random errors due to shot noise in backscatter lidar observations*, Appl. Opt., vol. 45, no. 18, pp. 4437–4447, **2006**

[295] A. **Biswas** and W. **Farr**, *Laboratory Characterization and Modeling of a Near-Infrared Enhanced Photomultiplier Tube*, Tech. Rep. JPL-IPN-PR-42-152, Jet Propulsion Laboratory, California Institute of Technology, **2003**

[296] J. R. **Prescott**, *A statistical model for photomultiplier single-electron statistics*, Nucl. Instrum. Methods, vol. 39, no. 1, pp. 173–179, **1966**

[297] P. B. **Coates**, *The edge effect in electron multiplier statistics*, J. Phys. D: Appl. Phys., vol. 3, no. 8, pp. 1290–1296, **1970**

[298] G. **Wimmer** and G. **Altmann**, *The Multiple Poisson Distribution, Its Characteristics and a Variety of Forms*, Biom. J., vol. 38, no. 8, pp. 995–1011, **1996**

[299] F. **Neves**, **et al.**, *Calibration of photomultiplier arrays*, Astropart. Phys., vol. 33, no. 1, pp. 13–18, **2010**

[300] A. G. **Wright**, *A Monte Carlo Simulation of Photomultiplier Resolution*, IEEE Trans. Nucl. Sci., vol. 34, no. 1, pp. 414–417, **1987**

[301] F. **Kaether** and C. **Langbrandtner**, *Transit time and charge correlations of single photoelectron events in R7081 photomultiplier tubes*, J. Instrum., vol. 7, no. 09, p. P09002, arXiv:1207.0378, **2012**

[302] P. A. **Bromiley**, *Products and Convolutions of Gaussian Probability Density Functions*, Tech. Rep. Tina Memo No. 2003-003, Division of Informatics, Imaging and Data Sciences, School of Health Sciences, University of Manchester, **2018**

[303] W. **Viehmann**, **et al.**, *Photomultiplier window materials under electron irradiation: fluorescence and phosphorescence*, Appl. Opt., vol. 14, no. 9, p. 2104, **1975**

[304] G. **Collazuol**, *The Silicon Photo-Multiplier – Status and Perspectives – talk at the 15th Vienna Conference on Instrumentation (VCI2019), 2019-02-18*, **2019**

[305] **Hamamatsu Photonics K.K. Solid State Division**, *MPPC – Technical note*, Tech. Rep. Cat. No. KAPD9008E01, Hamamatsu Photonics K.K. Solid State Division, **2021**

[306] N. **Barros**, **et al.**, *Characterization of the ETEL D784UKFLB 11 in. photomultiplier tube*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 852, pp. 15–19, arXiv:1512.06916, **2017**

[307] H. **Zhang**, *The progress of PMT test (the data updated to 2018.12.31) – talk at the 13th JUNO Collaboration Meeting, Shanghai, 2019-01-14*, **2019**

[308] W. H. **Press**, **et al.**, *Numerical Recipes – The Art of Scientific Computing*, Cambridge University Press, 3rd ed., ISBN 978-0521880688, **2007**

[309] D. **Liu**, *PMT Evaluation for the Daya Bay Neutrino Experiment – talk at the 2008 IEEE Nuclear Science Symposium and Medical Imaging Conference (IEEE NSS/MIC 2008),2008-10-23*, **2008**

[310] **Becker & Hickl GmbH**, *How (and why not) to Amplify PMT Signals – Application Note*, Tech. Rep., Becker & Hickl GmbH, Berlin, **2000**

[311] E. **Andres**, *The AMANDA neutrino telescope: principle of operation and first results*, Astropart. Phys., vol. 13, no. 1, pp. 1–20, **2000**

[312] O. Y. **Smirnov**, *Setting of the predefined multiplier gain of a photomultiplier*, Instrum. Exp. Tech., vol. 45, no. 3, pp. 363–370, arXiv:1904.07353, **2002**

[313] S. **Brice**, **et al.**, *Photomultiplier tubes in the MiniBooNE experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 562, no. 1, pp. 97–109, arXiv:1005.3525, **2006**

[314] A. S. **Singh** and A. G. **Wright**, *The determination of photomultiplier temperature coefficients for gain and spectral sensitivity using the photon counting technique*, IEEE Trans. Nucl. Sci., vol. 34, no. 1, pp. 434–437, **1987**

[315] **Hamamatsu Photonics K.K. Electron Tube Division**, *Photomultiplier Tubes and Assemblies – for Scintillation Counting & High Energy Physics [catalog]*, **2009**

[316] W.-Q. **Jiang**, **et al.**, *Suppressing ringing caused by large photomultiplier tube signals*, Chin. Phys. C, vol. 36, no. 3, pp. 235–240, **2012**

[317] **ET Enterprises Limited**, *130mm (5″) photomultiplier 9823B series data sheet*, **2012**

[318] M. **Schultze**, **et al.**, *Delay in Photoemission*, Sci., vol. 328, no. 5986, pp. 1658–1662, **2010**

[319] L. **Gallmann**, **et al.**, *Photoemission and photoionization time delays and rates*, Struct. Dyn., vol. 4, no. 6, p. 061502, **2017**

[320] I. V. **Bazarov**, **et al.**, *Thermal emittance and response time measurements of negative electron affinity photocathodes*, J. Appl. Phys., vol. 103, no. 5, p. 054901, **2008**

[321] R. V. **Poleshchuk**, **et al.**, *An observation of a new class of afterpulses with delay time in the range of 70-200 μs in classical vacuum photomultipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 695, pp. 362–364, **2012**

[322] G. **Hungerford** and D. J. S. **Birch**, *Single-photon timing detectors for fluorescence lifetime spectroscopy*, Meas. Sci. Technol., vol. 7, no. 2, pp. 121–135, **1996**

[323] W.-P. **Huang**, **et al.**, *An accurate measurement of PMT TTS based on the photoelectron spectrum*, Chin. Phys. C, vol. 39, no. 6, p. 066003, arXiv:1409.1146v2, **2015**

[324] G. **Collazuol**, **et al.**, *Single photon timing resolution and detection efficiency of the IRST silicon photo-multipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 581, no. 1-2, pp. 461–464, **2007**

[325] M. **Jiang**, *Performance evaluation of the 50 cm box-and-line dynode photomultiplier tube – talk at the 6th Open Meeting for the Hyper-Kamiokande Project, University of Tokyo, 2015-01-30*, **2015**

[326] A. **Kamiya** and **Hamamatsu Photonics K.K.**, *Photomultiplier Tubes – Recent Trend of PMTs for HEP & Scintillation Application – talk at the Research Center for Electron Photon Science (ELPH), Tohoku University, 2015-03-09*, **2015**

[327] S. **Korpar**, *Advanced vacuum photodetectors and their applications – talk at the 5th International Workshop on New Photon-Detectors (PD18), 2018-11-29*, **2018**

[328] **ET Enterprises Limited**, *51 mm (2″) photomultiplier 9814B series data sheet*, **2016**

[329] A. **Cormack**, *Photomultiplier development at ET Enterprises and ADIT Electron Tubes – talk at the LIGHT11, 2011-11-01*, **2011**

[330] **Hamamatsu Photonics K.K.**, *Voltage divider circuit for E7693-01 MOD (for R11780)*, **2013**

[331] M. **Bonitz**, **et al.**, *Prepulses in photomultipliers*, Nucl. Instrum. Methods, vol. 29, no. 2, pp. 314–318, **1964**

[332] H. R. **Krall**, *Extraneous Light Emission from Photomultipliers*, IEEE Trans. Nucl. Sci., vol. 14, no. 1, pp. 455–459, **1967**

[333] S. S. **Stevens** and J. W. **Longworth**, *Late output pulses from fast photomultipliers*, IEEE Trans. Nucl. Sci., vol. 19, no. 1, pp. 356–359, **1972**

[334] G. **Ranucci**, **et al.**, *Performances of the photomultiplier EMI 9351 for underground physics applications*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 333, no. 2-3, pp. 553–559, **1993**

[335] B. **Fleming**, **et al.**, *Photomultiplier tube testing for the MiniBooNE experiment*, IEEE Trans. Nucl. Sci., vol. 49, no. 3, pp. 984–988, **2002**

[336] A. **Ianni**, **et al.**, *The measurements of 2200 ETL9351 type photomultipliers for the Borexino experiment with the photomultiplier testing facility at LNGS*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 537, no. 3, pp. 683–697, **2005**

[337] M. **Mayer**, *Rutherford backscattering spectrometry (RBS) – Lecture at the Workshop on Nuclear Data for Science and Technology: Materials Analysis, Abdus Salam International Centre for Theoretical Physics (ICTP), 2003-05-28*, **2003**

[338] V. **Singh** and N. **Badiger**, *Study of mass attenuation coefficients, effective atomic numbers and electron densities of carbon steel and stainless steels*, Radioprot., vol. 48, no. 3, pp. 431–443, **2013**

[339] L. **Li**, *Advances and characteristics of high-power diode laser materials processing*, Opt. Lasers Eng., vol. 34, no. 4-6, pp. 231–253, **2000**

[340] B. **Tummers**, *DataThief III [software]*, **2006**

[341] V. **Baglin**, **et al.**, *The secondary electron yield of technical materials and its variations with surface treatment*, in *Proceedings of the 7th European Particle Accelerator Conference (EPAC 2000)*, vol. 1-3, pp. 217–221, **2000**

[342] O. **Kalekin**, *PMT Characterisation for the KM3NeT Project – poster at the 7th International Workshop on Ring Imaging Cherenkov Detectors (RICH 2010), 2010-05-02–07*, **2010**

[343] G. **Alimonti**, **et al.**, *The liquid handling systems for the Borexino solar neutrino detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 609, no. 1, pp. 58–78, **2009**

[344] S. **Pape** and **Physikalisch-Technische Bundesanstalt (PTB)**, *Fluoreszenz bei PTFE-beschichteten Ulbrichtkugeln und Rezepte zur Beseitigung ihres Einflusses [Fluorescence for PTFE-coated integrating spheres and remedies, in German]*, **2010**

[345] B. K. **Lubsandorzhiev**, **et al.**, *Studies of prepulses and late pulses in the 8" electron tubes series of photomultipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 442, no. 1-3, pp. 452–458, **2000**

[346] C. **Bronner**, **et al.**, *Performance of the Hyper-Kamiokande 20" PMT – talk at the 5th International Workshop on New Photon-Detectors (PD18), 2018-11-29*, **2018**

[347] S. **Aiello**, **et al.**, *The Measurement of Late-Pulses and After-Pulses in the Large Area Hamamatsu R7081 Photomultiplier with Improved Quantum-Efficiency Photocathode*, Tech. Rep. INFN/TC-04/11, INFN, **2011**

[348] L. **Bezrukov**, **et al.**, *Large area photodetectors for astroparticle physics Cherenkov arrays: PMTs vs. HPDs*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 639, no. 1, pp. 65–69, **2011**

[349] J. R. N. **Cameron**, *The Photomultiplier Tube Calibration of the Sudbury Neutrino Observatory*, Ph.D. thesis, University of Oxford, **2001**

[350] J. A. **Haser**, *Die Datennahmekette des Double Chooz Experiments und ihre Komponenten – Eigenschaften der Photomultiplier und der Frontend-Elektronik [The Data Acquisition Chain of the Double Chooz Experiment and its Components – Properties of the Photomultipliers and the Front-End Electronics, in German]*, Diploma thesis, Ruprecht-Karls-Universität Heidelberg, **2010**

[351] C. C. **Lo** and B. **Leskovar**, *Afterpulse Time Spectrum Measurement of RCA 8850 Photomultiplier*, IEEE Trans. Nucl. Sci., vol. 30, no. 1, pp. 445–450, **1983**

[352] B. **Lubsandorzhiev**, *Large area photodetectors for the next generation of neutrino detectors – talk at the E15 seminar, Technische Universität München, 2008-12-16*, **2008**

[353] T. **Szczęśniak**, **et al.**, *Fast photomultipliers for TOF PET*, IEEE Trans. Nucl. Sci., vol. 56, no. 1, pp. 173–181, **2009**

[354] S. **Gundacker**, **et al.**, *SiPM time resolution: From single photon to saturation*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 718, pp. 569–572, **2013**

[355] B. K. **Lubsandorzhiev**, *LED based powerful nanosecond light sources for calibration systems of deep underwater neutrino telescopes – talk at the International Workshop on a Very Large Volume Neutrino Telescope for the Mediterranean Sea 2008 (VLVnT08), 2008-04-24*, **2008**

[356] J. **Kapustinsky**, **et al.**, *A fast timing light pulser for scintillation detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 241, no. 2-3, pp. 612–613, **1985**

[357] B. K. **Lubsandorzhiev** and Y. E. **Vyatchin**, *Studies of "Kapustinsky's" light pulser timing characteristics*, J. Instrum., vol. 1, no. 06, p. T06001, arXiv:physics/0410281, **2006**

[358] C. J. **Bruegge**, *Use of Spectralon as a diffuse reflectance standard for in-flight calibration of earth-orbiting sensors*, Opt. Eng., vol. 32, no. 4, p. 805, **1993**

[359] A. **Trunschke**, *UV-vis-NIR Diffuse Reflectance and Photoluminescence Spectroscopy in Heterogeneous Catalysis – Lecture in Modern Methods in Heterogeneous Catalysis Research, Fritz-Haber-Institut,2012-11-23*, **2012**

[360] R. **Alfaro**, **et al.**, *A robot to characterize the photocathode response of the HAWC 8" photomultipliers*, in *Proceedings of the 31st International Cosmic Ray Conference (ICRC 2009)*, University of Łódź and Andrzej Sołtan Institute for Nuclear Studies, ISBN 978-83-929057-1-4, **2009**

[361] N. **Anfimov**, *Large photocathode 20-inch PMT testing methods for the JUNO experiment*, J. Instrum., vol. 12, no. 06, p. C06017, arXiv:1705.05012, **2017**

[362] P. **Vanegas**, **et al.**, *Characterization of the HAWC R5912 photomultipliers*, in *Proceedings of the 33rd International Cosmic Rays Conference*, p. 0809, ISBN 978-85-61516-15-4, **2013**

[363] L. **Steppat**, *Development and Commissioning of the Setup for Masstesting of JUNO PMTs*, Master thesis, Universität Hamburg, **2017**

[364] P. **Agnes**, **et al.**, *Characterization of large area PMTs at cryogenic temperature for rare event physics experiments*, J. Instrum., vol. 9, no. 03, p. C03009, **2014**

[365] T. **Koblesky**, **et al.**, *Cathode position response of large-area photomultipliers under a magnetic field*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 670, pp. 40–44, **2012**

[366] F. **Bauer**, **et al.**, *Performance Study of the New Hamamatsu R9779 and Photonis XP20D0 Fast 2" Photomultipliers*, IEEE Trans. Nucl. Sci., vol. 54, no. 3, pp. 422–426, **2007**

[367] **ORTEC**, *Fast-Timing Discriminator Introduction*, Tech. Rep., ORTEC, **2009**

[368] H. **Akima**, *A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures*, J. ACM, vol. 17, no. 4, pp. 589–602, **1970**

[369] H. O. **Meyer**, *Spontaneous electron emission from a cold surface*, EPL (Europhysics Lett.), vol. 89, no. 5, p. 58001, **2010**

[370] A. **Ankowski**, **et al.**, *Characterization of ETL 9357FLA photomultiplier tubes for cryogenic temperature applications*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 556, no. 1, pp. 146–157, **2006**

[371] M. C. **Tarun**, **et al.**, *Persistent Photoconductivity in Strontium Titanate*, Phys. Rev. Lett., vol. 111, no. 18, p. 187403, **2013**

[372] A. B. **Gerasimov**, **et al.**, *On the physical mechanism of exoelectron emission*, Phys. Status Solidi (a), vol. 35, no. 2, pp. K131–K134, **1976**

[373] P. **Eckert**, **et al.**, *Characterisation studies of silicon photomultipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 620, no. 2-3, pp. 217–226, arXiv:1003.6071, **2010**

[374] A. **Williams** and D. **Smith**, *Afterpulses in liquid scintillation counters*, Nucl. Instrum. Methods, vol. 112, no. 1-2, pp. 131–135, **1973**

[375] N. **Akchurin** and H. **Kim**, *A study on ion initiated photomultiplier afterpulses*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 574, no. 1, pp. 121–126, **2007**

[376] Y. **Cheng**, **et al.**, *The high-speed after-pulse measurement system for PMT*, J. Instrum., vol. 13, no. 05, p. P05014, **2018**

[377] R. **Mirzoyan**, **et al.**, *On the influence of afterpulsing in PMTs on the trigger threshold of multichannel light detectors in self-trigger mode*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 387, no. 1-2, pp. 74–78, **1997**

[378] T. N. K. **Godfrey**, **et al.**, *Satellite Pulses from Photomultipliers*, Phys. Rev., vol. 84, no. 6, pp. 1248–1249, **1951**

[379] R. J. **Lanter** and R. W. **Corwin**, *Spurious Pulses from the Type 5819 Photomultiplier Tube*, Rev. Sci. Instrum., vol. 23, no. 9, pp. 507–508, **1952**

[380] J. S. **Allen**, *After-pulses in photomultiplier tubes*, Tech. Rep. Los Alamos report LA-1459, Los Alamos Scientific Laboratory, University of California, **1952**

[381] G. A. **Morton**, **et al.**, *Afterpulses in Photomultipliers*, IEEE Trans. Nucl. Sci., vol. 14, no. 1, pp. 443–448, **1967**

[382] P. B. **Coates**, *The origins of afterpulses in photomultipliers*, J. Phys. D: Appl. Phys., vol. 6, no. 10, pp. 1159–1166, **1973**

[383] D. **Rapp** and P. **Englander-Golden**, *Total Cross Sections for Ionization and Attachment in Gases by Electron Impact. I. Positive Ionization*, J. Chem. Phys., vol. 43, no. 5, pp. 1464–1479, **1965**

[384] C. M. **Mollo**, *Test and Calibration of a large 3″ Hamamatsu PMT sample – talk at the New and Enhanced Photosensor Technologies for Underground/underwater Neutrino Experiments (NEPTUNE) workshop, Naples, 2018-07-19*, **2018**

[385] P. B. **Coates**, *A theory of afterpulse formation in photomultipliers and the prepulse height distribution*, J. Phys. D: Appl. Phys., vol. 6, no. 16, p. 306, **1973**

[386] J. **Incandela**, **et al.**, *The performance of photomultipliers exposed to helium*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 269, no. 1, pp. 237–245, **1988**

[387] S. Y. **Lai**, **et al.**, *The relationship between electron and ion induced secondary electron imaging: A review with new experimental observations*, Surf. Interface Anal., vol. 8, no. 3, pp. 93–111, **1986**

[388] W. **Lotz**, *Electron-impact ionization cross-sections and ionization rate coefficients for atoms and ions from scandium to zinc*, Z. Phys. A: Hadron. nuclei, vol. 220, no. 5, pp. 466–472, **1969**

[389] W. **Zhong**, *Photomultiplier Tube System for the Daya Bay Experiment*, Nucl. Phys. B - Proc. Suppl., vol. 229-232, p. 438, **2012**

[390] S. **Yin**, **et al.**, *A novel PMT test system based on waveform sampling*, J. Instrum., vol. 13, no. 01, p. T01005, **2018**

[391] M. **Sené**, **et al.**, *Afterpulsing in 130 mm diameter photomultiplier tubes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 254, no. 3, pp. 591–599, **1987**

[392] R. E. W. **Pettifer** and P. G. **Healey**, *Signal induced noise in a 56 TUVP photomultiplier*, J. Phys. E: Sci. Instrum., vol. 7, no. 8, pp. 617–620, **1974**

[393] B. R. **Clemesha**, *Transient noise in photomultiplier tubes*, J. Phys. E: Sci. Instrum., vol. 10, no. 8, pp. 814–816, **1977**

[394] M. **Yamashita**, **et al.**, *Probability and time distribution of afterpulses in GaP first dynode photomultiplier tubes*, Nucl. Instrum. Methods Phys. Res., vol. 196, no. 1, pp. 199–202, **1982**

[395] Y. **Zhao**, *Signal-induced fluorescence in photomultipliers in differential absorption lidar systems*, Appl. Opt., vol. 38, no. 21, p. 4639, **1999**

[396] J. Q. **Meindl**, *Personal communications*, **2010**

[397] M. L. **Ahnen**, **et al.**, *Light Induced Afterpulses in Photomultipliers*, IEEE Trans. Nucl. Sci., vol. 62, no. 3, pp. 1313–1317, **2015**

[398] P. **Adhikari**, **et al.**, *The liquid-argon scintillation pulseshape in DEAP-3600*, Eur. Phys. J. C, vol. 80, no. 4, p. 303, **2020**

[399] H. J. **Queisser**, *Persistent Photoconductivity in Semiconductors*, in *Proceedings of the 17th International Conference on the Physics of Semiconductors*, pp. 1303–1308, Springer, New York, ISBN 0387961089, **1985**

[400] S. M. **Sze**, *Physics of Semiconductor Devices*, John Wiley & Sons, 2nd ed., ISBN 9780471056614, **1981**

[401] P. B. **Coates**, *Noise sources in the C31000D photomultiplier*, J. Phys. E: Sci. Instrum., vol. 4, no. 3, pp. 201–207, **1971**

[402] L. **Kalarasse**, **et al.**, *Optical properties of the alkali antimonide semiconductors Cs_3Sb, Cs_2KSb, CsK_2Sb and K_3Sb*, J. Phys. Chem. Solids, vol. 71, no. 3, pp. 314–322, **2010**

[403] V. A. **Sashin**, **et al.**, *Electronic band structure of beryllium oxide*, J. Physics: Condens. Matter, vol. 15, no. 21, pp. 3567–3581, **2003**

[404] C. E. **Mandeville** and H. O. **Albrecht**, *Luminescence of Beryllium Oxide*, Phys. Rev., vol. 94, no. 2, pp. 494–494, **1954**

[405] M. **Sommer**, **et al.**, *Beryllium oxide as optically stimulated luminescence dosimeter*, Radiat. Meas., vol. 43, no. 2-6, pp. 353–356, **2008**

[406] M. **Petrenko**, **et al.**, *Thermoluminescence and low-temperature luminescence of beryllium oxide*, Radiat. Meas., vol. 90, pp. 14–17, **2016**

[407] G. C. **Brown**, *Fluorescence lifetimes of ruby*, J. Appl. Phys., vol. 35, no. 10, pp. 3062–3063, **1964**

[408] J. D. **Bauer**, **et al.**, *Ruby fluorescence lifetime measurements for temperature determinations at high (p, T)*, High Press. Res., vol. 32, no. 2, pp. 1–8, **2012**

[409] K. **Okamoto**, **et al.**, *A Study to Reduce Backgrounds of the Hyper-Kamiokande Photodetector*, in *Proceedings of the 5th International Workshop on New Photon-Detectors (PD18) – JPS Conference Proceedings*, vol. 27, p. 012007, Journal of the Physical Society of Japan, ISBN 4-89027-137-6, **2019**

[410] G. H. **Zschornack**, *Handbook of X-Ray Data*, Springer, Berlin, Heidelberg, ISBN 978-3-540-28618-9, **2006**

[411] A. **Fridman**, **et al.**, *Plasma Physics and Engineering*, CRC Press, 2nd ed., ISBN 9780429190759, **2011**

[412] R. J. **Ford**, *Calibration of SNO for the Detection of ^{8}B Neutrinos*, Ph.D. thesis, Queen's University, **1998**

[413] D. **D'Angelo**, *Towards the detection of low energy solar neutrinos in BOREXino: data readout, data reconstruction and background identification*, Ph.D. thesis, Technische Universität München, **2006**

[414] W. **Maneschg**, *Low-energy solar neutrino spectroscopy with Borexino - Towards the detection of the solar pep and CNO neutrino flux*, Ph.D. thesis, Ruperto-Carola University of Heidelberg, **2011**

[415] F. **An**, **et al.**, *The detector system of the Daya Bay reactor neutrino experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 811, pp. 133–161, **2016**

[416] A. **Kozlov**, *Search for new physics at KamLAND – talk at the Mount Elbrus Conference: from Deep Underground up to the Sky, 2017-09-13*, **2017**

[417] K. **Abe**, **et al.**, *Direct dark matter search by annual modulation in XMASS-I*, Phys. Lett. B, vol. 759, pp. 272–276, **2016**

[418] M. **Kurz**, *Precision Measurements of Ultra-fast Low Light Level Candidate Sensors for the CTA Project*, Diploma thesis, Technische Universität München, **2011**

[419] M. **Shayduk**, **et al.**, *Light Sensor selection for Cherenkov Telescope Array – talk at the 6th International Conference on New Developments In Photodetection (NDIP 2011), 2011-07-05*, **2011**

[420] M. **Knötig**, *Light Sensor Candidates for the Cherenkov Telescope Array*, Diploma thesis, Technische Universität München, **2012**

[421] T. H. **Maiman**, *Stimulated Optical Radiation in Ruby*, Nat., vol. 187, no. 4736, pp. 493–494, **1960**

[422] W. H. **Fonger** and C. W. **Struck**, *Temperature dependences of Cr^{+3} radiative and nonradiative transitions in ruby and emerald*, Phys. Rev. B, vol. 11, no. 9, pp. 3251–3260, **1975**

[423] J. E. **Sansonetti** and W. C. **Martin**, *Handbook of Basic Atomic Spectroscopic Data*, J. Phys. Chem. Ref. Data, vol. 34, no. 4, pp. 1559–2259, **2005**

[424] P. **Linstrom** and W. **Mallard** (editors), *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, **2018**

[425] **National Instruments Corp., Austin (TX), USA**, *LabVIEW [software]*

[426] F. **Trittmaack**, *Investigations of background light sources in photomultiplier tubes operated at negative high voltage*, Master thesis, Westfälische Wilhelms-Universität Münster, **2018**

[427] G. **Sperrin**, *Personal communications*, **2013**

[428] P.-A. **Amaudruz**, **et al.**, *Design and construction of the DEAP-3600 dark matter detector*, Astropart. Phys., vol. 108, pp. 1–23, arXiv:1712.01982, **2019**

[429] P. **Jagam** and J. **Simpson**, *Measurements of Th, U and K concentrations in a variety of materials*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 324, no. 1-2, pp. 389–398, **1993**

[430] S. **Aiello**, **et al.**, *Influence of the Earth's Magnetic Field on Large Area Photomultipliers*, IEEE Trans. Nucl. Sci., vol. 59, no. 4, pp. 1259–1267, **2012**

[431] E. **Calvo**, **et al.**, *Characterization of large-area photomultipliers under low magnetic fields: Design and performance of the magnetic shielding for the Double Chooz neutrino experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 621, no. 1-3, pp. 222–230, arXiv:0905.3246, **2010**

[432] T. J. **Sumner**, **et al.**, *Conventional magnetic shielding*, J. Phys. D: Appl. Phys., vol. 20, no. 9, pp. 1095–1101, **1987**

[433] **Sekels GmbH**, *Magnetic shielding – Principles, Delivery Programme, Measurement*, Sekels GmbH, Ober-Mörlen, **2020**

[434] E. **Calvo**, **et al.**, *Passive magnetic cylindrical shielding at gauss-range static fields*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 600, no. 3, pp. 560–567, **2009**

[435] P. **DeVore**, **et al.**, *Light-weight flexible magnetic shields for large-aperture photomultiplier tubes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 737, pp. 222–228, arXiv:1309.5415, **2014**

[436] **Aperam Alloys Imphy**, *PERMIMPHY Cold Rolled Strip datasheet*, **2019**

[437] K.-H. **Ackermann**, **et al.**, *The GERDA experiment for the search of $0\nu\beta\beta$ decay in ^{76}Ge*, Eur. Phys. J. C, vol. 73, no. 3, p. 2330, arXiv:1212.4067, **2013**

[438] L. **Oberauer**, *Personal communications*, **2020**

[439] **Hamamatsu Photonics K.K. Electron Tube Division**, *Technical information: Photomultiplier tube R11780 [datasheet]*, **2012**

[440] Y. **Nishimura**, *New 50 cm Photo-Detectors for Hyper-Kamiokande*, in *Proceedings of 38th International Conference on High Energy Physics — PoS(ICHEP2016)*, vol. 282, p. 303, Sissa Medialab, Trieste, Italy, **2017**

[441] P. **Bourgeois** and J. **Va'vra**, *Corrosion of Glass Windows in DIRC PMTs*, Tech. Rep. SLAC-PUB-8877, SLAC National Accelerator Laboratory (SLAC), Menlo Park, CA (United States), **2001**

[442] G. **Bonfini**, *[Borexino] Detector Stability – talk at the Borexino/SOX General Meeting 2014-05-14–15, Rome*, **2014**

[443] D. **Bravo Berguño**, *Precision Background Stability And Response Calibration in Borexino: Prospects For Wideband, Precision Solar ν Spectroscopy and BSM ν Oscillometry Through a Deeper Detector Understanding*, Ph.D. thesis, Virginia Polytechnic Institute and State University, **2016**

[444] C. **Ghiano**, *[Borexino] Detector Stability – talk at the Borexino Analysis and Collaboration Meeting 2019-12, Milano, 2019-12-10*, **2019**

[445] G. **Bonfini**, *[Borexino] Detector Stability – talk at the Borexino General Meeting 2013-05, Virginia Polytechnic Institute and State University*, **2013**

[446] C. **Oppenheimer**, *Determination of the light yield and transparency of LAB-scintillators as a function of the Bis-MSB concentration*, Bachelor thesis, Technische Universität München, **2013**

[447] **Hamamatsu Photonics K.K. Electron Tube Division**, *Large photocathode area photomultiplier tubes [catalog]*, **2019**

[448] E. **Leonora**, *Characterization of two prototypes for a new large area four anode photomultiplier*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 602, no. 1, pp. 217–219, **2009**

[449] A. **Bersani**, *Instrumentation of a new direction sensitive segmented optical module*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 602, no. 1, pp. 213–216, **2009**

[450] A. **Brigatti**, **et al.**, *The photomultiplier tube testing facility for the Borexino experiment at LNGS*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 537, no. 3, pp. 521–536, arXiv:physics/0406106, **2005**

[451] P. **Agnes**, **et al.**, *First results from the DarkSide-50 dark matter experiment at Laboratori Nazionali del Gran Sasso*, Phys. Lett. B, vol. 743, pp. 456–466, **2015**

[452] B. **Herold** and O. **Kalekin**, *Characterisation of PMTs for KM3NeT*, in *Proceedings of the 31st International Cosmic Ray Conference (ICRC 2009)*, University of Łódź and Andrzej Sołtan Institute for Nuclear Studies, ISBN 978-83-929057-1-4, **2009**

[453] P. **Lombardi**, *Laboratory measurements: Photomultipliers and scintillators – talk at the LENA Meeting 2010-07-05–06, INFN Milano*, **2010**

[454] S. **Nakayama**, *Recent progress in the development of Photomultiplier Tubes – talk at the International Conference on New Photo-detectors 2015 (PD15), 2015-07-07*, **2015**

[455] **Hamamatsu Photonics K.K. Electron Tube Division**, *Final test sheet for photomultiplier tube type R11780HQE, serial number ZN0120*, **2014**

[456] D.-H. **Liao**, **et al.**, *Study of TTS for a 20-inch dynode PMT*, Chin. Phys. C, vol. 41, no. 7, p. 076001, **2017**

[457] X.-C. **Lei**, **et al.**, *Evaluation of new large area PMT with high quantum efficiency*, Chin. Phys. C, vol. 40, no. 2, p. 026002, arXiv:1504.03174, **2016**

[458] M. **Ziembicki**, *Photosensors and Front-end Electronics for the Hyper-Kamiokande Experiment – talk at the 10th International Workshop on Ring Imaging Cherenkov Detectors (RICH 2018), 2018-07-31*, **2018**

[459] A. **Yang**, **et al.**, *The study of linearity and detection efficiency for 20" photomultiplier tube*, Radiat. Detect. Technol. Methods, vol. 3, no. 2, p. 11, **2019**

[460] G. **Bellini**, **et al.**, *Final results of Borexino Phase-I on low-energy solar neutrino spectroscopy*, Phys. Rev. D, vol. 89, no. 11, p. 112007, arXiv:1308.0443, **2014**

[461] B. **Neumair**, *Personal communications*, **2020**

[462] **Hamamatsu Photonics K.K. Systems Division**, *Blue Picosecond Light Pulser PLP-02 SLDH-041 datasheet*, **2000**

[463] R. **Brun** and F. **Rademakers**, *ROOT — An object oriented data analysis framework*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 389, no. 1-2, pp. 81–86, **1997**

[464] F. **Rademakers**, **et al.**, *ROOT v6.15/01, 2018-06-19 [software]*, **2018**

[465] H. A. **Wolfsperger**, *Elektromagnetische Schirmung [Electromagnetic Shielding, in German]*, VDI-Buch, Springer, Berlin, Heidelberg, ISBN 978-3-540-76912-5, **2008**

[466] B. K. **Lubsandorzhiev**, *Extensive studies of Ultra Bright LEDs and Light Sources Based on Them – talk at the LIGHT11, 2011-10-31*, **2011**

[467] D. **Marcuse**, *Light Transmission Optics*, Van Nostrand Reinhold Company, 2nd ed., ISBN 0442263090, **1982**

[468] H. **Kellermann**, *Präzise Vermessung der fokussierten Reflektivität der MAGIC Teleskopspiegel und Charakterisierung des hierfür verwendeten diffusen Reflektors [Precise Measurements of the Focused Reflectivity of the MAGIC Telescope Mirrors and Characterization of the Diffuse Reflector Used for This, in German]*, Diploma thesis, Hochschule München, **2011**

[469] M. **Nöbauer**, *Entwicklung eines Testaufbaus für Photomultiplier im Rahmen des LENA-Projekts [Development of a Photomultiplier Test Setup for the LENA Project, in German]*, Diploma thesis, Technische Universität München, **2011**

[470] **Hamamatsu Photonics K.K. Electron Tube Division**, *Large photocathode area photomultiplier tubes [catalog]*, **2008**

[471] A. R. **Back**, **et al.**, *Accelerator Neutrino Neutron Interaction Experiment (ANNIE): Preliminary Results and Physics Phase Proposal*, arXiv e-prints, arXiv:1707.08222, **2017**

[472] C. **Joram**, *Large area hybrid photodiodes*, Nucl. Phys. B - Proc. Suppl., vol. 78, no. 1-3, pp. 407–415, **1999**

[473] S. **Donati**, *Photodetectors: Devices, Circuits and Applications*, Prentice Hall, 1st ed., ISBN 978-0130203373, **2000**

[474] R. A. **Yotter** and D. M. **Wilson**, *A review of photodetectors for sensing light-emitting reporters in biological systems*, IEEE Sens. J., vol. 3, no. 3, pp. 288–303, **2003**

[475] C. **Schäfer**, *Particle Interactions with Matter and Detector Design Prinicples – Lecture 3: More interactions & photodetectors, Vilniaus universitetas*, **2016**

[476] **Hamamatsu Photonics K.K. Solid State Division**, *Hamamatsu Opto-semiconductors – Condensed Catalog*, **2020**

[477] **Hamamatsu Photonics K.K.**, *Hamamatsu Opto-semiconductor Handbook*, Hamamatsu Photonics K.K., **2014**

[478] N. **Britun** and A. **Nikiforov** (editors), *Photon Counting - Fundamentals and Applications*, InTech, ISBN 978-953-51-4081-8, **2018**

[479] J. **Chadwick** and M. **Goldhaber**, *A 'Nuclear Photo-effect': Disintegration of the Diplon by γ-Rays*, Nat., vol. 134, no. 3381, pp. 237–238, **1934**

[480] J. **Fraden**, *Handbook of Modern Sensors – Physics, Designs, and Applications*, Springer, New York, 4th ed., ISBN 978-1-4419-6465-6, **2010**

[481] A. **Grinberg**, *Theory of the Photoelectric and Photomagnetic Effect Produce by Light Pressure*, Sov. J. Exp. Theor. Phys., vol. 31, no. 3, p. 531, **1970**

[482] J. **Frejlich**, *Photorefractive Materials: Fundamental Concepts, Holographic Recording and Materials Characterization*, John Wiley & Sons, Inc., Hoboken, NJ, USA, ISBN 9780470089064, **2007**

[483] P. **Günter** and J.-P. **Huignard** (editors), *Photorefractive Materials and Their Applications 3 – Applications*, vol. 115 of *Springer Series in Optical Sciences*, Springer, New York, ISBN 978-0-387-34728-8, **2007**

[484] K. C. **Liddiard**, *The active microbolometer: a new concept in infrared detection*, in D. **Abbott**, **et al.** (editors), *Proc. SPIE 5274, Microelectronics: Design, Technology, and Packaging*, p. 227, **2004**

[485] G. S. **Buller** and R. J. **Collins**, *Single-photon generation and detection*, Meas. Sci. Technol., vol. 21, no. 1, p. 012002, **2010**

[486] D. D. E. **Martin** and P. **Verhoeve**, *Observing Photons in Space*, in M. C. E. **Huber**, **et al.** (editors), *Observing Photons in Space*, Springer, New York, ISBN 978-1-4614-7803-4, **2013**

[487] P. K. **Day**, **et al.**, *A broadband superconducting detector suitable for use in large arrays*, Nat., vol. 425, no. 6960, pp. 817–821, **2003**

[488] A. **Fleischmann**, **et al.**, *Metallic Magnetic Calorimeters*, in *Cryogenic Particle Detection*, pp. 151–216, Springer, Berlin, Heidelberg, ISBN 978-3-540-31478-3, **2005**

[489] J. **Fraden**, *Handbook of Modern Sensors – Physics, Designs, and Applications*, Springer, Cham, 5th ed., ISBN 978-3-319-19303-8, **2016**

[490] S. N. **Thakur**, *Optogalvanic spectroscopy and its applications*, in *Molecular and Laser Spectroscopy*, pp. 217–244, Elsevier, ISBN 9780128188705, **2020**

[491] S. B. **Rai** and D. K. **Rai**, *Optogalvanic Spectroscopy*, Proc. Indian Natl. Sci. Acad., vol. 62, no. 1A, pp. 475–512, **1996**

[492] **Hamamatsu Photonics K.K. Electron Tube Division**, *Hamamatsu Opto-Galvanic Sensor: Laser Galvatron L2783 Series [datasheet]*, **2008**

[493] K. **Iizuka**, *Engineering Optics*, Springer, Cham, 4th ed., ISBN 978-3-319-69251-7, **2019**

[494] M. **Land**, *Photoreception – Structure And Function Of Photoreceptors [version 2020-06-01]*, Encyclopædia Britannica, **2020**

[495] A. F. **Holleman**, **et al.**, *Lehrbuch der Anorganischen Chemie [Textbook of Inorganic Chemistry, in German]*, Walter de Gruyter, 101st ed., ISBN 978-3110126419, **1995**

[496] C. **Joram**, *Hybrid Photon Detectors (HPD) & Multi Anode Photo Multiplier Tubes (MAPMT) – talk at the Innovative Detectors for Supercolliders workshop, Erice, 2003-10-03*, **2003**

[497] M. **Minot**, **et al.**, *Large Area Picosecond Photodetector (LAPPD™) - Pilot production and development status*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 936, pp. 527–531, **2019**

[498] S. **Nishida**, **et al.**, *Development of an HAPD with 144 channels for the aerogel RICH of the Belle upgrade*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 595, no. 1, pp. 150–153, **2008**

[499] A. **Ball**, **et al.**, *C2GT: intercepting CERN neutrinos to Gran Sasso in the Gulf of Taranto to measure θ_{13}*, Eur. Phys. J. C, vol. 49, no. 4, pp. 1117–1142, **2007**

[500] G. **Anton**, **et al.**, *A hybrid photodetector using the Timepix semiconductor assembly for photoelectron detection*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 602, no. 1, pp. 205–208, **2009**

[501] A. **Suzuki**, **et al.**, *Novel large aperture EBCCD*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 628, no. 1, pp. 260–263, **2011**

[502] R. **Bagduev**, **et al.**, *The optical module of the Baikal deep underwater neutrino telescope*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 420, no. 1-2, pp. 138–154, **1999**

[503] A. **Braem**, **et al.**, *Development, fabrication and test of a highly segmented hybrid photodiode*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 478, no. 1-2, pp. 400–403, **2002**

[504] E. **Pantic**, **et al.**, *Status of QUPID, a novel photosensor for noble liquid detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 695, pp. 121–124, **2012**

[505] F. **Barbato**, *The 2-inches VSiPMT industrial prototype – talk at the 15th Vienna Conference on Instrumentation (VCI2019), 2019-02-21*, **2019**

[506] T. **Gys**, **et al.**, *A new position-sensitive photon detector based on an imaging silicon pixel array (ISPA-tube)*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 355, no. 2-3, pp. 386–389, **1995**

[507] X. **Michalet**, **et al.**, *Development of new photon-counting detectors for single-molecule fluorescence microscopy*, Philos. Trans. R. Soc. B: Biol. Sci., vol. 368, no. 1611, p. 20120035, **2013**

[508] J. B. **Pawley**, *Appendix 3 – More Than You Ever Really Wanted to Know About Charge-Coupled Devices*, in J. B. **Pawley** (editor), *Handbook Of Biological Confocal Microscopy*, Springer, Boston, MA, 3rd ed., ISBN 978-0-387-25921-5, **2006**

[509] **Hamamatsu Photonics K.K. Solid State Division**, *Technical Information SD-37 - Characteristics and use of Charge amplifier*, **2001**

[510] J. **Ma**, **et al.**, *Photon-number-resolving megapixel image sensor at room temperature without avalanche gain*, Opt., vol. 4, no. 12, p. 1474, **2017**

[511] S. **Sze** and K. N. **Kwok**, *Physics of Semiconductor Devices*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 3rd ed., ISBN 9780470068328, **2006**

[512] R. **Vaz**, **et al.**, *Measurement of the secondary electron emission from CVD diamond films using phosphor screen detectors*, J. Instrum., vol. 10, no. 03, p. P03004, **2015**

[513] K. **Kuroda**, **et al.**, *New type of position sensitive photomultiplier*, Rev. Sci. Instrum., vol. 52, no. 3, pp. 337–346, **1981**

[514] V. **Agoritsas**, **et al.**, *A new type of position-sensitive electron multiplier*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 277, no. 1, pp. 237–241, **1989**

[515] H. **Kolanoski** and N. **Wermes**, *Teilchendetektoren [Particle Detectors, in German]*, Springer, Berlin, Heidelberg, ISBN 978-3-662-45349-0, **2016**

[516] D. A. **Orlov**, **et al.**, *High-detection efficiency MCP-PMTs with single photon counting capability for LIDAR applications*, in N. **Karafolas**, **et al.** (editors), *International Conference on Space Optics — ICSO 2018*, vol. 11180, p. 1118031, SPIE, ISBN 9781510630772, **2019**

[517] P. **Chen**, **et al.**, *Design of the large area MCP-PMT and a novel bowl-shaped MCP – poster at the 15th Vienna Conference on Instrumentation (VCI2019), 2019-02-18–22*, **2019**

[518] V. **Prodanovic**, *Ultra-thin MEMS fabricated tynodes for electron multiplication*, Ph.D. thesis, Delft University of Technology, **2019**

[519] D. **Ferenc** and E. **Lorenz**, *Novel photosensors for neutrino detectors and telescopes*, Earth, Moon, Planets, vol. 100, no. 3-4, pp. 241–257, **2007**

[520] S. **Masciocchi**, *Gaseous detectors – Lecture at the 39th Heidelberg Physics Graduate Days, HGSFP, Heidelberg, 2017-10-10*, **2017**

[521] L. **Cerrito**, *Radiation and Detectors – Introduction to the Physics of Radiation and Detection Devices*, Graduate Texts in Physics, Springer, Cham, ISBN 978-3-319-53179-3, **2017**

[522] F. **Sauli**, *Principles of Operation of Multiwire Proportional and Drift Chambers – Lectures given in the Academic Training programm of CERN 1975-1976 (CERN 77-09)*, CERN, Geneva, **1977**

[523] A. **Lyashenko**, **et al.**, *Efficient ion blocking in gaseous detectors and its application to gas-avalanche photomultipliers sensitive in the visible-light range*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 598, no. 1, pp. 116–120, arXiv:0804.4396, **2009**

[524] H. **Kumagai**, **et al.**, *Development of Parallel Plate Avalanche Counter (PPAC) for BigRIPS fragment separator*, Nucl. Instrum. Methods Phys. Res. Sect. B, vol. 317, pp. 717–727, **2013**

[525] G. **Cattani**, *The Resistive Plate Chambers of the ATLAS experiment: performance studies*, J. Physics: Conf. Ser., vol. 280, no. 1, p. 012001, **2011**

[526] F. **Angelini**, **et al.**, *The micro-gap chamber*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 335, no. 1-2, pp. 69–77, **1993**

[527] A. **Ochi**, **et al.**, *Development of a Micro Pixel Chamber for the ATLAS Upgrade*, Phys. Procedia, vol. 37, pp. 554–560, **2012**

[528] F. **Sauli**, *The gas electron multiplier (GEM): Operating principles and applications*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 805, pp. 2–24, **2016**

[529] R. **Chechik** and A. **Breskin**, *Advances in gaseous photomultipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 595, no. 1, pp. 116–127, arXiv:0807.2086, **2008**

[530] V. **Peskov**, **et al.**, *Progress in the development of photosensitive GEMs with resistive electrodes manufactured by a screen printing technology*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 610, no. 1, pp. 169–173, **2009**

[531] F. D. **Amaro**, **et al.**, *The Thick-COBRA: a new gaseous electron multiplier for radiation detectors*, J. Instrum., vol. 5, no. 10, p. P10002, **2010**

[532] M. **Iodice**, *Micromegas detectors for the Muon Spectrometer upgrade of the ATLAS experiment*, J. Instrum., vol. 10, no. 02, p. C02026, **2015**

[533] I. **Giomataris**, **et al.**, *Micromegas in a bulk*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 560, no. 2, pp. 405–408, **2006**

[534] S. **Franchino**, *Micro Pattern Gas Detectors (MPGD) fabrication processes – Lecture at the HighRR Lecture Week, Universität Heidelberg, 2016-04-13*, **2016**

[535] A. **Ochi**, **et al.**, *A new design of the gaseous imaging detector: Micro Pixel Chamber*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 471, no. 1-2, pp. 264–267, **2001**

[536] A. **Ochi**, *MPGD R&D activities in JAPAN*, J. Physics: Conf. Ser., vol. 469, no. 1, pp. 1–6, **2013**

[537] F. **Yamane**, **et al.**, *Development of μ-PIC with resistive electrodes using sputtered carbon*, EPJ Web Conf., vol. 174, p. 04001, **2018**

[538] P. **Martinengo**, **et al.**, *Position Sensitive Gaseous Photomultipliers*, arXiv e-prints, arXiv:1008.1327, **2010**

[539] R. **Bellazzini**, **et al.**, *Imaging with the invisible light*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 581, no. 1-2, pp. 246–253, **2007**

[540] K. **Terasaki**, *Study of Ion Back Flow suppression with thick COBRA GEM*, J. Instrum., vol. 9, no. 03, p. C03014, **2014**

[541] G. **Charpak**, **et al.**, *Micromegas, a multipurpose gaseous detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 478, no. 1-2, pp. 26–36, **2002**

[542] M. **Titov**, *Current Trends in the Micro-Pattern Gas Detectors – talk at the workshop Micro Pattern Gas Detectors. Towards an R&D Collaboration., CERN, 2007-09-10*, **2007**

[543] S. **Andriamonje**, **et al.**, *Development and performance of Microbulk Micromegas detectors*, J. Instrum., vol. 5, no. 02, p. P02001, **2010**

[544] H. **Lin**, **et al.**, *Design and commissioning of a 600 L Time Projection Chamber with Microbulk Micromegas*, J. Instrum., vol. 13, no. 06, p. P06012, arXiv:1804.02863, **2018**

[545] T. **Gys**, *Status and perspectives of vacuum-based photon detectors for single photon detection*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 595, no. 1, pp. 136–141, **2008**

[546] S. H. **Byun**, *Radioisotopes and Radiation Methodology I,II – Lecture 6: Pulse Processing, McMaster University*, **2017**

[547] O. **Siegmund**, **et al.**, *Development of Sealed Tube Microchannel Plate Detectors with Cross Strip Readouts*, in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, pp. 1–7, IEEE, ISBN 978-1-5386-8494-8, **2018**

[548] X. **Wei**, **et al.**, *Development of Parallel Plate Avalanche Counter for heavy ion collision in radioactive ion beam*, Nucl. Eng. Technol., vol. 52, no. 3, pp. 575–580, **2020**

[549] X. **Michalet**, **et al.**, *Photon-counting H33D detector for biological fluorescence imaging*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 133–136, **2006**

[550] A. **Lyashenko**, **et al.**, *Performance of Large Area Picosecond Photo-Detectors (LAPPD™)*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 958, p. 162834, **2020**

[551] O. H. W. **Siegmund**, **et al.**, *Microchannel plate imaging photon counters for ultraviolet through NIR detection with high time resolution*, in M. A. **Itzler** and J. C. **Campbell** (editors), *Proc. SPIE 8033, Advanced Photon Counting Techniques V*, p. 80330V, **2011**

[552] S. **Gundacker**, **et al.**, *On the comparison of analog and digital SiPM readout in terms of expected timing performance*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 787, pp. 6–11, **2015**

[553] B. **Lubsandorzhiev**, *Hybrid phototubes in neutrino telescopes: Experience and perspectives*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 602, no. 1, pp. 201–204, **2009**

[554] F. **Barbato**, **et al.**, *Recent development on the realization of a 1-inch VSiPMT prototype*, EPJ Web Conf., vol. 136, p. 02016, **2017**

[555] G. **Barbarino**, **et al.**, *Another step towards photodetector innovation: The first 1-inch industrial VSiPMT*, Astropart. Phys., vol. 101, pp. 27–35, **2018**

[556] Y. **Wang**, **et al.**, *Feasibility study of SiGHT: a novel ultra low background photosensor for low temperature operation*, J. Instrum., vol. 12, no. 02, p. P02019, arXiv:1611.04713, **2017**

[557] A. **Braem**, **et al.**, *Design, fabrication and performance of the 10-in. TOM HPD*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 518, no. 1-2, pp. 574–578, **2004**

[558] C. **Joram**, **et al.**, *Proof of principle of G-APD based hybrid photodetectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 621, no. 1-3, pp. 171–176, **2010**

[559] M. **Taiuti**, *A project for a high-efficiency direction-sensitive photo-detector to be used in underwater neutrino telescopes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 525, no. 1-2, pp. 137–140, **2004**

[560] R. **Winston**, **et al.**, *Nonimaging Optics*, Academic Press, ISBN 9780127597515, **2005**

[561] D. **Ferenc**, *Single-photon color-sensitive photosensor TransReFerence*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 497, no. 1, pp. 198–201, **2003**

[562] O. H. W. **Siegmund**, **et al.**, *Large Area and High Efficiency Photon Counting Imaging Detectors with High Time and Spatial Resolution for Night Time Sensing and Astronomy*, in S. **Ryan** (editor), *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference 2012 (AMOS 2012)*, p. 92, The Maui Economic Development Board, Inc., **2012**

[563] F. **Omnes**, *Introduction to Semiconductor Photodetectors*, in *Optoelectronic Sensors*, pp. 1–14, ISTE, London, UK, ISBN 9781848210783, **2010**

[564] C. H. **Lin** and C. W. **Liu**, *Metal-insulator-semiconductor photodetectors*, Sens., vol. 10, no. 10, pp. 8797–8826, **2010**

[565] R. **Paschotta**, *Position-sensitive Detectors [version 2020-03-14]*, RP Photonics Encyclopedia, **2020**

[566] L. **Hirvonen** and K. **Suhling**, *Photon Counting Imaging with an Electron-Bombarded Pixel Image Sensor*, Sens., vol. 16, no. 5, p. 617, **2016**

[567] M. **Suyama**, **et al.**, *Single-photon-sensitive EBCCD with additional multiplication*, in R. E. **Longshore** and A. **Sood** (editors), *Proc. SPIE 6294, Infrared and Photoelectronic Imagers and Detector Devices II*, p. 629407, **2006**

[568] V. **Prodanović**, **et al.**, *Ultra-thin alumina and silicon nitride MEMS fabricated membranes for the electron multiplication*, Nanotechnol., vol. 29, no. 15, p. 155703, **2018**

[569] M. **McClish**, **et al.**, *Recent advances of planar silicon APD technology*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 36–40, **2006**

[570] D. **Renker** and E. **Lorenz**, *Advances in solid state photon detectors*, J. Instrum., vol. 4, no. 04, p. P04004, **2009**

[571] A. **Bross**, **et al.**, *Characterization and performance of visible light photon counters (VLPCs) for the upgraded DØ detector at the Fermilab Tevatron*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 477, no. 1-3, pp. 172–178, **2002**

[572] M. **Wayne**, *Visible light photon counters and the DØ scintillating fiber tracker*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 387, no. 1-2, pp. 278–281, **1997**

[573] E. **Shefer**, *Study of novel stable photocathode materials for gaseous photon detectors in the near-UV to visible spectral range*, Ph.D. thesis, Weizmann Institute of Science, **2000**

[574] R. H. **Haitz**, **et al.**, *Avalanche Effects in Silicon p–n Junctions. I. Localized Photomultiplication Studies on Microplasmas*, J. Appl. Phys., vol. 34, no. 6, pp. 1581–1590, **1963**

[575] M. D. **Eisaman**, **et al.**, *Invited Review Article: Single-photon sources and detectors*, Rev. Sci. Instrum., vol. 82, no. 7, p. 071101, **2011**

[576] Y. **Haemisch**, **et al.**, *Fully Digital Arrays of Silicon Photomultipliers (dSiPM) – a Scalable Alternative to Vacuum Photomultiplier Tubes (PMT)*, Phys. Procedia, vol. 37, pp. 1546–1560, **2012**

[577] R. **Paschotta**, *Image Sensors [version 2020-03-14]*, RP Photonics Encyclopedia, **2020**

[578] C. **Coates**, **et al.**, *sCMOS – Scientific CMOS Technology – A High-Performance Imaging Breakthrough, White Paper*, Tech. Rep., Andor Technology, Fairchild Imaging, PCO imaging, **2009**

[579] G. **Schmaler**, *Characterisation of PNCCDs and Analysis of Pixel Defects*, Ph.D. thesis, Technische Universität München, **2012**

[580] J. **Tiffenberg**, **et al.**, *Single-Electron and Single-Photon Sensitivity with a Silicon Skipper CCD*, Phys. Rev. Lett., vol. 119, no. 13, p. 131802, arXiv:1706.00028, **2017**

[581] E. **Fossum**, **et al.**, *The Quanta Image Sensor: Every Photon Counts*, Sens., vol. 16, no. 8, p. 1260, **2016**

[582] S. M. **Sze** and M.-K. **Lee**, *Semiconductor Devices: Physics and Technology*, John Wiley & Sons, 3rd ed., ISBN 978-0-470-91407-6, **2012**

[583] R. **Paschotta**, *Phototubes [version 2020-03-14]*, RP Photonics Encyclopedia, **2020**

[584] **The Editors of Encyclopaedia Britannica**, *Photoelectric cell [version 2014-05-13]*, Encyclopædia Britannica, **2014**

[585] D. A. **DeWolf**, **et al.**, *RCA Electro-Optics Handbook*, Atlantic Books, ISBN 978-0913970119, **1974**

[586] **Hamamatsu Photonics K.K. Electron Tube Division**, *Phototubes [catalog]*, **2016**

[587] R. **Brown**, **et al.**, *The Vacuum Phototriodes for the CMS Electromagnetic Calorimeter*, in *Proceedings of International Europhysics Conference on High Energy Physics — PoS(hep2001)*, vol. 7, p. 255, Sissa Medialab, Trieste, Italy, **2001**

[588] S. **Zahid**, **et al.**, *A segmented anode vacuum phototriode with position sensitivity*, J. Instrum., vol. 13, no. 01, p. C01014, **2018**

[589] M. C. E. **Huber**, **et al.** (editors), *Observing Photons in Space – A Guide to Experimental Space Astronomy*, vol. 9 of *ISSI Scientific Report Series*, Springer, New York, 2nd ed., ISBN 978-1-4614-7803-4, **2013**

[590] J. **Xie**, **et al.**, *MCP-PMT development at Argonne for particle identification*, J. Instrum., vol. 15, no. 04, p. C04038, **2020**

[591] F. **Gao**, **et al.**, *Status of the 20 inch MCP-PMT prototype development for JUNO experiment*, J. Physics: Conf. Ser., vol. 888, no. 1, p. 012050, **2017**

[592] Y. **Chang**, **et al.**, *The R&D of the 20 in. MCP–PMTs for JUNO*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 824, pp. 143–144, **2016**

[593] C. **Bronner**, **et al.**, *Performances of the Hyper-Kamiokande 20" PMT*, in *Proceedings of the 5th International Workshop on New Photon-Detectors (PD18)*, vol. 27, p. 011013, Journal of the Physical Society of Japan, ISBN 4-89027-137-6, **2019**

[594] M. J. **Minot**, **et al.**, *Large area picosecond photodetector (LAPPD™) offers fast timing for nuclear physics and medical imaging*, Nuovo Cim. C, vol. 43, no. 1, p. 11, **2020**

[595] M. **Wetstein**, *LAPPD Commercial Availability, Status, and Opportunities – talk at the Theia workshop 2018-04-12–14, University of California, Davis, 2018-04-13*, **2018**

[596] B. **Adams**, **et al.**, *Timing characteristics of Large Area Picosecond Photodetectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 795, pp. 1–11, **2015**

[597] B. W. **Adams**, **et al.**, *A Brief Technical History of the Large-Area Picosecond Photodetector (LAPPD) Collaboration*, arXiv e-prints, arXiv:1603.01843, **2016**

[598] J. **Xie**, *Fast-timing microchannel plate photodetectors: design, fabrication and characterization – talk at the 5th International Workshop on New Photon-Detectors (PD18), 2018-11-29*, **2018**

[599] A. **Elagin**, *New Opportunities with Large-Area Picosecond Photo-Detectors – talk at the HEE Seminar, Boston University, 2019-04-04*, **2019**

[600] O. H. W. **Siegmund**, **et al.**, *Large area event counting detectors with high spatial and temporal resolution*, J. Instrum., vol. 9, no. 04, p. C04002, **2014**

[601] G. R. **Jocher**, **et al.**, *Multiple-photon disambiguation on stripline-anode Micro-Channel Plates*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 822, pp. 25–33, arXiv:1805.01077, **2016**

[602] L. **Ren**, **et al.**, *Mass Production of MCP-PMT for JUNO and Development of 20-inch MCP-PMT with TTS Improved*, in *Proceedings of the 5th International Workshop on New Photon-Detectors (PD18)*, vol. 27, p. 011014, Journal of the Physical Society of Japan, ISBN 4-89027-137-6, **2019**

[603] Q. **Wu**, **et al.**, *The Status of the 20 inch MCP-PMT and its APR Test Result*, in *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–4, IEEE, ISBN 978-1-7281-4164-0, **2019**

[604] Y. **Zhu**, **et al.**, *The mass production and batch test result of 20" MCP-PMTs*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 952, p. 162002, **2020**

[605] P. **Chen**, **et al.**, *A novel bowl-shaped microchannel plate with high electron collection efficiency and good time performance*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 936, pp. 534–536, **2019**

[606] Y. **Nishimura**, *New large aperture photodetectors for water Cherenkov detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 958, p. 162993, **2020**

[607] Y. **Kudenko**, *Hyper-Kamiokande*, J. Instrum., vol. 15, no. 07, p. C07029, arXiv:2005.13641, **2020**

[608] B. **Moomaw**, *Chapter 13: Camera Technologies for Low Light Imaging: Overview and Relative Advantages*, in G. **Sluder** and D. E. **Wolf** (editors), *Digital Microscopy*, vol. 81 of *Methods in Cell Biology*, pp. 251–283, Academic Press, 3rd ed., **2007**

[609] A. **Tremsin** and J. **Vallerga**, *Unique capabilities and applications of Microchannel Plate (MCP) detectors with Medipix/Timepix readout*, Radiat. Meas., vol. 130, p. 106228, **2020**

[610] M. **Fiorini**, **et al.**, *Single-photon imaging detector with $\mathcal{O}(10)$ ps timing and sub-10 μm position resolutions*, J. Instrum., vol. 13, no. 12, p. C12005, **2018**

[611] R. **Bellazzini**, **et al.**, *Single photon imaging at ultra-high resolution*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 591, no. 1, pp. 125–128, **2008**

[612] T. **Gys** and C. **Joram**, *Position-sensitive vacuum photon detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 970, p. 163373, **2020**

[613] B. **Moomaw**, *Chapter 11 - Camera Technologies for Low Light Imaging: Overview and Relative Advantages*, in G. **Sluder** and D. E. **Wolf** (editors), *Digital Microscopy*, vol. 114 of *Methods in Cell Biology*, pp. 243–283, Academic Press, 4th ed., **2013**

[614] B. W. **Miller**, **et al.**, *The iQID camera: An ionizing-radiation quantum imaging detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 767, pp. 146–152, **2014**

[615] T. **Ohnuki**, **et al.**, *Development of an ultrafast single photon counting imager for single molecule imaging*, in J. **Enderlein** and Z. K. **Gryczynski** (editors), *Proc. SPIE 6092, Ultrasensitive and Single-Molecule Detection Technologies*, p. 60920P, **2006**

[616] T. **van der Reep**, **et al.**, *Measurement of the transmission secondary electron yield of nanometer-thick films in a prototype Timed Photon Counter*, J. Instrum., vol. 15, no. 10, p. P10022, arXiv:2010.12462, **2020**

[617] J. **Pouthas**, *Large Photodetector Developments in Europe – talk at the Next Generation of Nucleon Decay and Neutrino Detectors conference (NNN05), 2005-04-08*, **2005**

[618] A. **Braem**, **et al.**, *The X-HPD—A modern implementation of a SMART concept*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 602, no. 1, pp. 193–196, **2009**

[619] M. **Mahgoub**, **et al.**, *Longitudinal phase space studies at the PITZ facility*, in *Proceedings of International Particle Accelerator Conference 2012*, pp. 631–633, ISBN 9783954501151, **2012**

[620] J. **Song**, **et al.**, *Development of a new type of hybrid photo-detector involving photocathode, scintillator, and silicon photomultiplier*, J. Instrum., vol. 15, no. 07, p. C07041, **2020**

[621] C. H. V. **Wiebusch**, *The Detection of Faint Light in Deep Underwater Neutrino Telescopes*, Ph.D. thesis, RWTH Aachen, **1995**

[622] G. **van Aller**, **et al.**, *A "SMART" 35 cm Diameter Photomultiplier*, Helv. Phys. Acta, vol. 59, no. 6-7, p. 1119, **1986**

[623] A. **Roberts**, *The birth of high-energy neutrino astronomy: A personal history of the DUMAND project*, Rev. Mod. Phys., vol. 64, no. 1, pp. 259–312, **1992**

[624] B. **Lubsandorzhiev**, *QUASAR-370 hybrid phototube as a prototype of a photodetector for the next generation of deep underwater neutrino telescopes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 610, no. 1, pp. 68–71, **2009**

[625] B. K. **Lubsandorzhiev**, *New Developments of Photodetectors for the Lake Baikal Neutrino Experiment*, in M. **Barone**, **et al.** (editors), *Advanced Technology and Particle Physics*, pp. 79–84, World Scientific, ISBN 978-981-238-180-4, **2002**

[626] B. **Lubsandorzhiev**, *Photodetectors of Lake Baikal Neutrino experiment and Tunka Air Cherenkov Array*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 442, no. 1-3, pp. 368–373, **2000**

[627] B. K. **Lubsandorzhiev** and B. **Combettes**, *The Quest for the Ideal Scintillator for Hybrid Phototubes*, IEEE Trans. Nucl. Sci., vol. 55, no. 3, pp. 1333–1337, **2008**

[628] B. **Lubsandorzhiev**, **et al.**, *Development of a hybrid phototube with ZnO:Ga luminescent screen and GaN photocathode*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 695, pp. 118–120, **2012**

[629] O. **Kalekin**, *Optical modules for the neutrino telescope KM3NeT*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 623, no. 1, pp. 312–315, **2010**

[630] U. **Katz**, *Status of the KM3NeT project*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 602, no. 1, pp. 40–46, **2009**

[631] R. **Paschotta**, *Streak Cameras [version 2020-03-14]*, RP Photonics Encyclopedia, **2020**

[632] **Hamamatsu Photonics K.K. Systems Division**, *Guide to streak cameras*, **2008**

[633] D. **Ferenc**, *Vacuum photosensor device with electron lensing – US Patent 9064678 B2*, **2015**

[634] F. C. T. **Barbato** and G. **Barbarino**, *Understanding VSiPMT: a comparison with other large area hybrid photodetectors*, arXiv e-prints, arXiv:2004.12627, **2020**

[635] S. **Jeon** and S. **Kim**, *Physics Sensitivity Studies at Korean Neutrino Observatory*, in *Proceedings of The 39th International Conference on High Energy Physics — PoS(ICHEP2018)*, p. 794, Sissa Medialab, Trieste, Italy, **2019**

[636] I. **Rodionov**, **et al.**, *Hybrid gaseous photomultipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 478, no. 1-2, pp. 384–390, **2002**

[637] T. **Francke**, **et al.**, *Novel position-sensitive gaseous detectors with solid photocathodes*, IEEE Trans. Nucl. Sci., vol. 49, no. 3, pp. 977–983, arXiv:physics/0206071v1, **2002**

[638] M. **Moritz**, **et al.**, *Performance study of new pixel hybrid photon detector prototypes for the LHCb RICH counters*, IEEE Trans. Nucl. Sci., vol. 51, no. 3, pp. 1060–1066, **2004**

[639] T. **Gys**, *Production of 500 pixel hybrid photon detectors for the RICH counters of LHCb*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 176–179, **2006**

[640] S. **Bradbury**, **et al.**, *Test of the new hybrid INTEVAC intensified photocell for the use in air Cherenkov telescopes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 387, no. 1-2, pp. 45–49, **1997**

[641] D. **Ferenc**, *A novel photosensor concept*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 471, no. 1-2, pp. 229–233, **2001**

[642] A. **Braem**, **et al.**, *Highly segmented large-area hybrid photodiodes with bialkali photocathodes and enclosed VLSI readout electronics*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 442, no. 1-3, pp. 128–135, **2000**

[643] N. **Kanaya**, **et al.**, *Performance study of hybrid photon detectors for the LHCb RICH*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 553, no. 1-2, pp. 41–45, **2005**

[644] M. **Adinolfi**, **et al.**, *Performance of the LHCb RICH photo-detectors and readout in a system test using charged particles from a 25 ns-structured beam*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 603, no. 3, pp. 287–293, **2009**

[645] A. **Braem**, **et al.**, *Development of HPDs for applications in physics and medical imaging*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 162–165, **2006**

[646] E. **Albrecht**, **et al.**, *Performance of hybrid photon detector prototypes with 80% active area for the rich counters of LHCB*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 442, no. 1-3, pp. 164–170, **2000**

[647] C. **Joram**, *Photodetection in the LHC experiments*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 695, pp. 13–22, **2012**

[648] P. B. **Cushman** and A. H. **Heering**, *CMS HCAL Hybrid Photodiode Design and Quality Assurance Stations*, ICFA Instrum. Bull., vol. 25, pp. 1–16, **2002**

[649] P. **Cushman**, **et al.**, *Custom HPD readout for the CMS HCAL*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 442, no. 1-3, pp. 289–294, **2000**

[650] P. **Cushman**, **et al.**, *Crosstalk properties of the CMS HCAL hybrid photodiode*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 504, no. 1-3, pp. 62–69, **2003**

[651] J. **Freeman**, *Innovations for the CMS HCAL*, Int. J. Mod. Phys. A, vol. 25, no. 12, pp. 2421–2436, **2010**

[652] F. **De Guio**, *First results from the CMS SiPM-based hadronic endcap calorimeter*, J. Physics: Conf. Ser., vol. 1162, no. 1, p. 012009, **2019**

[653] A. **Piccioli**, **et al.**, *HPD imaging properties in Cherenkov telescopes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 525, no. 1-2, pp. 144–147, **2004**

[654] A. **Braem**, **et al.**, *Development of a 10-inch HPD with integrated readout electronics*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 504, no. 1-3, pp. 19–23, **2003**

[655] C. **Joram**, **et al.**, *Development of HPDs for applications in physics and medical imaging – talk at the 4th International Conference on New developments in Photodetection (NDIP05), 2005-06-23*, **2005**

[656] A. **Ball**, **et al.**, *A large spherical HPD for a novel deep-sea neutrino experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 553, no. 1-2, pp. 85–90, **2005**

[657] T. K. **Rügheimer**, **et al.**, *Experimental demonstration of a hybrid photon detector concept based on the Timepix detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 595, no. 2, pp. 353–358, **2008**

[658] A. **Mac Raighne**, **et al.**, *Development of a high-speed single-photon pixellated detector for visible wavelengths*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 607, no. 1, pp. 166–168, **2009**

[659] A. **Mac Raighne**, **et al.**, *Imaging visible light with Medipix2*, Rev. Sci. Instrum., vol. 81, no. 11, p. 113103, **2010**

[660] T. **Abe**, **et al.**, *Development of large-aperture hybrid avalanche photo-detector*, in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pp. 2187–2189, IEEE, ISBN 978-1-4244-3961-4, **2009**

[661] A. **Teymourian**, **et al.**, *Characterization of the QUartz Photon Intensifying Detector (QUPID) for noble liquid detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 654, no. 1, pp. 184–195, arXiv:1103.3689, **2011**

[662] Y. **Yusa**, **et al.**, *Test of the HAPD light sensor for the Belle II Aerogel RICH*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 876, pp. 149–152, **2017**

[663] M. **McClish**, **et al.**, *A hybrid position sensitive avalanche photodiode detector for scintillation spectroscopy and imaging*, in *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No. 03CH37515)*, vol. 2, pp. 1358–1362, IEEE, ISBN 0-7803-8257-9, **2003**

[664] D. **Ferenc**, **et al.**, *First ReFerence photosensor prototype*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 504, no. 1-3, pp. 359–363, **2003**

[665] M. **Suyama**, **et al.**, *A compact hybrid photodetector (HPD)*, IEEE Trans. Nucl. Sci., vol. 44, no. 3, pp. 985–989, **1997**

[666] F. **Retière**, *Photo-detector R&D for next generation NNN detectors – talk at the 14th International Workshop on Next generation Nucleon Decay and Neutrino Detectors (NNN13), 2013-11-12*, **2013**

[667] D. **Ferenc**, **et al.**, *A new method for vacuum sealing of flat-panel photosensors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 205–208, **2006**

[668] A. **Lyashenko**, *QUartz Photon Intensifying Detector (QUPID): a possible alternative for noble liquid experiments – talk at the Light Detection In Noble Elements conference (LIDINE 2013), Fermilab, 2013-05-29*, **2013**

[669] K. **Arisaka**, **et al.**, *Studies of a three-stage dark matter and neutrino observatory based on multi-ton combinations of liquid xenon and liquid argon detectors*, Astropart. Phys., vol. 36, no. 1, pp. 93–122, arXiv:1107.1295, **2012**

[670] Y. **Wang**, **et al.**, *The development of SiGHT: an ultra low background photosensor*, in R. B. **James**, **et al.** (editors), *Proc. of SPIE Vol. 9968, Hard X-Ray, Gamma-Ray, and Neutron Detector Physics XVIII*, p. 99680U, **2016**

[671] G. **Barbarino**, **et al.**, *A new generation photodetector for astroparticle physics: The VSiPMT*, Astropart. Phys., vol. 67, pp. 18–25, arXiv:1407.2805v2, **2015**

[672] S. **Korpar**, **et al.**, *A 144-channel HAPD for the Aerogel RICH at Belle II*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 766, pp. 145–147, **2014**

[673] **Hamamatsu Photonics K.K. Electron Tube Division**, *Hamamatsu High Speed Compact HPD (Hybrid Photo Detector) Series R10467U / R11322U-40 / H13223-40 [datasheet]*, **2018**

[674] T. **Lou**, *Development and Performance Evaluation of Large-Aperture Hybrid Photo-Detector for Hyper-Kamiokande*, J. Physics: Conf. Ser., vol. 888, p. 012160, **2017**

[675] J. **Wang**, *Research and Development of a Digital Hybrid Avalanche Photo-Detector*, Master thesis, University of Tokyo, **2012**

[676] T. **Abe**, *R&D status of HAPD*, in *Proceedings of International Workshop on New Photon Detectors — PoS(PD09)*, vol. 090, p. 014, Sissa Medialab, Trieste, Italy, **2010**

[677] H. **Nakayama**, **et al.**, *Development of a 13-in. Hybrid Avalanche Photo-Detector (HAPD) for a next generation water Cherenkov detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 567, no. 1, pp. 172–175, **2006**

[678] Y. **Kawai**, **et al.**, *Large-aperture hybrid photo-detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 579, no. 1, pp. 42–45, **2007**

[679] T. **Abe**, **et al.**, *R&D status of a large-aperture hybrid avalanche photo-detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 623, no. 1, pp. 279–281, **2010**

[680] A. **Taketa** and K. **Hoshina**, *Calibration and Development of Preamplifier for 8-inch Hybrid Photo Detector – talk at the 5th International Workshop on New Photon-Detectors (PD18), 2018-11-29*, **2018**

[681] F. D. **Lodovico**, *Hyper-Kamiokande R&D - NNN13 – talk at the 14th International Workshop on Next generation Nucleon Decay and Neutrino Detectors (NNN13), 2013-11-12*, **2013**

[682] H. **Aihara**, *R&D of a Large Format Hybrid Photo-Detector for a Next Generation Water Cherenkov Detector – talk at the Next Generation of Nucleon Decay and Neutrino Detectors conference (NNN05), 2005-04-08*, **2005**

[683] A. **Kusaka**, *Research and Development of a Hybrid Photo Sensor for a Water Cerenkov Detector*, Master thesis, University of Tokyo, **2004**

[684] E. **de Wolf** (editor), *Proceedings of the Workshop on Technical Aspects of a Very Large Volume Neutrino telescope in the Mediterranean Sea (VLVnT), Amsterdam, Netherlands, 5–8 Oct. 2003*, NIKHEF, ISBN 9789064880261, **2004**

[685] M. **Suyama**, **et al.**, *Development of a multi-pixel photon sensor with single-photon sensitivity*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 523, no. 1-2, pp. 147–157, **2004**

[686] G. **Barbarino**, **et al.**, *A new high-gain vacuum photomultiplier based upon the amplification of a Geiger-mode p–n junction*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 594, no. 3, pp. 326–331, **2008**

[687] F. **Barbato**, **et al.**, *The 2-inches VSiPMT industrial prototypes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 958, p. 162144, **2020**

[688] **Hamamatsu Photonics K.K. Electron Tube Division**, *Technical information – How to Use UVTRON*, **2009**

[689] **Hamamatsu Photonics K.K. Electron Tube Division**, *Hamamatsu Electron Tube Products – Condensed Catalog*, **2020**

[690] C. **Traunsteiner**, *Investigation of gas photosensors for LENA and development of a hybrid-gas photomultiplier – Diploma thesis defense, Technische Universität München, (n.d.)*, **2010**

[691] F. **Tessarotto**, *Evolution and recent developments of the gaseous photon detectors technologies*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 912, pp. 278–286, arXiv:1710.09309, **2018**

[692] A. **Lyashenko**, **et al.**, *High-gain continuous-mode operated gaseous photomultipliers for the visible spectral range*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 610, no. 1, pp. 161–163, **2009**

[693] S. D. **Torre**, *Single photon detection by gaseous counters: Status and perspectives*, in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–4, IEEE, ISBN 9781509016433, **2016**

[694] F. **Tokanai**, **et al.**, *Newly developed gaseous photomultiplier*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 766, pp. 176–179, **2014**

[695] V. **Peskov**, **et al.**, *Liquid and solid organic photocathodes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 269, no. 1, pp. 149–160, **1988**

[696] Y. **Giomataris**, **et al.**, *A ring-imaging detector with liquid and solid radiators using a multistep parallel-plate avalanche chamber at atmospheric pressure with optical readout*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 279, no. 1-2, pp. 322–330, **1989**

[697] V. **Peskov**, **et al.**, *First attempts to combine capillary tubes with photocathodes*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 433, no. 1-2, pp. 492–501, **1999**

[698] M. **Balcerzyk**, **et al.**, *Methods of preparation and results of sealed gas photomultipliers for visible light*, in *2002 IEEE Nuclear Science Symposium Conference Record*, vol. 1, pp. 302–307, IEEE, ISBN 0-7803-7636-6, **2002**

[699] A. V. **Lyashenko**, **et al.**, *Further progress in ion back-flow reduction with patterned gaseous hole-multipliers*, J. Instrum., vol. 2, no. 08, p. P08004, **2007**

[700] A. V. **Lyashenko**, **et al.**, *Development of high-gain gaseous photomultipliers for the visible spectral range*, J. Instrum., vol. 4, no. 07, p. P07005, **2009**

[701] F. **Tokanai**, **et al.**, *Development of gaseous PMT with micropattern gas detector*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 610, no. 1, pp. 164–168, **2009**

[702] T. **Sumiyoshi**, **et al.**, *Development of a gaseous PMT with micro-pattern gas detectors*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 639, no. 1, pp. 121–125, **2011**

[703] T. **Moriya**, **et al.**, *A concise quantum efficiency measurement system for gaseous photomultipliers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 732, pp. 269–272, **2013**

[704] J. H. **Carver** and P. **Mitchell**, *Ionization chambers for the vacuum ultra-violet*, J. Sci. Instrum., vol. 41, no. 9, pp. 555–557, **1964**

[705] A. **Breskin**, **et al.**, *A highly efficient low-pressure UV-RICH detector with optical avalanche recording*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 273, no. 2-3, pp. 798–804, **1988**

[706] A. **Breskin**, **et al.**, *A 3-stage gated UV-photon gaseous detector with optical imaging*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 286, no. 1-2, pp. 251–261, **1990**

[707] J. **Baechler**, **et al.**, *A large area optical imaging UV detector operating with TMAE at 48°C*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 324, no. 3, pp. 449–460, **1993**

[708] J. **Baechler**, **et al.**, *An optical readout RICH detector for the NA35 heavy-ion experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 343, no. 1, pp. 213–217, **1994**

[709] J. **Va'vra**, *Photon detectors with gaseous amplification*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 387, no. 1-2, pp. 137–145, **1997**

[710] R. **Baur**, **et al.**, *In-beam experience from the CERES UV-detectors: prohibitive spark breakdown in multi-step parallel-plate chambers as compared to wire chambers*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 343, no. 1, pp. 231–240, **1994**

[711] A. **Rubin**, **et al.**, *Optical readout: a tool for studying gas-avalanche processes*, J. Instrum., vol. 8, no. 08, p. P08001, arXiv:1305.1196, **2013**

[712] G. **Cavoto**, **et al.**, *Micro pattern gas detector optical readout for directional dark matter searches*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 958, p. 162400, **2020**

[713] J. **Benziger**, **et al.**, *The nylon scintillator containment vessels for the Borexino solar neutrino experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 582, no. 2, pp. 509–534, arXiv:physics/0702162, **2007**

[714] A. **Pocar**, **et al.**, *Solar neutrino physics with Borexino*, in *SciPost Physics Proceedings*, vol. 1, p. 25, SciPost, **2019**

[715] H. **Hanada**, **et al.**, *A highly sensitive optical detector for use in deep underwater*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 408, no. 2-3, pp. 425–437, **1998**

[716] E. **Anassontzis**, **et al.**, *The optical module for the NESTOR neutrino telescope*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 479, no. 2-3, pp. 439–455, **2002**

[717] A. **Achterberg**, **et al.**, *First year performance of the IceCube neutrino telescope*, Astropart. Phys., vol. 26, no. 3, pp. 155–173, arXiv:astro-ph/0604450, **2006**

[718] I. **Amore**, **et al.**, *NEMO: A Project for a km^3 Underwater Detector for Astrophysical Neutrinos in the Mediterranean Sea*, Int. J. Mod. Phys. A, vol. 22, no. 21, pp. 3509–3520, arXiv:0709.3991v1, **2007**

[719] M. **Ageron**, **et al.**, *ANTARES: The first undersea neutrino telescope*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 656, no. 1, pp. 11–38, **2011**

[720] M. A. **Knapp**, *Design, Simulation und Aufbau des GERDA-Myonvetos [Design, simulation and construction of the GERDA muon veto, in German]*, Ph.D. thesis, Eberhard-Karls-Universität Tübingen, **2009**

[721] L. **Agostino**, *MEMPHYS R&D – talk at the 14th International Workshop on Next generation Nucleon Decay and Neutrino Detectors (NNN13), 2013-11-12*, **2013**

[722] G. **Alimonti**, **et al.**, *A large-scale low-background liquid scintillation detector: the counting test facility at Gran Sasso*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 406, no. 3, pp. 411–426, **1998**

[723] K. **Chow**, **et al.**, *Waterproofed photomultiplier tube assemblies for the Daya Bay reactor neutrino experiment*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 794, pp. 25–32, **2015**

[724] P. **Alivisatos**, **et al.**, *Proposal for US Participation in KamLAND*, U.S. KamLAND Collaboration, **1999**

[725] G. **Beischler**, *Engineering of an Optical Module for the LENA Detector*, Master thesis, Technische Universität München, **2013**

[726] G. **Beischler**, *Design Study and FEM Simulations of Pressure Resistant Photomultiplier Encapsulations for the LENA Detector*, Bachelor thesis, Technische Universität München, **2011**

[727] C. **Lendvai**, *Photomultiplier-Einkapselung für das Myon-Veto des Borexino-Experiments und Messung myoninduzierter Neutronen in der CTFII [Photomultiplier Housing for the Muon*

Veto of the Borexino Experiment and Measurement of Muon-Induced Neutrons in the CTFII, in German], Diploma thesis, Technische Universität München, **2001**

[728] A. **de Bellefon**, **et al.**, *MEMPHYS:A large scale water Cerenkov detector at Fréjus*, arXiv e-prints, arXiv:hep-ex/0607026, **2006**

[729] S. K. **Agarwalla**, **et al.**, *LAGUNA-LBNO Design Study – Design of a pan-European infrastructure for Large Apparatus for Grand Unification, Neutrino Astrophysics and Long Baseline Neutrino Oscillations – Part 1: Water Cerenkov Detector - Underground Layout, Tank, Instrumentation, Liquid Handling, Safety*, Tech. Rep., The LAGUNA-LBNO Consortium, **2014**

[730] L. **Oberauer**, *Personal communications*, **2014**

[731] S. **Ahlen**, **et al.**, *First supermodule of the MACRO detector at Gran Sasso*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 324, no. 1-2, pp. 337–362, **1993**

[732] M. E. **Moorhead**, *Reflectors in Cherenkov Detectors*, Ph.D. thesis, Oxford University, **1992**

[733] G. **Doucas**, **et al.**, *Light concentrators for the Sudbury Neutrino Observatory*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 370, no. 2-3, pp. 579–596, **1996**

[734] L. **Oberauer**, **et al.**, *Light concentrators for Borexino and CTF*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 530, no. 3, pp. 453–462, arXiv:physics/0310076, **2004**

[735] B. **Huber**, **et al.**, *Solid light concentrators for small-sized photosensors used in Cherenkov telescopes*, in *Proceedings of the 32nd International Cosmic Ray Conference, Beijing 2011 (ICRC2011)*, vol. 9, p. 2, **2011**

[736] H. **Anderhub**, **et al.**, *Design and operation of FACT – the first G-APD Cherenkov telescope*, J. Instrum., vol. 8, no. 06, p. P06008, **2013**

[737] F. **Hénault**, **et al.**, *Design of light concentrators for Cherenkov telescope observatories*, in R. **Winston** and J. **Gordon** (editors), *Nonimaging Optics: Efficient Design for Illumination and Solar Concentration X*, vol. 8834, pp. 13–24, SPIE, arXiv:1309.4252, **2013**

[738] J. **Aguilar**, **et al.**, *Design, optimization and characterization of the light concentrators of the single-mirror small size telescopes of the Cherenkov Telescope Array*, Astropart. Phys., vol. 60, pp. 32–40, arXiv:1404.2734, **2015**

[739] M. **Actis**, **et al.**, *Design concepts for the Cherenkov Telescope Array CTA: an advanced facility for ground-based high-energy gamma-ray astronomy*, Exp. Astron., vol. 32, no. 3, pp. 193–316, arXiv:1008.3703, **2011**

[740] O. **Kavatsyuk**, **et al.**, *Expansion cone for the 3-inch PMTs of the KM3NeT optical modules*, J. Instrum., vol. 8, no. 03, p. T03006, **2013**

[741] S. **Aiello**, **et al.**, *KM3NeT Technical Design Report for a Deep-Sea Research Infrastructure in the Mediterranean Sea Incorporating a Very Large Volume Neutrino Telescope*, The KM3NeT collaboration, ISBN 978-90-6488-033-9, **2009**

[742] M. B. **Avanzini**, **et al.**, *Occulting Light Concentrators in Liquid Scintillator Neutrino Detectors*, J. Physics: Conf. Ser., vol. 888, no. 1, p. 012055, arXiv:1612.05444, **2017**

[743] M. **Shiozawa**, *Hyper-Kamiokande project – talk at the Seminar of the High Energy Physics Department of INR RAS, Moscow, 2019-04-25*, **2019**

[744] **The MathWorks Inc., Natick (MA), USA**, *MATLAB and MuPAD, release 2016b [software]*

[745] M. **Hohenwarter**, **et al.**, *GeoGebra 5.0 [software]*, **2017**

[746] S. **Agostinelli**, **et al.**, *Geant4—a simulation toolkit*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 506, no. 3, pp. 250–303, **2003**

[747] J. **Allison**, **et al.**, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci., vol. 53, no. 1, pp. 270–278, **2006**

[748] T. **Marrodán Undagoitia**, *Search for the Proton Decay in the large liquid Scintillator Detector LENA - A Feasibility Study*, Diploma thesis, Technische Universität München, **2005**

[749] T. **Marrodán Undagoitia**, *Measurement of light emission in organic liquid scintillators and studies towards the search for proton decay in the future large-scale detector LENA*, Ph.D. thesis, Technische Universität München, **2008**

[750] J. M. A. **Winter**, *Phenomenology of Supernova Neutrinos, Spatial Event Reconstuction, and Scintillation Light Yield Measurements for the Liquid-Scintillator Detector LENA*, Diploma thesis, Technische Universität München, **2007**

[751] J. M. A. **Winter**, *Detection of Supernova Neutrinos in the Liquid-Scintillator Experiment LENA*, Ph.D. thesis, Technische Universität München, **2013**

[752] R. **Möllenberg**, *Monte Carlo Study of the Fast Neutron Background in LENA*, Diploma thesis, Technische Universität München, **2009**

[753] D. **Hellgartner**, *Lepton track reconstruction in LENA and attenuation length measurements in liquid scintillators*, Diploma thesis, Technische Universität München, **2011**

[754] S. **Lorenz**, *Discrimination of Neutral Current Background in a Future Long-Baseline Experiment with LENA*, Diploma thesis, Universität Hamburg, **2012**

[755] S. M. **Franke**, *Optical Purification Study of the LAB-based Liquid Scintillator for the JUNO Experiment*, Ph.D. thesis, Technische Universität München, **2019**

[756] T. **Benesch**, *Schlüsselkonzepte zur Statistik [Key Concepts for Statistics, in German]*, Spektrum Akademischer Verlag, Heidelberg, ISBN 978-3-8274-2771-7, **2013**

[757] J. **Krumm**, *Savitzky-Golay Filters for 2D Images [website]*, **2001**

[758] B. **Naranjo**, *Light Concentrators for Spherical Detectors: Tiling and Timing – talk at the Advances in Neutrino Technology 2014 conference (ANT 2014), 2014-09-24*, **2014**

[759] W. **Zinth** and H.-J. **Körner**, *Physik III: Optik, Quantenphänomene und Aufbau der Atome [Physics III: Optics, Quantum Phenomena and Atomic Structure, in German]*, Oldenbourg Wissenschaftsverlag, 3rd ed., ISBN 978-3-486-24054-2, **1997**

[760] **Evonik Röhm GmbH – Business Unit Performance Polymers**, *Lichttechnik: Richtwerte und Anwendungen – PLEXIGLAS GS, PLEXIGLAS XT [Lighting technology: Standard Values and Applications – PLEXIGLAS GS, PLEXIGLAS XT, in German]*, Tech. Rep., Evonik Röhm GmbH, **2000**

[761] A. C. **Ruggeri**, *Materials in optical sensors – talk at the New and Enhanced Photosensor Technologies for Underground/underwater Neutrino Experiments (NEPTUNE) workshop, Naples 2018-07-19*, **2018**

[762] **Röhm GmbH – Acrylic Products**, *Technical information: PLEXIGLAS GS, UV transmitting – Clear 2458, Clear 2458 SC*, **2020**

[763] **Filmetrics (KLA Corporation)**, *Reflectance Calculator [website]*, **2021**

[764] F. **Pedrotti**, **et al.**, *Optik für Ingenieure – Grundlagen [Optics for Engineers – Basics, in German]*, Springer, Berlin, Heidelberg, 3rd ed., ISBN 978-3-540-27379-0, **2005**

[765] J. **Brack**, **et al.**, *Characterization of the Hamamatsu R11780 12 in. photomultiplier tube*, Nucl. Instrum. Methods Phys. Res. Sect. A, vol. 712, pp. 162–173, arXiv:1210.2765, **2013**

[766] **Dassault Systèmes**, *SolidWorks 2013 SP3.0 [software]*, **2013**

[767] **Project Gutenberg** (editor), *The Project Gutenberg eBook of The King James Bible, Plain text UTF-8*, Project Gutenberg, **2011**

Acknowledgments

First, I would like to thank Prof. Stefan Schönert and Prof. Franz von Feilitzsch for the opportunity to write this thesis at E15.
I am much obliged to my supervisor Prof. Lothar Oberauer for the guidance and freedom that enabled me to pursue this work, the valuable input and extensive discussions that contributed to it, and the perseverance to help me complete it.
My special thanks go to German Beischler for his substantial contributions to the development of the optical module, which took a far greater amount of time than apparent by the size of the corresponding chapter, and his cheerful disposition. I greatly enjoyed our collaboration. Harald Hess and Rainer Stoepler provided us valuable pointers on how to properly conduct FEA pressure simulations.
Many have contributed to the making of SPAX, and all are deserving of my deepest gratitude. Mario Schwarz, Konstantin Schweizer, Hanna Kellermann and Jill Kaindl helped in the construction and characterization of the setup in numerous ways. Laszlo Papp built all LED pulsers and repaired a voltage divider in a way which should not be physically possible. Harald Hess and the E15 workshop performed the upgrade to a Faraday cage, manufactured a multitude of components for the optical setup, and relocated the entire dark box way too many times – once in pieces, once from the sudden impending doom of fire prevention, and once airborne. Gunther Korschinek kindly allowed SPAX to find shelter in the MLL in between two of these relocations, Hermann Hagn provided his formidable experience to help eliminate noise sources, and Patrick Krause and Felix Henningsen lent me fast SiPMs to characterize SPAX.
The photomultiplier characterization at the LNGS would not have been possible without the incredible help of Gyorgy Korga. It is as much his achievement as it is mine. The time of our collaboration was one of the happiest of my life and I have learned much from him. Oleg Smirnov, Laszlo Papp, and Paolo Lombardi also contributed greatly to these measurements, which were enabled through the generous permission of Gioacchino Ranucci in the first place. I owe my thanks also to Sean Grullon, who provided me with a sample of the R11780-HQE for the characterization.
I am grateful to Razmik Mirzoyan for having had the chance to work at the Max-Planck-Institut für Physik for nearly a year to conduct our joint research of the light emission from PMTs and to Hanna Kellermann and Jill Kaindl which aided me in this endeavor.
All of this would not have been feasible without the hard work of the LENA group, which laid the framework for much of this thesis, with too many to mention.
My special thanks go to my proofreaders and advisors, who dared to face the beast. Most notably, Konstantin Schweizer, who bore the brunt of the work and with whom I have enjoyed many a discussion about philosophy; Quirin Meindl, who advised me what to leave out; Andreas Zöller, who read 130 pages which were then omitted following Quirin's council; and Birgit Neumair, who made certain that the part about neutrino physics was based on actual physics.

I would also like to thank Dieter Renker, Razmik Mirzoyan, Bayarto Lubsandorzhiev and Oleg Kalekin for fruitful discussions about photomultipliers and photosensors which have taught me much.
Sincere thanks go to all members of E15 who have helped me throughout the years and made the time of my dissertation very enjoyable.
I would particularly like to thank the many companions with whom I have had the pleasure to share my office, most of all: Quirin – one of my best and oldest friends – for his humor, Birgit Neumair, Hanna Kellermann and German Beischler for their cheerful nature, Jürgen Winter for being Jürgen (words cannot describe it), and Timo Lewke for providing much needed distractions from tedious tasks.
Our team assistants, most notably Maria Bremberger, Sabine Kaps, Paola Mucciarelli, Sabine Wenzel and Elke Krüger, deserve my heartfelt thanks for removing all that red tape from the cogs of the chair and helping me with all things administration.
In addition, I want to express my thanks to my friends, who have accompanied me, listened to me and laughed with me through all this time. You have helped me more than you may think.
None deserve my gratitude more than my family. Without the continuous support of my parents Erika and Kurt as well as my sister Vanessa this would not have been possible. Above all, my thanks go to my girlfriend Carina, who helped me see this through and who never stopped believing in me.

Statistics

- 0.80 metric tons of coffee consumed (conservative estimate)
- 1503 literature sources (Mendeley entries only; 767 used in work)
- 19 ring binders of material, 4 lab notebooks (plus digital notes)
- 0.538 Bibles [767] of code (PEST and ROOT++ only, not counting deprecated code, comparison by bytes)
- 5.0 TB of data
- 0.546 Bibles of thesis text (printed characters only)