TUM School of Natural Sciences
Technische Universität München

Bioluminescence in the Pacific Ocean Neutrino Experiment: Shedding Light on the Deep Sea

Kilian Holzapfel

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:
Prof. Dr. Andreas Weiler

Prüfer*innen der Dissertation:
1. Prof. Dr. Elisa Resconi
2. Prof. Dr. Peter Fierlinger

Die Dissertation wurde am 27.06.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 06.07.2023 angenommen.

Abstract

Neutrino astronomy is a highly promising field in high-energy particle and astrophysics, requiring the use of cubic-kilometer scale detectors to capture the light signatures of rare neutrino interactions. Currently, the future Pacific Ocean Neutrino Experiment (P-ONE) is being developed to transform the depths of the ocean abyss off the west coast of Vancouver Island into a neutrino observatory. Despite being positioned at depths exceeding 1.5 km, the P-ONE remains inhabited by various forms of life, including the captivating phenomenon of bioluminescence. Bioluminescence refers to the emission of light by living organisms and presents a significant challenge as a light background for neutrino telescopes. However, deep-sea neutrino telescopes also offer a unique opportunity for biologists to study the diverse life thriving in this vast and largely unexplored habitat, which encompasses the largest expanse on Earth. To assess the feasibility of the future P-ONE site, pathfinder missions such as Strings for Absorption Length in Water (STRAW) and STRAW-b (Strings for Absorption Length in Water b) have been deployed. This thesis focuses on the development of the data acquisition system for STRAW-b and introduces the camera system employed, providing insights into the camera optics. Additionally, an innovative image recognition algorithm aided by machine learning is presented, enabling the detection and characterization of bioluminescent events. This advanced image recognition capability facilitates a comprehensive analysis of bioluminescence, including its correlation with water currents and the quantification of bioluminescent organism concentrations. Notably, this thesis represents the first measurement of the spectral population of bioluminescent organisms in the deep sea.

Zusammenfassung

Neutrinoastronomie ist ein vielversprechendes Forschungsfeld in der Hochenergie- und Astrophysik, das den Einsatz von Detektoren mit einem Volumen von einem Kubikkilometer und mehr erfordert, um die Lichtsignale seltener Neutrino-Wechselwirkungen zu erfassen. Derzeit wird das künftige Pacific Ocean Neutrino Experiment (P-ONE)[1] entwickelt, um die Tiefsee vor der Westküste von Vancouver in ein Neutrinoobservatorium zu verwandeln. Trotz der Tiefe von über 1.5 km beherbergt P-ONE verschiedene Formen von Leben, einschließlich des faszinierenden Phänomens der Biolumineszenz. Biolumineszenz bezieht sich auf die Lichtemission lebender Organismen und stellt eine bedeutende Herausforderung als Hintergrund für Neutrinoteleskope dar. Neutrinoteleskope für die Tiefsee bieten jedoch auch eine einzigartige Gelegenheit für Biologen Biolumineszenz, in diesem weltweit größten Lebensraum, der weitgehend unerforschten ist, zu untersuchen. Um die Machbarkeit von P-ONE am geplanten Standort zu bewerten, wurden zwei Erkundungsmissionen Strings for Absorption Length in Water (STRAW)[2] und STRAW-b (Strings for Absorption Length in Water b)[2] durchgeführt. Diese Arbeit konzentriert sich auf das entwickelte DAQ-Systems[3] und für STRAW-b, stellt das eingesetzte Kamerasystem vor und gibt Einblicke in die Kameraoptik. Darüber hinaus wird ein innovativer Bilderkennungsalgorithmus mit Hilfe von maschinellem Lernen vorgestellt, der die Erkennung und Charakterisierung von biolumineszenten Ereignissen ermöglicht. Diese Bilderkennung erleichtert eine umfassende Analyse der Biolumineszenz, einschließlich ihrer Korrelation mit den Wasserströmungen und der Quantifizierung der Konzentration biolumineszenter Organismen. Besonders bemerkenswert ist, dass diese Arbeit die erste Messung der spektralen Population biolumineszenter Organismen in der Tiefsee darstellt. Diese Bilderkennung ermöglicht eine umfassende Analyse der Biolumineszenz, einschließlich ihrer Korrelation mit Wasserströmungen und der Quantifizierung der Konzentrationen biolumineszenter Organismen. Des Weiteren stellt diese Arbeit die erste Messung der spektralen Population biolumineszenter Organismen in der Tiefsee dar.

[1] P-ONE (engl.: Pacific Ocean Neutrino Experiment) Neutrino-Experiment im Pazifischen Ozean

[2] STRAW (engl.: Strings for Absorption Length in Water): Vertikale Leinen für die Absorptionslänge im Wasser

[3] DAQ (engl.: data acquisition): Datenerfassung

Contribution

The following presents a summary of the author's contributions within the scope of this thesis.

Section 2.1 and 2.2: part of the development team in both P-ONE pathfinder missions: STRAW and STRAW-b (Strings for Absorption Length in Water b).

Section 2.2.1: development of the PMT-Spectrometer modules, including simulation, optimization, design, production, testing, and calibration. Moreover, project lead of the art integrations in STRAW-b called UNDERCURRENTS [1] in collaboration with Fine-arts Academy in Munich and the Sonderforschungsbereich 42 (SFB42) [2].

Section 2.2.2, 2.2.3, and 2.2.4: development of the master control software (MCTL) that operates nine of ten[4] STRAW-b instruments from the low-level module operations towards the data acquisition (DAQ) system which is fully integrated into the *Oceans 2.0* platform[5].

Section 2.2.5: development of a user-friendly software package to simplify data handling, processing, and analyzing the data of STRAW-b.

Chapter 3: integration a camera system in two STRAW-b modules, including the optimization of the control software to guarantee a high performance which resulted in an uptime of over 95 % since the commissioning.

Chapter 4: part of the development team of the flasher system for the optical modules in STRAW-b. Responsible for software integration, including safety mechanisms to prevent excessive exposure to sensitive PMTs.

Chapter 5: derivation of the optical projection of the camera system, which includes the lens optics and the distortion of the spherical pressure housing.

Chapter 6: leading the development and optimization of an machine learning (ML) aided bioluminescence detection algorithm. This algorithm exhibits exceptional performance by accurately identifying bioluminescence signals that

[4] The Wavelength-shifting Optical Module (WOM) is operated individually.

[5] *Oceans 2.0* a publicly available database that is part of the Ocean Networks Canada (ONC) infrastructure

extend across a minimum of five pixels within the images captured by the camera.

Chapter 7: developed the analysis software that utilizes recorded images to track the relative movement of moorings between two modules and quantify the contamination caused by biofouling and sedimentation based on the reduction in brightness of bioluminescence emissions. A literature study was also conducted to compile potential countermeasures for addressing biofouling in P-ONE.

Chapter 8: development of the simulation and analysis framework that quantifies significant aspects of bioluminescence observed in the recorded images. Derivation of the relationship between the bioluminescence rate at the detector, the water flow speed, and the concentration of bioluminescent organisms.

Contents

1 Introduction

The vast expanse of the Universe remains an enigma that has fascinated humanity for centuries. At the high-energy end of this mysterious domain, the cosmos is awash with cosmic radiation of various types. These particles include photons in the form of gamma rays, neutrinos, and charged particles, which continuously bombard the Earth's atmosphere without leaving a clear indication of their origins or the mechanisms that produce them. Cosmic rays have energies of up to 10^{20} eV, making them more than a million times more energetic than those generated in human-made particle accelerators such as the Large Hadron Collider (LHC) at CERN which reaches a collision energy of 13.6 TeV after the latest upgrade in April 2022 [3]. Despite intensive research, fundamental questions regarding high-energy cosmic radiation still need to be answered, such as its origin, production mechanisms, and role in cosmic evolution.

Neutrinos, unlike photons, only interact weakly, making them closely linked to the acceleration and interaction of cosmic rays in astrophysical environments. Therefore, neutrino astronomy is one of the most promising fields for high-energy particle and astrophysics. However, the low cross-section of neutrinos requires enormous detectors, and even with the cubic-kilometer volume of IceCube, the number of astrophysical neutrinos detected remains relatively low. 2013 the first high-energy astrophysical neutrinos were detected [4, 5]. However, it took another four years to observe the first compelling neutrino source candidate, a flaring gamma-ray blazar[1] called TXS 0506+056 [6–8]. As recently presented by the IceCube collaboration, the first evidence of high energy neu-

[1] Blazars are a type of elliptical galaxy that is characterized by a supermassive black hole at their center which produces a high-energy jet of particles that points in the direction of the Earth.

trino emission from the nearby galaxy NGC1068 has been shown, supporting the case for broad astronomy based on neutrinos [9].

These astrophysical observations provide strong motivation for the implementation of additional neutrino telescopes on a cubic-kilometer scale, such as KM3NeT (Cubic Kilometer Neutrino Telescope) [10] and P-ONE (Pacific Ocean Neutrino Experiment) [11]. Despite being deployed in the ocean abyss at depths below 1.5 km kilometers, neutrino telescopes such as KM3NeT and P-ONE are still in the presence of life in the abyss. One phenomenon of the deep sea is bioluminescence, which refers to the emission of light by living organisms such as bacteria, plankton, and fish. While bioluminescence is a crucial aspect of oceanography, it poses a significant background for neutrino telescopes which requires a deep understanding of the physical and biological processes that contribute to bioluminescence, as well as the use of advanced modeling techniques to simulate and account for its effects. Despite this challenge, bioluminescence in neutrino telescopes offers a unique opportunity for biologists to study life in the deep sea, which is by far the most extensive habitat on Earth and still remains largely unexplored.

Two pathfinder missions, STRAW and STRAW-b (Strings for Absorption Length in Water b), have been deployed at the future site of P-ONE to investigate environmental parameters and feasibility for a cubic-kilometer scale neutrino detector [12, 13]. While STRAW measured the optical properties of water in Cascadia Basin [13], STRAW-b utilized cameras and spectrometers to investigate bioluminescence properties in the deep sea.

The primary focus of this thesis centers around the key elements of the STRAW-b pathfinder. It delves into the low-level module operations, the DAQ system, and the hardware components. A significant emphasis is placed on analyzing bioluminescence using the cameras installed in select modules of STRAW-b. A novel technique utilizing machine learning has been developed for bioluminescence detection. This detection technique enables the analysis of bioluminescence from various perspectives. Notably, this thesis presents, for the first time, the measurement of the spectral population of bioluminescent organisms in the deep sea. The obtained results support the feasibility of a long-term neutrino

telescope based at the Cascadia Basin.

The thesis is structured as follows:

In Chapter 2, a concise introduction to the P-ONE detector is given, which covers its two pathfinder missions, namely STRAW and STRAW-b, that have been accomplished successfully. The chapter puts special emphasis on the data pipeline developed for STRAW-b, which involves a reliable DAQ system and a user-friendly interface that simplifies the process of data analysis.

In Chapter 3, a summary of the cameras incorporated in STRAW-b is given, encompassing their hardware elements, the circumstances in which images are recorded, and a synopsis of the measurement's overall performance. It is noteworthy that these cameras have exhibited an uptime of more than 99 % since December 2021, indicating their high degree of operational reliability.

In Chapter 4, the LED flasher system used in conjunction with STRAW-b's cameras is introduced. This includes a detailed discussion of the hardware components involved, the DAQ system used to collect data from the flasher, and the conditions that trigger flashes.

In Chapter 5, the camera optics is discussed, wherein an analytical model is presented that considers the lens and distortion of the pressure housing's glass. This model serves as a foundation for interpreting the images discussed in chapters 7 and 8, where it establishes a metric space for specific measurements.

In Chapter 6, the method for image processing to extract bioluminescence emissions from the STRAW-b cameras is outlined. This process combines deterministic cluster detection with novel machine-learning classification techniques, enabling automated adaptation to intrinsic camera differences. Furthermore, with the ability to achieve a high bioluminescence detection probability with as few as 5 pixels in the image, this approach is highly effective for bioluminescence analysis.

In Chapter 7, the focus is on characterizing the movement of the STRAW-b mooring system and the effects of biofouling and sedimentation. These two crucial aspects affecting the camera-based analysis will be relevant for the future P-ONE

telescope. Furthermore, the chapter presents practical strategies for mitigating and detecting bio-fouling in future installations.

In Chapter 8, the bioluminescence phenomenon is explored through the use of images captured by the cameras in STRAW-b. The chapter delves into various aspects of bioluminescence, including the periodic variations in the baseline emission rates, the spatial distribution of the bioluminescent features in the images, the spectral population of the organisms emitting the light, and the processes that lead to bioluminescence and how they relate to the current speed and concentration of the bioluminescent organisms.

The appendix of this thesis contains supplementary material to aid in understanding the methodology and results presented in the main text. Specifically, detailed documentation of the DAQ file structure of STRAW-b is provided in Appendix A. Furthermore, Appendix B summarizes and discusses the sensor and module malfunctions encountered during the STRAW-b mission. Additional information on cluster detection and classification is available in Appendix C. Moreover, the bioluminescence catalog, which summarizes the emission spectra of bioluminescent organisms, is found in Appendix D. Finally, detailed information on water flow simulation and measurements can be found in Appendix E.

2 Pacific Ocean Neutrino Experiment and its Pathfinder Missions

The Pacific Ocean Neutrino Experiment (P-ONE) is a joint initiative among universities and research institutions in the USA, Canada, UK, Poland, and Germany to construct a multi-cubic-kilometer neutrino telescope in the Pacific Ocean that will provide new insights into the Universe's highest energies [11, 14]. The project began in 2017 and has since developed into an international collaboration. The primary partner of P-ONE is Ocean Networks Canada (ONC), an initiative from the University of Victoria that operates the most extensive oceanographic infrastructure globally, with the NEPTUNE

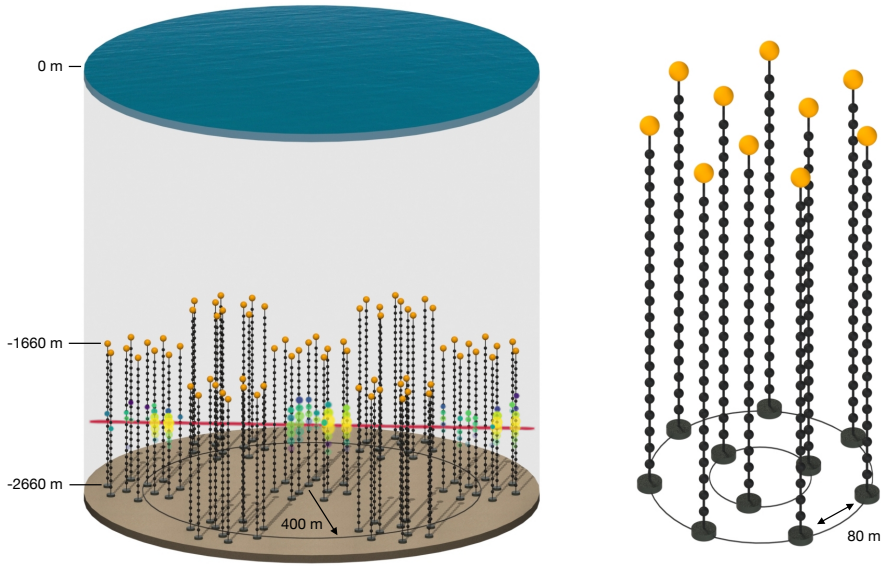

Figure 2.1: The P-ONE detector is designed as a modular structure consisting of seven clusters, as shown in the left image, with a visualization of a track event. A single cluster, as depicted on the right, comprises ten strings, each equipped with twenty optical modules.

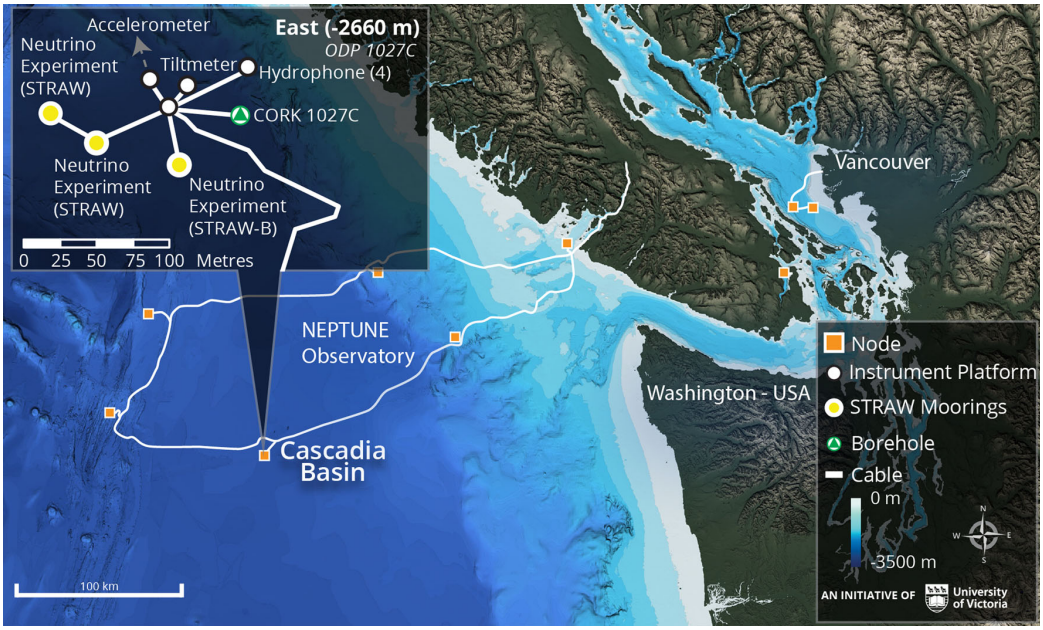

Figure 2.2: Map of ONC's NEPTUNE Observatory. P-ONE will be located at the Cascadia Basin node, where the pathfinder projects STRAW and STRAW-b are currently in operation. Image courtesy of ONC.

[15] ocean observatory providing power and data connections to various deep ocean sites off the west coast of Vancouver Island. NEPTUNE is based on an approximately 800 km long cable for power transmission and fiber optic communication and plugs devices in the deep sea directly into the Internet [11, 14–16]. A map of the cable and the ocean sites are depicted in Figure 2.2.

P-ONE has been primarily designed to detect astrophysical neutrinos at multi-TeV energies, with a particular emphasis on horizontal tracks. Its modular structure enables easy scalability and allows for the installation of the array in varying sizes and stages. As illustrated in Figure 2.1, the detector is structured into individual clusters. Each P-ONE cluster comprises ten strings with 20 optical modules [11, 14].

The P-ONE neutrino observatory has been selected to be situated in the Cascadia Basin, located off the coast of British Columbia, Canada, as shown in Figure 2.2. This site has been under observation since 2018 by two pathfinder experiments, namely Strings for Absorption Length in Water (STRAW) (Section 2.1)

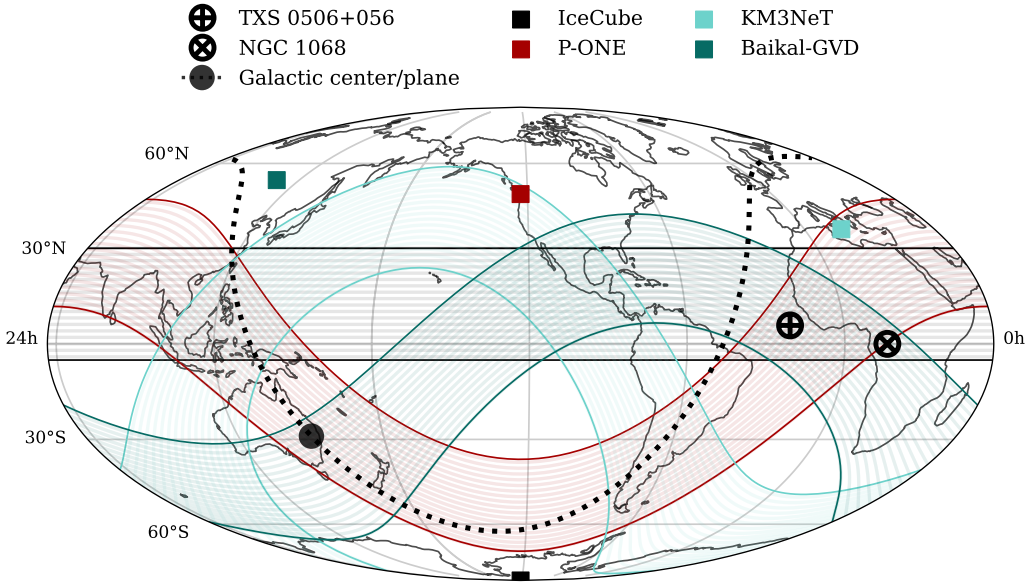

Figure 2.3: Global map of existing and under-construction neutrino telescopes and their corresponding horizontal coverage, where the detectors have the highest sensitivity for high-energy neutrinos and are not affected by Earth absorption [17]. Plot courtesy of Lisa Schumacher.

and STRAW-b (Section 2.2). The pathfinder missions have provided critical insights into the optical and environmental suitability of the P-ONE large-scale neutrino telescope. Following up on the two pathfinder experiences, the design of P-ONE started. The blueprint of the P-ONE mooring line is currently under design, with a deployment scheduled in 2024 [11].

The P-ONE prototype will include P-ONE digital optical modules to measure the emerging Cherenkov radiation by muon and neutrino-induced processes, as well as P-ONE calibration devices to provide in-situ calibration of the detector. The prototype line will also feature external geometry calibration units to verify the envisioned calibration principles [11].

The P-ONE project capitalizes on the strengths of the Canadian oceanographic community, and the successful deployment of the STRAW and STRAW-b (Strings for Absorption Length in Water b) projects have demonstrated the potential of the Cascadia Basin site for large-scale neutrino telescope experiments. By expanding our observable window of the Universe to the highest energies, the

P-ONE neutrino observatory is expected to significantly contribute to our understanding of astrophysical phenomena, alongside other neutrino telescopes such as ANTARES [18], KM3NeT [10], GVD [19], and IceCube [20]. Combining these observatories into a single distributed planetary instrument called the high-energy Planetary Neutrino Monitoring System (PLEνM) would enhance detection probability by up to two orders of magnitude compared to IceCube's current capabilities, covering almost the entire sky as depicted in Figure 2.3 [17].

2.1 First Pathfinder: Strings for Absorption Length in Water (STRAW)

The Strings for Absorption Length in Water (STRAW) is the first pathfinder mission towards P-ONE. As the name indicates, STRAW primary purpose is to assess the attenuation length of water in the wavelength range between 350 nm and 600 nm and the optical background at the site where the P-ONE will be constructed. The optical background is a combination of the bioluminescence of deep-sea living organisms and the ^{40}K dissolved in the salty water. Besides, gaining experience in the challenges of deep-sea experiments is another purpose of the STRAW mission [12, 13].

STRAW has two 146 m long mooring lines spaced 37 m apart horizontally where each mooring hosts four modules as shown in Figure 2.4. From the overall eight modules on the two mooring lines, three modules are called Precision Optical Calibration Module (POCAM) and emit calibrated, intense, adjustable, isotropic, nanosecond light flashes [21, 22]. The remaining five modules are called STRAW Digital Optical Module (sDOM) and detect the light with one upwards and one downwards facing 3-inch photomultiplier tube (PMT) [22]. The modules are placed along the mooring line above the sea floor at a depth of 30 m–110 m. STRAW 's geometry allows to record the light intensity of each POCAM at different distances with the sDOM, which translates to a relative intensity measurement instead of a more complex absolute measurement [12, 13, 21, 22].

In June 2018, after eight months of development, the two strings were success-

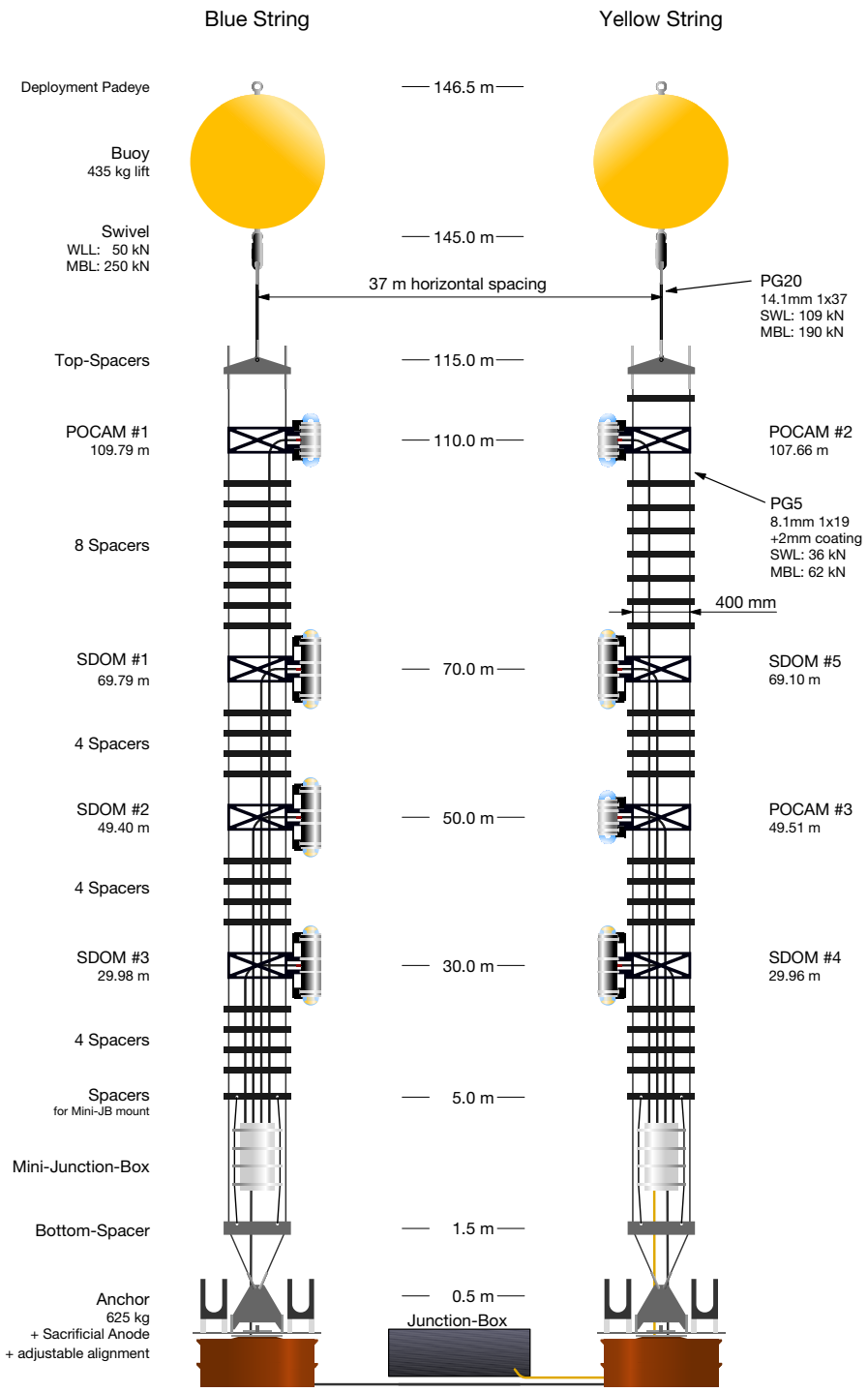

Figure 2.4: Sketch of the two STRAW mooring lines. To measure the water's attenuation length, a POCAM emits an isotropic, intense, and ns-long light pulse which gets detected by two PMTs in each sDOM.

fully deployed by ONC [12]. After the commissioning phase, the detector is continuously operating with an exemplary lifetime of 98.3 % from March 2019 to March 2021 [13]. The analysis of two years of data yields an optical attenuation length of about (28 ± 2) m at 450 nm [13]. By extracting the signature of ^{40}K decays from the data, a salinity of (2.5 ± 1.4) % is found, which covers the salinity measured by ONC of (3.482 ± 0.001) % [13]. Bioluminescence was identified as the significant light background, which can even saturate the PMT and data acquisition (DAQ) capabilities [13].

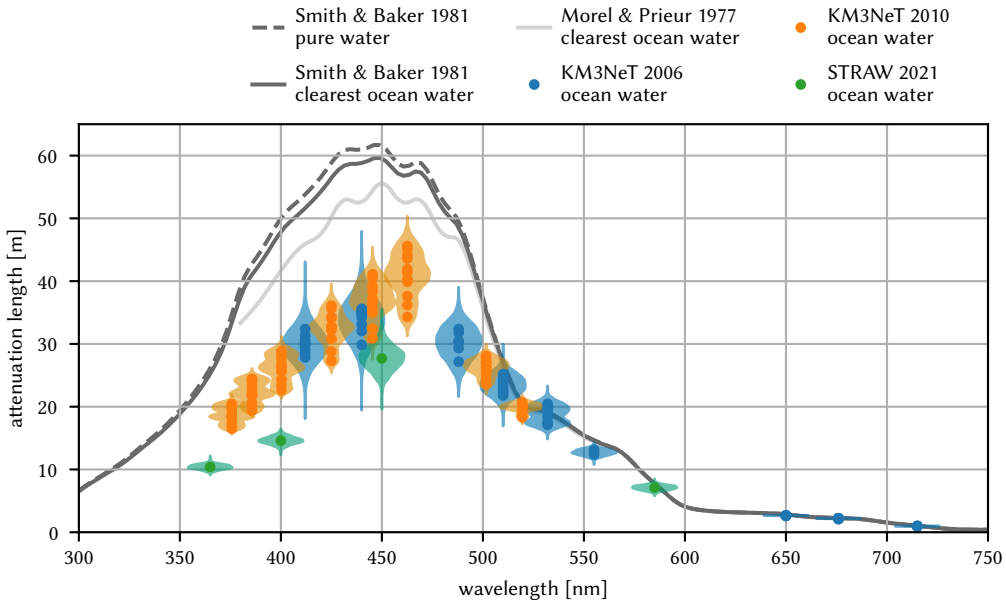

Figure 2.5: Attenuation length measurement of STRAW[13] compared to data from Cubic Kilometer Neutrino Telescope (KM3NeT)[23, 24] and Smith and Baker [25]. The results are presented using violin plots that display the Kernel Density Estimation (KDE) of the published attenuation values (points), along with their associated uncertainties. In 2006, KM3NeT published data from the AC9 device manufactured by WETLabs, which monitored the attenuation length between 1999 and 2003. The 2010 KM3NeT data refers to measurements taken at various sites, covering a depth range of 2000 m–4900 m. Smith and Baker's method involved purifying water to distinguish the different components of water absorption. Their reported attenuation length exceeds those of STRAW and KM3NeT.

2.2 Second Pathfinder: Strings for Absorption Length in Water b (STRAW-b)

STRAW-b is the second pathfinder for P-ONE deployed in 2020, positioned approximately 40 m away from STRAW. It consists of ten modules mounted on a 444-meter-long mooring line, as illustrated in Figure 2.6. Each module, except the Wavelength-shifting Optical Module (WOM), is equipped with a 13-inch high-pressure-resistant glass housing for added protection. To ensure redundancy, each module is connected to its own data cable, which provides network connection via glass fibers and power supply with copper cables and is connected to the junction box positioned at the base of the line.

Section 2.2.1 presents an overview of various modules utilized in a range of measurements that are detailed in Section 2.2.2. These modules and measurements are controlled and operated by a distributed control system (DCS) described in Section 2.2.3, which employs a time-over-threshold (ToT) method for detecting nanosecond-range signals, as explained in Section 2.2.4.

2.2.1 Instrumentation of STRAW-b

Various types of instruments have been installed on the STRAW-b mooring line, each utilizing a standardized module setup that incorporates electronics for power management, networking, and a single-board computer. The standardized module setup also incorporates environmental sensors, including for internal pressure, temperature, and humidity which serve as checkpoints for the successful deployment and long-term monitoring of the module's health. In addition, an accelerometer and electronic compass measure the module's orientation, shielding information of the mooring movements with the currents [28].

Two LiDAR modules aim to verify the attenuation length measurements from STRAW and separately determine scattering length and absorption length. It has a pulsed laser diode as an emitter, and a lens focuses the back-scattered photons onto a micro PMT device. The laser and PMT are mounted on a two-axis gimbal allowing a change in the orientation in azimuth and zenith [28, 29].

Figure 2.6: Sketch of the STRAW-b mooring line (left) with the position of each module. The Standard Module 3 (upper right) that hosts the art project *radioamnion* [26] and the PMT-Spectrometer (middle right and lower right [27]) are shown exemplarily.

One Muon-Tracker module utilizes two separate 2x2 arrays of plastic scintillator tiles positioned between a certain distance. If a muon is propagating through the scintillator, light is emitted. Two SiPM matrices measure the induced light in each scintillator at diagonal corners. Approximate directional information can be obtained in the case of time-correlated events in separate scintillator tiles [28, 30].

Two PMT-Spectrometer modules are characterizing the bioluminescence emissions. Each module hosts 12 PMTs with different wavelength filters to monitor the spectrum and intensity of nearby bioluminescent organisms. Furthermore, a low-light camera is included to capture the bioluminescence, including the position. This thesis focuses mainly on the camera (Chapter 3), processing the images (Chapter 6) and the interpretation of the images (Chapter 8). In addition, the module hosts a Mini-Spectrometer from Hamamatsu similar to the Mini-Spectrometer module outlined in the following [28, 31].

The Mini-Spectrometer module is composed of five Mini-Spectrometer from Hamamatsu [32] and the same low light camera as the PMT-Spectrometer module. The CMOS[1]-based Mini-Spectrometer is a complementary device to the spectrum measurement. It has lower sensitivity to light but a finer resolution on the spectra than the PMT-Spectrometer. However, it turned out that the sensitivity is too low to measure bioluminescence emissions [28, 33].

In addition to its scientific objectives, specific modules of the project also serve as hosts for art installations created in collaboration with the Sonderforschungsbereich 42 (SFB42) [2], a research group consisting of both artists and physicists based in Munich. This collaborative project, known as UNDERCURRENTS [1], involves a group of artists from the Fine-arts Academy in Munich under the guidance of Diogo da Cruz and Jol Thomson. An aspect of this collaboration is to engage and educate the general public about P-ONE. Another aspect is to facilitate a creative partnership between the two disciplines which can turn science into captivating art and forge a deeper public understanding of science [34]. As one result, the sound art project *radioamnion* (https://radioamnion.net was developed and built together with Jol Thomson. During each full moon, *ra-*

[1] CMOS: complementary metal-oxide-semiconductor

dioamnion broadcasts sonic transmissions of invited artists. Because the module hosting *radioamnion*[2] has a connection failure, the initial planned acoustic transmission is now fulfilled by LEDs in other modules where a Fourier transform decodes the acoustic signal into a light pattern.

2.2.2 Operation of the Instruments in STRAW-b

In STRAW-b, the data acquisition (DAQ) records the data either continuously or schedules measurements to collect data. All modules contain ambient sensors to monitor the conditions inside the pressure housing, and some modules host additional sensors, as outlined in Section 2.2.1. This section presents a summary of the various measurements, while Table 2.1 provides information on the measurements, including their readout rate, sensor types, and corresponding hosting modules.

The standardized module, which each module builds on, is equipped with a 3-axis accelerometer and a 3-axis magnetometer that provide information on movement and orientation. Additionally, the module includes sensors for monitoring pressure, humidity, temperature, and power consumption within the pressure housing. The DAQ also keeps track of various software and system settings. Sensor data is collected every 10 s, while software and system settings are only saved when changes are made to reduce unnecessary data. The collected data is labeled with the data-product-code *SMRD*[3] and is stored in *Oceans 2.0*[4], a publicly available database that is part of the ONC infrastructure.

The Mini-Spectrometer and PMT-Spectrometer modules contain a camera (Chapter 3) and a mini-spectrometer from Hamamatsu [32], respectively. These sensors operate synchronously because they share the same serial peripheral interface (SPI)-bus for communicating. The camera has a readout limitation, resulting in one image and spectrum being recorded every 99.5 s. More detailed information on the measurement sequence can be found in Section 3.4. Moreover, the PMT-Spectrometer 2 captures a set of images with enabled LEDs (Chapter 4) in the neighboring module (LiDAR 2) once per hour

[2] *radioamnion* is hosted by the *Standard Module 3* (Figure 2.6)
[3] The internal file structure of all data-product-codes is summarized in Appendix A
[4] *Oceans 2.0* - or the newer version *Oceans 3.0*

Table 2.1: Summary of STRAW-b measurements including readout rate, sensor type, and hosting module.

Sensor	Modules	Rate	Measurement
3-Axis Accelerometer	All Modules	0.1 Hz	continuously
3-Axis Magnetometer	All Modules	0.1 Hz	continuously
PTH[a]	All Modules	0.1 Hz	continuously
6 Powermeter[b]	All Modules	0.1 Hz	continuously
3 Temperature	All Modules	0.1 Hz	continuously
Camera	Mini- & PMT-Spec.	~1/90 Hz	continuously
Mini-Spectrometer	Mini- & PMT-Spec.	~1/90 Hz	continuously
Mini-Spectrometer[c]	LiDAR	2 times[c]	hourly
16 PMTs (ToT counts)	PMT-Spec.	1 kHz	continuously
16 SiPMs (ToT events)	Muon-Tracker	>1 GHz[d]	continuously
LiDAR (ToT events)	LiDAR	>1 GHz[d]	hourly

[a]Pressure-Temperature-Humidity sensor
[b]The Powermeter measures for each channel current and voltage separately
[c]use for Laser calibration before and after each LiDAR measurement
[d]Only the events are stored. The time resolution of the time-to-digital converter (TDC) is below 1 ns.

to track its position (Chapter 7). The measurement with enabled LEDs ceased after the module failure of the LiDAR 2 at the end of 2021. The LED activation and configuration information is logged by the DAQ and stored in a file with data-product-code *SMRD*[3]. Additionally, the camera and Mini-Spectrometer each have their own files, which are labeled with data-product-codes *MSSCD*[3] and *MSSD*[3], respectively.

The PMT-Spectrometer continuously records the PMT signals in the low precision mode of the Trigger Readout Board (TRB) [35], which involves counting ToT events rather than storing the time and duration of each event separately as described in Section 2.2.4. The ToT counts are stored with a frequency of 1 kHz in the PMT-Spectrometer. Once a day (around midnight UTC), a threshold scan is conducted for the ToT measurement, and the measurement is interrupted. During the threshold scan, the TRB is rebooted to prevent communication issues

that may arise. The DAQ system continuously monitors communication, and if an issue arises, it triggers a reboot of the TRB and restarts the measurement. The PMT-DAQ files (data-product-code *PMTSD*[3]) rotate every hour to ensure file sizes remain below 100 MB. The PMT-Spectrometer 2 is not used for this measurement as it has a broken FPGA that applies the thresholds, making it impossible to read out the PMTs.

The Muon-Tracker utilizes the high precision mode to track photon detection in the scintillators with the SiPM. In this mode, all ToT events, i.e., events with timestamps and durations over the threshold, are stored as described in Section 2.2.4. Due to the low dark rate of the SiPMs, the files rotate only once a day, remaining below 100 MB, with the data-product-code *MTSD*[3]. As with the PMT-Spectrometer, a threshold scan is performed once a day, and the DAQ system continuously monitors the TRB, triggering a restart of the measurement if needed.

The LiDARs perform various measurements, including a daily scan of the surrounding volume with the gimbal rotating over 2π. This scan is repeated without the enabled laser to gather information on in-situ afterpulsing characteristics. Additionally, a daily laser scan is performed, where the gimbal moves to a fixed position, and the laser changes it's pointing relative to the optical axis of the detection optics. When no other measurements are ongoing, two LiDAR measurements with the enabled laser and one without the enabled laser are taken per hour. During each measurement, events are recorded for 60 s as the gimbal moves in the same direction as the laser scan position. Each measurement is saved in one Hierarchical Data Format (HDF5) file identified by the data-product-code *LIDARSD*[3]. In addition, the raw HLD files (can be multiple for a single measurement) from the TRB are stored and synchronized with the ONC database.

Additional information is provided in Appendix A, which includes an overview of the data-product-codes, rollover intervals, and tables summarizing the internal HDF5 file structure of all files.

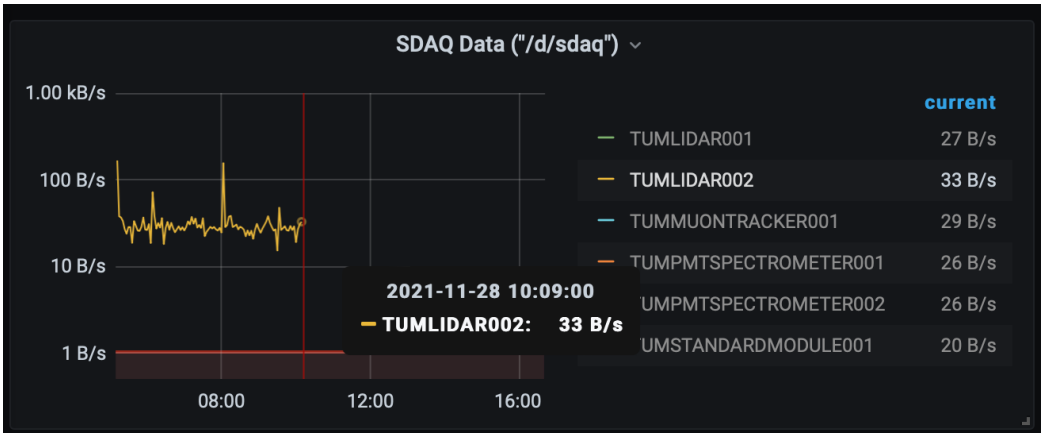

Figure 2.7: The figure shows a screenshot of the real-time monitoring system based on the *TIG Stack* (Telegraf, InfluxDB, Grafana) [37, 38], displaying the last connection to the LiDAR 2 module. This system collects sensor data in real time from various modules and allows for easy visualization and detection of any anomalies through customizable Grafana dashboards.

2.2.3 Distributed Control System and Data Acquisition

The master control software (MCTL) operates all STRAW-b modules (except for the WOM) and runs on each of them as an autonomous controller implemented as a distributed control system (DCS) [36]. Its primary goal is to ensure the safe and streamlined operation of all sensors, both in manual and automated modes. To this end, the MCTL includes various functionalities such as safety control, logging, sensor operation and monitoring, bus management, inter-module communication, DAQ, and measurement execution and scheduling.

The safety control prevents any actions that could potentially damage the sensors or cause data loss. For example, if a LED is activated in the same or neighboring module, the safety control will block the powering of PMTs to avoid exposing them to intense light. The logging function generates a detailed record of system events, warnings, and errors, which is essential for effective debugging, especially in case of rare issues. The sensor operation manages low-level functions required to operate individual sensors, while the sensor monitoring ensures their proper functioning and triggers corrective action in case of any malfunctions.

The monitoring system of each module is integrated into a real-time monitoring system for the entire detector, which uses the open-source tools Telegraph, InfluxDB, and Grafana to collect, store, and visualize sensor data in real-time [37, 38]. The combination of the tools is also called the *TIG Stack*[38]. Telegraph is used to gather data from various sources, including sensors and services, and send it to InfluxDB, which stores the data as a time series. Grafana is used to create customizable dashboards to visualize the data in real time, allowing for easy monitoring and detection of any anomalies. Additionally, the system has the ability to send alerts in real-time based on predefined thresholds, ensuring that any issues are addressed promptly [37, 38]. Figure 2.7 shows the last connection of the LiDAR 2 module as a demonstration of the monitoring system's performance. A summary and discussion of module and sensor malfunctions within the STRAW-b system can be found in Appendix B.

Some electronic components share a communication bus where only one component can send data at a time, such as SPI (serial peripheral interface) or I^2C (Inter-Integrated Circuit). The bus-management function keeps track of all communications and distributes the available resources to prevent interference. The inter-module communication relies on network protocols like secure shell protocol (SSH) and is used to coordinate measurements and ensure safety.

The DAQ collects sensor values and stores data in HDF5 files using lossless data compression and checksums natively supported by HDF5 [39]. The file rollover is initialized by the DAQ for continuous measurements after fixed periods, such as hourly or daily. The structure of HDF5 files used by the DAQ is designed to organize time series data. A time series is a collection of measurements made over time, and each HDF5 group in the file represents a unique time series. The first dimension of all datasets within a group corresponds to time, with each HDF5 group containing a *time* dataset that stores the timestamp of the time series. This structure ensures that all datasets within a group have the same size. Figure 2.8 illustrates the generic file structure. In addition, metadata are stored in the attributes of the HDF5 file, group, and datasets. Appendix A summarize the internal file structure for all generated files [39].

The measurement execution and scheduling function manages high-level tasks

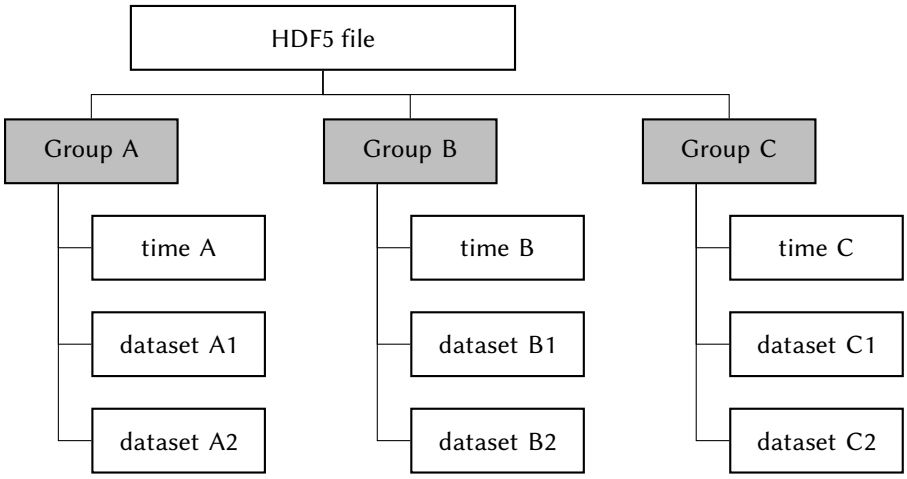

Figure 2.8: Generalized HDF5 file structure of STRAW-b's DAQ. It exemplarily shows three data groups (A, B, C), each with two datasets and the time dataset.

such as initiating, executing, and stopping measurements.

After recording the data on the modules, it is then downloaded and integrated into *Oceans 2.0*. *Oceans 2.0* - or the newer version 3.0 - is the data base (DB) for all experiments connected to the ONC infrastructure and the DB including all data is publicly available [40]. MCTL is equipped with an interface to ONC 's shutdown procedure in case of the modules need to be powered off. In such a scenario, MCTL stops all measurements, stores the data, and puts the module in a safe state concerning software and hardware. The final shutdown command to the operating system is triggered by ONC 's shutdown procedure.

2.2.4 Fast signal digitization

Both STRAW and STRAW-b utilize a similar method for digitizing signals from PMTs or SiPMs where a nanosecond precision is required. This method involves measuring the ToT of a signal to efficiently reduce data. When a single exceeds a user-defined threshold, the starting time and duration of each transgression are measured by a TDC with sub-nanosecond precision. To achieve sensitivity to single photons, the threshold is set at an appropriate level. The PMTs are read out using a TRB (version TRB3sc) that employs a TDC [35].

The readout system allows the detectors to operate in two modes: high-

precision and low-precision. In the high-precision mode, the exact timestamp of each pulse is recorded with sub-nanosecond precision relative to the master clock, enabling analysis at the single-photon level. In contrast, in the low-precision mode, only the pulse counts detected within a specified time interval are recorded.

However, storing the ToT data for long-term monitoring is not always feasible due to the large data size it generates (up to 100 MB per minute). Therefore, the DAQ operates in the low-precision mode where only the signals threshold transgressions are counted. The counters are then recorded at a frequency of approximately ~33 Hz and ~1 kHz in STRAW and STRAW-b, respectively.

To summarize, the readout system can operate in either high-precision mode, where each event is stored individually, or low-precision mode, where only the count of single events is recorded. The LiDAR and Muon-Tracker operate in high-precision mode since both require sub-nanosecond precision and the dark rates are optimized to a few Hz, allowing long-term monitoring in high-precision mode. Typically, the sDOM and PMT-Spectrometer operate in low-precision mode. During a measurement run of STRAW, where the POCAMs emit light, the sDOM records in high-precision mode to enable attenuation measurements [21, 22].

2.2.5 A Software Package for the Data Analysis

Managing and analyzing large datasets, like those generated by the STRAW-b detector, can be a challenging task. To simplify this process, a user-friendly software package called *strawb* has been developed in Python. This package provides tools for syncing data from the ONC database to a local database, as well as importing and analyzing the data. Currently, the *strawb* package repository is not public and is hosted on GitHub within the P-ONE group. This allows only members of the group to access and contribute to the package.

Strawb offers several key features, such as the ability to synchronize available data from the ONC database to a local database. This feature allows users to share their code with others, with the package taking care of synchronizing the required files without downloading all files from the ONC database. By down-

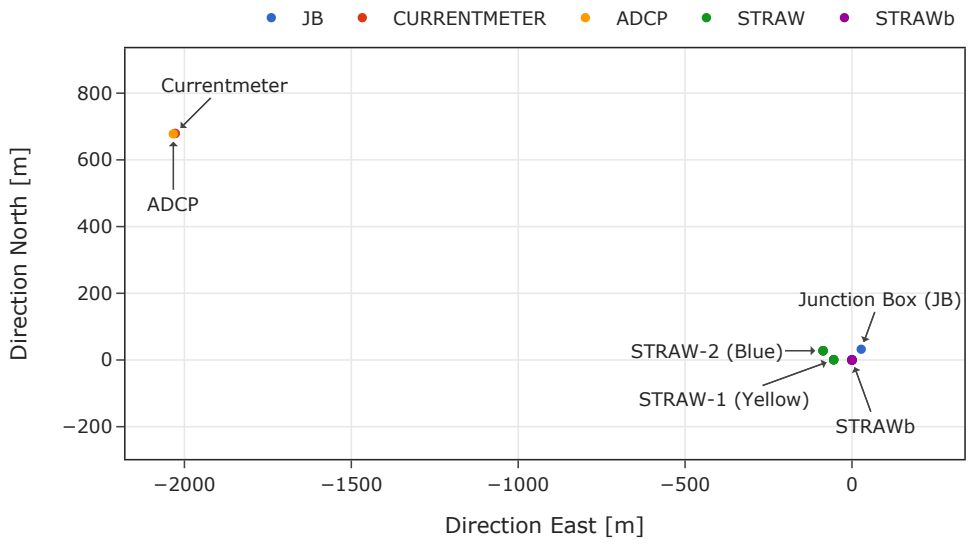

Figure 2.9: Locations of selected devices at the Cascadia Basin Node, including the STRAW and STRAW-b detectors, as well as the current meter and acoustic Doppler profiler (ADCP), which booth measure the water current. Position data accessed from the ONC database[40] using the *strawb* software package.

loading only the required files, *strawb* minimizes the amount of data transfer and optimizes working with large datasets.

Additionally, *strawb* provides basic import functionalities for different file types, making it easy to import and process data from the STRAW-b detector. This feature reduces the time and effort required for data management.

Strawb also includes basic and advanced analysis tools that every user can use to streamline their work with the data. These tools include functions for data visualization, quality control checks, and event selection, as well as advanced analysis techniques such as signal processing and machine learning.

Additionally, *strawb* can access includes functionalities to access other sensors available on the ONC database, such as current profilers and temperature sensors. It also includes functionalities to access metadata from the ONC database, such as position information of the devices, as shown in Figure 2.9. With its user-friendly interface and comprehensive suite of tools, *strawb* provides an efficient and effective solution for managing and analyzing large datasets from

the STRAW-b detector.

3 Camera System in STRAW-b

We'd still be getting only black-and-white pictures from our digital cameras.

Larry Scarff, The New York Times [41]

STRAW-b hosts in total three cameras in three different modules, both PMT-Spectrometer and the Mini-Spectrometer. A sketch of the STRAW-b mooring including the position of the three modules is illustrated in Figure 2.6. All cameras are oriented upwards in the direction of the mooring line, with the PMT-Spectrometer camera integration depicted in Figure 3.1. The cameras provide visual insights into the environment, such as sedimentation attached to the module housing (Chapter 7) and bioluminescence (Chapter 8).

This chapter aims to provide an overview of the hardware components of the cameras, the conditions under which images are captured, and a summary of the overall performance of this measurement, which has demonstrated an uptime of over 99 % since 1st December 2021.

3.1 The Camera: a Synergy with the IceCube-Upgrade

STRAW-b uses a customized camera, originally developed in collaboration with Arducam for the IceCube Upgrade and IceCube Gen2 [42, 43]. This decision was made to leverage the synergy between neutrino detectors, where hardware must operate in remote locations with limited space, low power consumption, and in

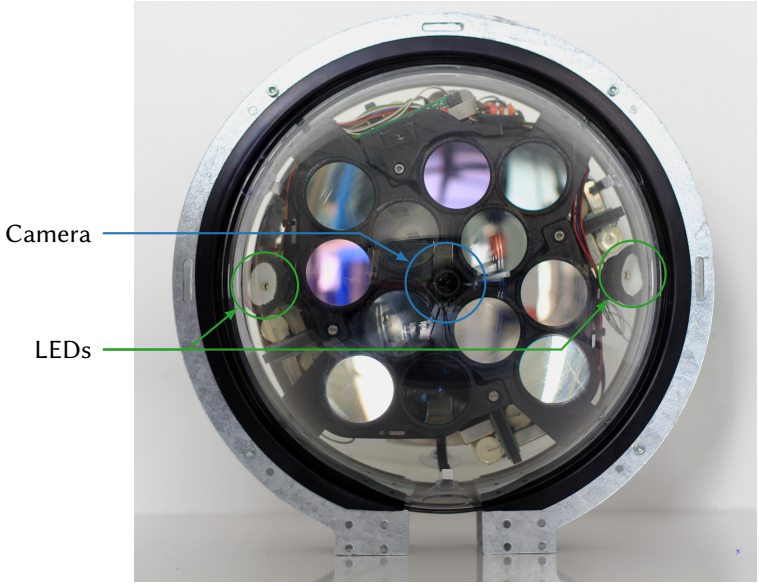

Figure 3.1: Pictures of a fully assembled PMT-Spectrometer with the marked position of the camera and light-emitting diodes (LEDs). The module connects to the string at the gaps of the metal rings (bottom). After the deployment, this side of the module is pointing upwards.

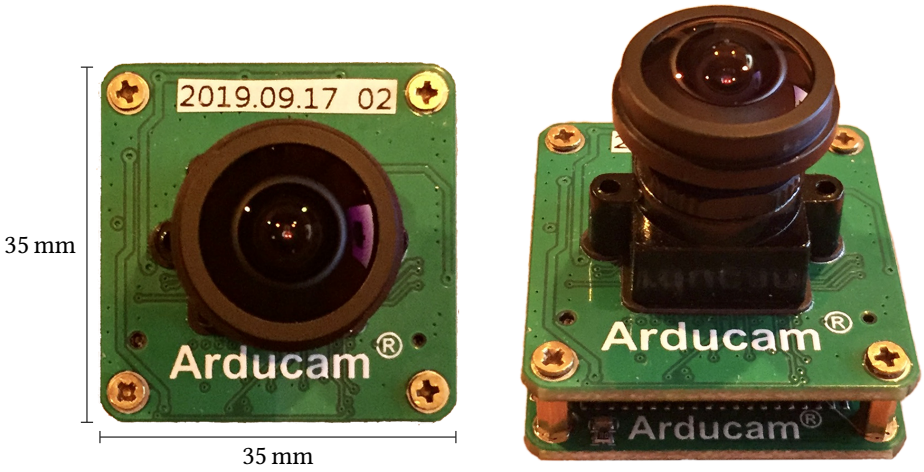

Figure 3.2: The camera developed for the IceCube Upgrade and IceCube Gen2 [42]. It is a customized system based on the Arducam with an IMX225 color image sensor from Sony and a 1312 x 979 pixels resolution. For the optics, the camera uses an Arducam lense M25170H12 with a field of view (FoV) of ~170° [43].

low light conditions. For reference, Figure 3.2 displays pictures of the camera, along with its dimensions.

In the IceCube Upgrade and IceCube Gen2, the cameras will be integrated into the modules deployed in the Antarctic ice to study various parameters, such as the optical properties of the glacial and drill hole ice, and detector geometry, including the location and orientation of the optical modules and cables. These parameters are crucial for calibrating IceCube Monte Carlo simulations and improving event reconstructions [42, 43].

3.1.1 Hardware Components

The camera is composed of several hardware components, as detailed in Table 3.1. At the core of the camera is the IMX225LQR-C CMOS image sensor, which is controlled by a complex programmable logic device (CPLD). The CPLD communicates with the module's data acquisition (DAQ) system via serial peripheral interface (SPI), while the DAQ system itself runs on an Odroid single-board computer with a Linux operating system.

A wide-angle M12 lens with a diameter of 22 mm and a focal length of 1.7 mm provide a FoV of approximately 170°. The refraction of the glass housing lowers the FoV by some degrees. In the PMT-Spectrometer modules, the FoV will be further reduced by the mounting due to the minimal space. In contrast, STRAW-b's third module with camera provides an obstacle-free view for the camera [42]. In Chapter 5, the system's optics are summarized with more details.

Table 3.1: Main components of the camera developed for the IceCube Upgrade and used in STRAW-b [42].

Component	Type	Manufacture
CMOS	IMX225LQR-C	SONY Semiconductor Solutions Co
CPLD	LCMXO2-1200HC	Lattice Semiconductor Co.
RAM	MT48LC4M16A2P-6A	Micron Technology Inc.
Lense	M25170H12	Arducam

A wide-angle M12 lens with a diameter of 22 mm and a focal length of 1.7 mm provides a FoV of approximately 170°. However, due to the refraction of the glass housing, the field of view (FoV) is reduced by a few degrees. In the PMT-Spectrometer modules, the FoV is further reduced by the mounting due to the limited space. On the other hand, the Mini-Spectrometer module provides an unobstructed view for the camera. For more details about the system's optics, please refer to Chapter 5.

One of the key features of the IMX225LQR image sensor is its high sensitivity, which Sony claims to be approximately twice that of the previous product IMX238LQJ. Moreover, Sony's image sensors have some of the most responsive pixels, including the one in this camera [44]. The sensor has a pixel size of 3.75 µm by 3.75 µm and an active diagonal of 6.09 mm, with a native resolution of 1312 x 993 pixels. After cutting black, inactive, and dummy pixels, the resolution becomes 1305 x 977 pixels.

Color processing, specifically the demosaicing method described in Section 3.3, introduces boundary effects that may affect the image quality. The datasheet recommends cutting a margin of 8-9 pixels symmetrically from all sides [45]. Moreover, the image sensor is designed for a final resolution of 1280 x 960 pixels, or a 4:3 aspect ratio, which is a standard for television screens and computer monitors [46]. To summarize, the resolution of 1305 x 977 pixels is reduced to 1297 x 977 pixels, and after color processing and the boundary effect margin are taken into account, the final image resolution becomes 1280 x 960 pixels. Table 3.2 provides a summary of the different resolutions and the dimensions of the image sensor.

3.1.2 Exposure Time and Gain

This section covers the camera configuration settings, which mainly consist of exposure time and gain. The exposure time parameter p is used to configure the exposure time of the camera. However, the translation of this parameter to actual time units is unknown for this camera. To determine this mapping, reverse engineering is required. One approach is to capture a scene from which the start and stop time of the picture can be extracted, for example, by using a

Table 3.2: Resolution and corresponding dimensions of a Sony IMX225LQR Image Sensor. Each pixel has a size of 3.75 µm by 3.75 µm. The image sensor has more pixels than the final image resolution, as some are black or dummies. Also, the color processing requires a margin to cut boundary effects, resulting in a recommended resolution of 1280x960 pixels [45].

	Pixel			Dimensions		
Type	Horiz.	Vertical	Total	Horiz.	Vertical	Diagonal
native	1312	993	1.30M	4.92 mm	3.72 mm	6.17 mm
effective	1305	977	1.27M	4.89 mm	3.66 mm	6.11 mm
active	1297	977	1.27M	4.86 mm	3.66 mm	6.09 mm
w/o margins	1280	960	1.23M	4.80 mm	3.60 mm	6.00 mm

monitor, moving parts, or blinking LEDs. However, this requires a specific setup and accessibility to the camera, which is not possible for the deployed cameras. Another approach is to time the command execution for different exposure time settings.

In this study, the timing approach was used. Hundreds of pictures were taken, and the timings were extracted. A linear mapping function was fitted to the timings, resulting in

$$t(p) = p * m, \tag{3.1}$$

where m is the best-fit parameter. The maximum exposure time for the camera was determined to be 63.5 s, with $p = 31751$. The value of m was calculated to be 1/500 s.

As previously stated, the second configuration parameter refers to gain. This can be directly applied to the image sensor to enhance the dynamic range of an image and utilize the camera's bit-depth in low-light scenarios such as capturing faint bioluminescence in the depths of the ocean. Two types of gain exist: analog gain, which amplifies the analog pixel signal before digitization, and digital gain, which enhances the values after digitization. Following the datasheet of the image sensor, the gain is measured in in decibel (dB) [45]. Gain A in dB is defined

as

$$A = 10 \cdot \log_{10}\left(\frac{X_{out}}{X_{in}}\right) \tag{3.2}$$

where X_{in} is the value before the gain and X_{out} the value after. Transforming the function yields the scaling between X_{in} and X_{out}

$$X_{out} = 10^{A/10} \cdot X_{in} \tag{3.3}$$

with non-linear scaling factor $10^{A/10}$. The image sensor applies only analog gain for gain up to 30 dB. Above 30 dB up to the maximum of 72 dB, it is a combination of the 30 dB analog gain plus the required digital gain. A drawback of the digital gain is that it decreases the recordable intensity range without increasing the resolution. Furthermore, a digital gain can be applied later offline with greater adaptability and without lowering the intensity range. Generalized, digital gain removes some information irreversibly, like digital zoom in photography. However, this 30 dB threshold was not known initially, and the gain was set to 50 dB initially to adapt to the low light environment.

The image sensor utilizes solely analog gain for gains of up to 30 dB. For values above 30 dB, up to the maximum of 72 dB, it employs a combination of the 30 dB analog gain and the necessary digital gain. One downside of digital gain is that it can narrow the recordable intensity range without improving resolution. Additionally, digital gain can be applied post-capture with greater flexibility and without diminishing the intensity range. Generally, digital gain causes irreversible information loss, similar to digital zoom in photography. Nonetheless, the 30 dB limit was not initially known, and the gain was initially set to 50 dB to adjust to the low-light environment. On the 22nd of December 2021 at 14:30 UTC, the correction took place by lowering the value to 30 dB and disabling the digital gain [45].

3.1.3 Limitations and Issues

The camera developed for the IceCube Upgrade, and IceCube Gen2 could benefit from improvements in several areas, including enhanced support through Arducam, improved documentation, firmware, and software. However, it is worth

noting that custom products, especially those produced in small quantities, may have limited support and documentation, which is a common challenge in the custom manufacturing industry.

One major issue with the camera is its slow readout speed. It takes approximately 36 s to download a 1.2M pixel image over the SPI-bus, which is relatively long. In some cases, it even takes as long as 47 s, although the reason for this difference is unknown. Attempts to speed up the download by increasing the SPI transmission frequency failed, resulting in corrupt pictures. Unfortunately, the cause of the limited readout speed remains unknown.

Another issue with the camera is the firmware, which has closed source code and contains a bug that causes the camera to freeze occasionally. The more images taken, the higher the chance of failure. While power-cycling the camera after each image taken works reasonably well, it cannot prevent image losses. It must be considered if this power-cycle has a non-optimal influence on individual components of the camera or if it speeds up the aging.

Attempting to resolve this issue and make the camera operational for module integration consumed valuable time, and there was not enough time remaining to calibrate the cameras before deployment. Switching to a more performant camera was not feasible at this stage due to late detection of issues and dependencies on other components in the module.

3.2 Color Image Sensor

CMOS and CCD color image sensors measure intensity and not color directly. This means that the image is initially black and white.

In order to capture and measure color, a Color Filter Array (CFA) is placed over the image sensor. This array comprises a series of color filters arranged in a repeated pattern, where each pixel is assigned a specific filtered light. Ideally, the different types of filters are evenly distributed in two perpendicular directions, alternating at every other pixel position. This means that each individual pixel captures the intensity of a particular filtered color, while neighboring pixels represent different colors in a regular grid pattern. The filter matrix is sometimes

referred to as a mosaic.

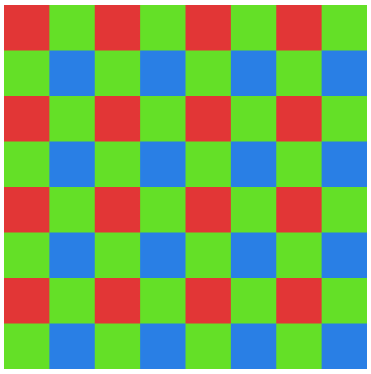

Figure 3.3: A Bayer Filter Mosaic on the pixel array of an image sensor. The filter mosaic transforms a natively black-and-white image sensor into a color image sensor [47].

In the present day, there exist various arrangements of filters, each offering its own set of advantages. When it comes to consumer products, the filters are usually designed to replicate the three primary colors of the human eye: red, green, and blue. One of the most widely recognized filter patterns is known as the Bayer Filter, which was patented by Bryce Bayer in 1976 [47]. This filter pattern, depicted in Figure 3.3, consists of a unit cell containing one red, one blue, and two green pixels, as the human eye is most receptive to the color green. The cameras used in STRAW-b incorporate the Bayer Filter technology.

Other filter types introduce pixels without a filter to enhance low-light performance while still maintaining color detection. These filters are called RGBW, where W stands for white. In one version, the white pixel replaces one of the two green pixels, resulting in a white density of 25 %. However, other arrangements with higher white densities are also available. Additionally, some filters introduce other colors, such as red, green, and blue.

For deep-sea applications, the enhanced low-light performance of white pixels and colors adapted to the environment can be advantageous. For instance, if a camera is to be integrated into P-ONE, it may require colors that are better suited for the environment.

3.3 Demosaicing of the Raw Image

To reconstruct a full-color image where all colors are calculated for all pixel positions, any raw image captured with a CFA must undergo demosaicing. This is a non-trivial process, and different algorithms have been developed, each perform-

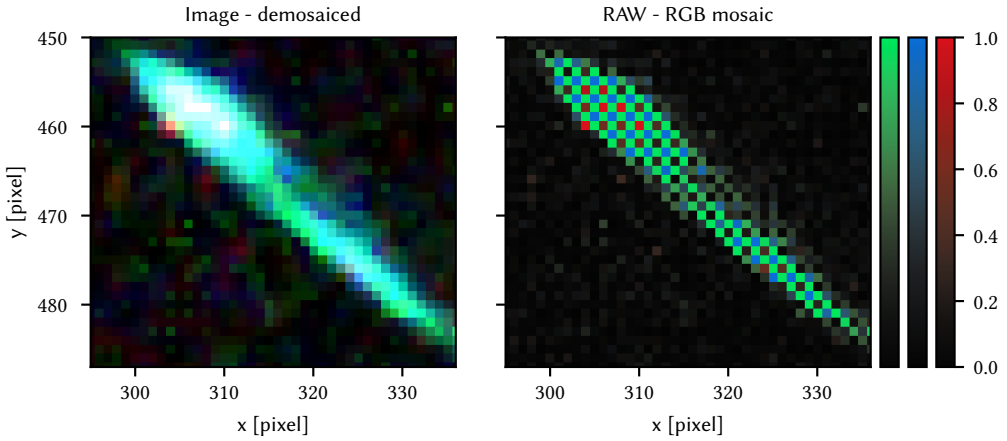

Figure 3.4: Demosacied image (left) and visualization of the RAW RGB Bayer mosaic (right), where both show the same scene with bioluminescence. Typically colored photos result from two steps (see Section 3.2). First, an image sensor with a Bayer mosaic of color filters records an image. Second, a demosaicing algorithm calculates the intensity of all colors at every pixel. In this picture, structures are smeared out as the exposure time is about one minute (see Section 3.4), currents are in the range of 10 cm/s, and bioluminescence is in the order of seconds usually.

ing better under certain conditions. Demosaicing algorithms should compensate for several characteristics, such as avoiding false color artifacts, preserving image resolution, low computational complexity for fast processing, and more. However, all methods face the inherent problem of estimating colors at pixels where the color is unknown.

To produce the demosaiced images shown in this thesis, the software package *OpenCV* and a bi-linear interpolation with edge detection [1] was used [48]. An example of a raw and a demosaiced image is provided in Figure 3.4.

However, it is worth noting that demosaicing is only used for image visualization purposes. Further analysis is based on the raw image data, where all pixel intensities are used independently of their color, or the different colors are treated separately. This approach has several advantages, including the fact that demosaicing artifacts do not need to be considered, and the analysis is independent of any specific algorithm and its strengths and weaknesses in different conditions.

[1] The bi-linear interpolation with edge detection called *COLOR_BAYER_BG2RGB_EA* in *OpenCV* [48].

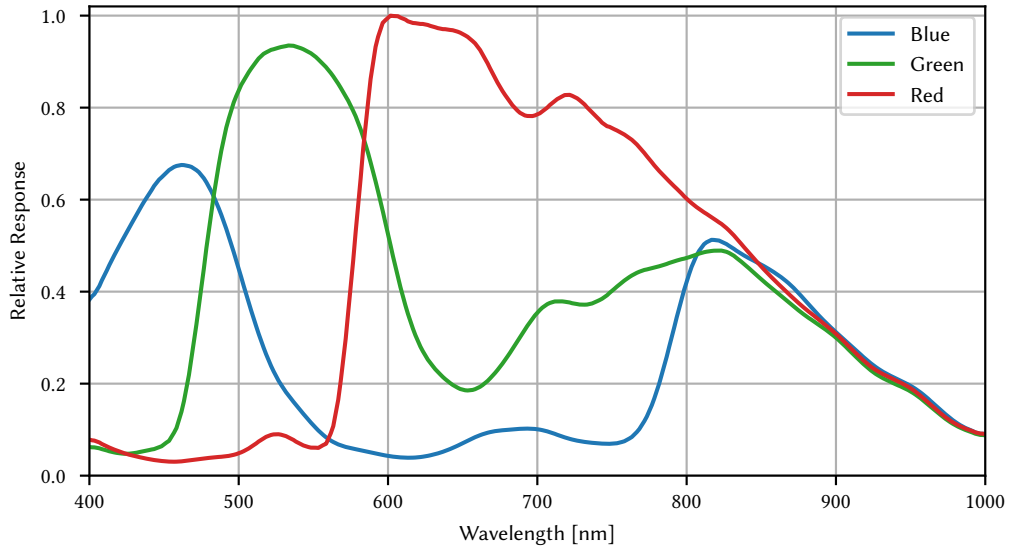

Figure 3.5: Transmittance spectra from the different color filters of the color image sensor (IMX225LQR-C), which is used in STRAW-b's cameras. Filters are arranged in a classical Bayer Pattern, allowing the camera to take colored images (see Section 3.2) [45].

Additionally, demosaicing often assumes perfect filtering of colors, whereas in reality, filters have a wider transmittance spectrum. An illustration of the spectra of the IMX225LQR-C image sensor is shown in Figure 3.5 [45].

3.4 Measurement Sequence

For communication, the camera uses a SPI-Bus, which it shares with other devices in the module. However, a SPI-Bus only permits communication with one device at a time. As a result, the camera's measurements must be synchronized with a Mini-Spectrometer from Hamamatsu [32]. The measurement scheme is based on the principle of maximizing exposure time, for two reasons. Firstly, because both devices are not highly sensitive to low-light conditions in the deep sea. Secondly, it minimizes the downtime ratio of the camera since the image transferred always takes the same time per image, irrespective of the exposure time. As detailed in Section 3.1.1, the camera's maximum exposure time is 63.5 s, and the image transfer takes around 36 s, resulting in 99.5 s for a full

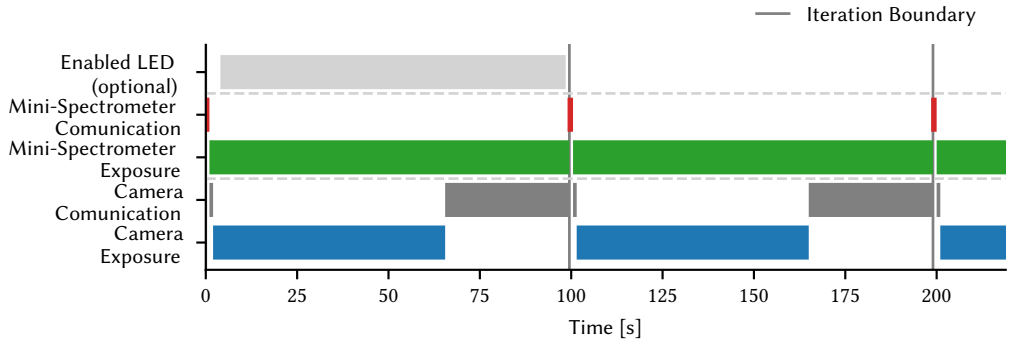

Figure 3.6: Gantt Chart of the Camera and Mini-Spectrometer measurements, which have to be synchronized as they use the same communication bus SPI. One iteration needs 99.5 s to maximize the camera's exposure to download time ratio. Occasionally, a measurement can be performed with enabled LEDs, illustrated in the first iteration of the sequence.

capture.

To ensure efficient and reliable measurements, the following sequence is implemented:

- The Mini-Spectrometer is configured and starts a 99.5 s exposure.
- The camera is configured and starts a 63.5 s exposure.
- The image is transferred from the camera, which takes approximately 36 s.
- The Mini-Spectrometer data is transferred in about 1 ms.
- The sequence restarts

The sequence is visualized in a Gantt chart in Figure 3.6.

Additionally, the camera captures images with enabled LEDs to track the neighboring module's position and sedimentation. For further details regarding the LEDs configuration and DAQ summaries, the reader should refer to Chapter 4.

Table 3.3: Total taken images per device and the measurement up-time. The up-time is calculated from the actual recorded images and the theoretical number of captured images based on the 99.5 s/image under ideal conditions for the measurement period.

Module	Images	Up-Time	Measurement Period
Mini-Spectrometer 1	36176	83.2 %	2021-04-22 to 2021-06-11 [a]
PMT-Spectrometer 1	519856	97.2 %	2021-04-28 to 2022-12-31 [b]
PMT-Spectrometer 2	512975	95.9 %	2021-04-22 to 2022-12-31 [b]

[a] Connection lost to Mini-Spectrometer 1 module on 2021-06-11.
[b] Considered period for up-time, measurement is ongoing at the time of writing this thesis.

3.5 Commissioning, Data Acquisition and Performance

The Data Acquisition (DAQ) system of STRAW-b, as described in section 2.2.3, is responsible for collecting and storing data from various devices in a module. To achieve reliable and automated storage of data, the DAQ uses *hdf5* files that allow for storing sensor data and metadata in a single file, with built-in compression support [39]. To avoid large file sizes, data is losslessly compressed, and a fixed file rollover interval of 1 h is defined for the camera DAQ. The hourly files have a size of approximately 50 MB, ensuring efficient storage of the data.

In the database Oceans 2.0 from Ocean Networks Canada (ONC), camera files are identified by the data-product code *MSSCD*. These files contain raw images, along with exposure and gain settings, capture times, and other metadata that are crucial for data interpretation. Since LED emissions are also relevant to the camera, the LED data are stored in a separate file with their data-product code, *SMRD*, in ONC's database. For more information about the flasher DAQ, readers are directed to Section 4.3. Additionally, details on the internal Hierarchical Data Format (HDF5) file structure can be found in Appendix A.

To ensure high-quality data and maximize uptime, multiple improvements have been made to the measurement procedure since the modules were first initialized. As depicted in Figure 3.7, the camera performance on all three modules has

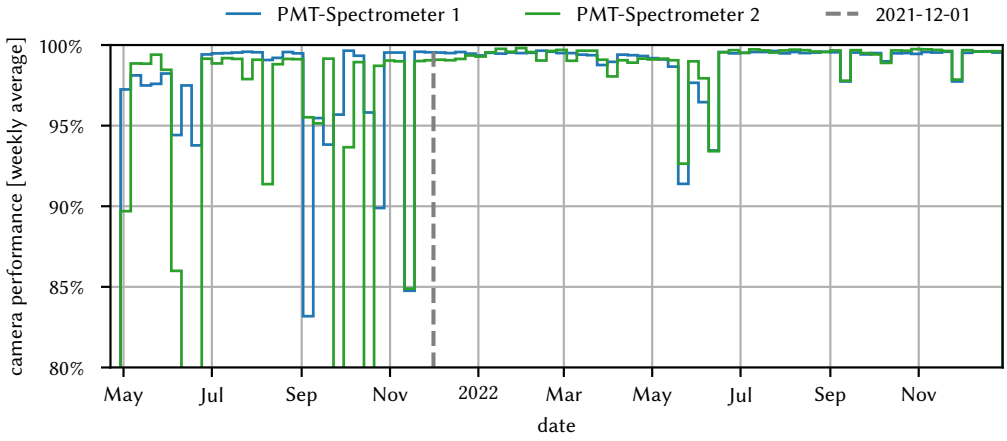

Figure 3.7: Performance of the cameras as an average per week. A camera measurement cycle is fixed to 99.5 s due to the synchronization with the Mini-Spectrometer. Both devices share the same SPI bus, where only one device can use the bus simultaneously. A performance of 100 % for a period represents that every 99.5 s one picture is stored in the DAQ over the entire period. Any issue which prevents image recordings causes fewer pictures to be taken and result in lower performance. This can be failures of the camera firmware, the DAQ, or broken data after the transmission to shore. Table 3.3 shows the overall performance per device.

notably enhanced over time, and the total number of captured images and overall uptime are listed in Table 3.3. PMT-Spectrometer 1 started capturing images a few days later than the Mini-Spectrometer and PMT-Spectrometer 2 to ensure compatibility with the active PMTs and related DAQ. Unfortunately, on June 11th, 2021, the Mini-Spectrometer module experienced an failure and stopped communicating. Despite this setback, the remaining two modules have continued to record data, including images, up to the current time of writing this thesis. A summary and discussion of module and sensor malfunctions within the STRAW-b system can be found in Appendix B. Investigations suggest that the malfunctioning modules are booting properly, and the issue lies specifically with the network connection. Furthermore, the the statistic of ONC about connector failures is approximate 10 %, something considered too high for P-ONE. Consequently, penetrators and connectors are not employed within the P-ONE system to mitigate this issue.

As summarized in Section 3.1.3, the camera has some firmware bugs that cause

picture losses. Unfortunately, the code is not open source, and the company producing the camera does not provide any support, making it impossible to solve the issue. Initially, the DAQ had some issues, especially with the delicate camera firmware, which resulted in lower performance. However, several updates have been made to the DAQ over time, solving the known issues and increasing its overall robustness against camera malfunctions. Most of these improvements were implemented until October 2021 when the commissioning was completed.

Unrelated to the DAQ, the modules have faced some power shutdowns that lasted from hours to days. Sometimes, ONC had to switch off the power for maintenance or infrastructure updates, and unplanned power outages also occurred. Historical uptimes and the current status of ONC infrastructure are available on ONC's status page[2].

As of the writing of this thesis, over one year of data has been recorded, and the measurement on both remaining devices is ongoing. Overall, the two modules have shown up-times of over 95 % for the whole period and over 99 % since the 1st of December 2021 (as marked in Figure 3.7). In total, the three cameras have taken more than 1 million images, adding approximately 850 images per camera every day.

[2] ONC's status page: https://oceannetworkscanada.statuspage.io and historical uptime: https://oceannetworkscanada.statuspage.io/uptime

4 Flasher System for the Optical Modules

A camera in Strings for Absorption Length in Water b (STRAW-b) requires artificial light to image its surrounding in the permanent darkens of the deep sea. Therefore, STRAW-b utilizes up to six LEDs per camera, as illustrated in Figure 4.1. The cameras face upwards in the deployed modules, with the neighboring module above being typically centered in the captured image. This neighboring module contains four light-emitting diodes (LEDs) split into two sets, each set equipped with a blue and a white LED. One set points downward toward the camera, while the other points horizontally. Additionally, both PMT-Spectrometer modules have two extra LEDs that function as a flasher system, providing illumination for the surrounding area. These LEDs are situated near the glass and point in the same direction as the camera, with the camera positioned between them. While a shield blocks direct light from the LEDs, reflections on the glass sphere are still visible. Figure 6.1 displays the appearance of enabled LEDs in the images, and Table 4.1 presents the modules that incorporate LEDs, including their orientation, color, and wavelength.

4.1 Hardware Components of the Flasher System

The flasher system consists of two main components: the LED and a powerful LED driver. Table 4.2 lists the three different LEDs and drivers used for STRAW-b. A single flasher is the combination of one driver and one LED.

LEDs use semiconductors to emit light, and the color of the emitted light is determined by the band gap of the semiconductor. The higher the band gap, the higher the photon energies and the shorter the wavelength of the emitted light.

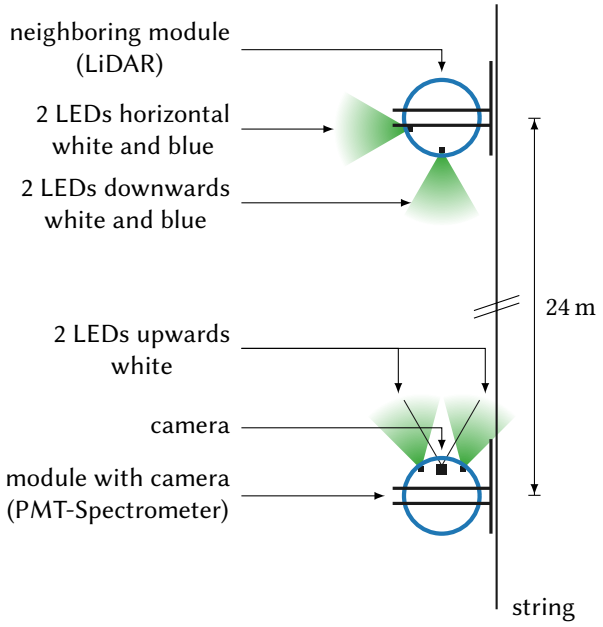

Figure 4.1: Arrangement of the neighboring modules LiDAR and PMT-Spectrometer with the camera and LEDs. The LEDs have white and green colors, and Table 4.1 provides the mapping between position and color.

Table 4.1: Summary of the different LED configurations for all modules in STRAW-b with a LED [49–51].

Module	Address	Pointing	Color	Wavelength
Muontracker &	51	horizontal	white	410 nm–710 nm
LiDARs	52	horizontal	UV	365 nm
	53	downwards	white	410 nm–710 nm
	54	downwards	UV	365 nm
PMT-Spectrometer	51	upwards	white	
	53	upwards	white	410 nm–710 nm
Standard module	51	horizontal	white	410 nm–710 nm
	52	horizontal	royal blue	444 nm

Table 4.2: Summary of the main flasher components [49–52].

Device	Product code	Wavelength
LED driver	LM2759	
LED UV	10-UV-B130-F365-00	365 nm
LED blue	XPEBRY-L1-R250-0R01	444 nm
LED white	XPGDWT-01-0000-00ME1	410 nm–710 nm

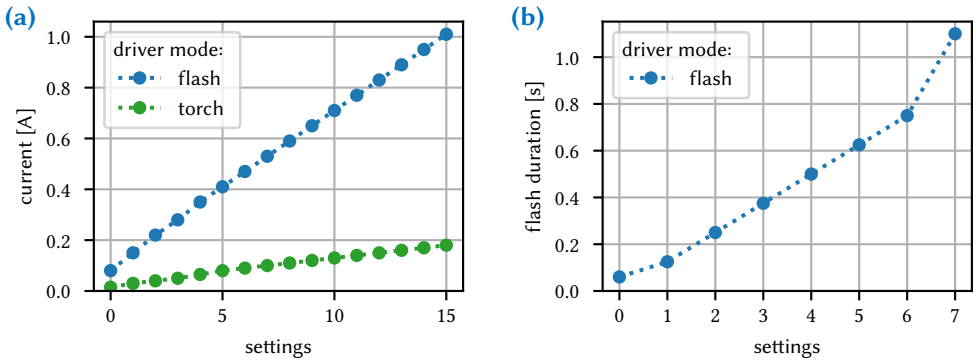

Figure 4.2: Overview of the different LED driver modes. There are two modes, *torch* and *flash*. The *torch* mode provides lower currents (a). Therefore it is not limited in time. In contrast, the *flash* mode limits the maximum duration (b) of the flash to reduce the risk of an overheating LED [52].

The band gap also determines the minimum voltage required for the LED to emit light. The current flowing through the LED increases exponentially with the applied voltage U, as described by the Shockley diode equation:

$$I(U) = I_S \left(e^{\frac{U}{nU_T}} - 1 \right) \tag{4.1}$$

with the reverse bias saturation current I_S, the *ideality factor* n, and the thermal voltage U_T. I_S, n depend on the specific type of the LED, and $U_T = kT/q$ is independent of the Boltzmann constant, electron charge, and temperature of the p–n junction [53]. Therefore, it is crucial to regulate the current when operating an LED. This can be achieved with a simple resistor, or with a more advanced LED driver that provides greater control and power efficiency.

The flasher system uses a configurable LED driver with two different modes,

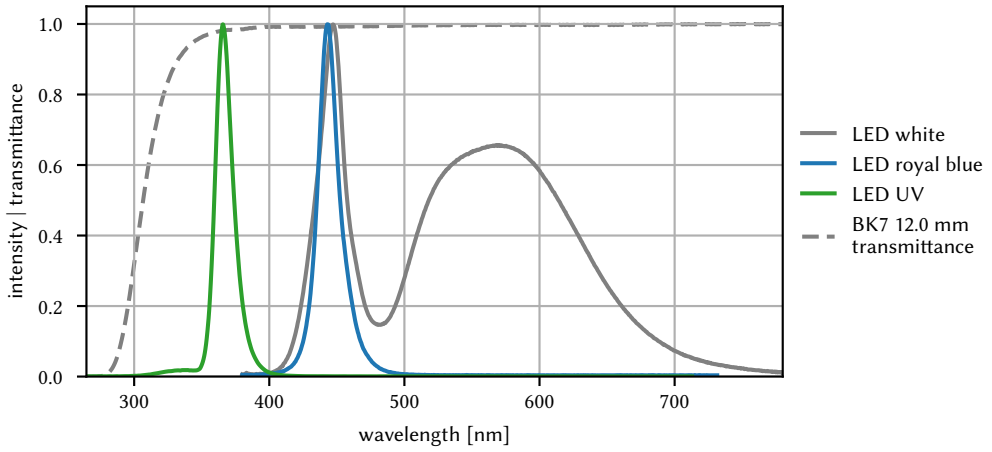

Figure 4.3: Radiation spectra of the three different LEDs used in STRAW-b together with the transmittance of the pressure housing [49–51, 54].

torch and *flash*. As the name indicates, the *flash* mode allows for a configurable maximum duration of up to 1 s and automatically disables itself after that time. In contrast, the *torch* mode turns on the LED without a time limit. To prevent overheating of the LED, which can drastically reduce the lifetime, the *torch* mode has lower current configurations than the *flash* mode.

The configuration mapping to a current and maximum duration is summarized in Figure 4.2. For every communication to the LED driver, the data acquisition (DAQ) stores the mode, configuration parameter, and execution time in the HDF5 (Hierarchical Data Format) file. Additionally, the switch-off signal is recorded as an event by the DAQ if it is sent to the LED driver. However, this is not the case for the *flash* mode, as it automatically disables itself after the configured flash duration [52].

In STRAW-b, the selection of ultraviolet (UV) and white LEDs was based on their ability to cover the range of maximum light transparency of water. These LEDs were strategically placed within the detection range of the PMTs and cameras to ensure optimal detection. In contrast, the blue flashers were not utilized for any measurement purposes.

The spectra of the three types of LEDs used in STRAW-b are illustrated in Figure 4.3. Notably, the spectra from the white LED have a different shape than

those from the UV and blue LEDs. This is because white LEDs utilize a blue LED with multiple layers of light-emitting phosphor in different colors to shift parts of the light to a greater wavelength, forming white light. As a result, the spectra exhibit a peak in the blue from the semiconductor, a broader peak in the yellow to red from the combination of phosphor layers, and a minimum in between.

4.2 Flasher Activations and Measurement Integration

The LEDs can be enabled manually or automatically during a measurement routine that drives the image recording and mini-spectrometer. This routine can communicate with the neighboring module to enable synchronized LED during image exposure. In a typical measurement sequence, the camera captures five pictures every hour with the neighboring module's LED enabled. In one picture, all four LEDs light up simultaneously, while in the other four, each LED lights up individually. More information about the measurement sequence can be found in Section 3.4. To protect the operating PMTs in the PMT-Spectrometer 1 module, only the PMT-Spectrometer 2 module executes this routine, enabling the LEDs in its neighbor, the LiDAR2. However, on October 28, 2021, at 10:09 UTC, the module monitoring system of STRAW-b lost connection to the LiDAR 2, and all subsequent connection attempts have failed. As a result, no images with enabled LEDs have been captured since this date, even though PMT-Spectrometer 2 continued to record images. A summary and discussion of module and sensor malfunctions within the STRAW-b system can be found in Appendix B. Investigations suggest that the malfunctioning modules are booting properly, and the issue lies specifically with the network connection. Furthermore, the the statistic of Ocean Networks Canada (ONC) about connector failures is approximate 10 %, something considered too high for Pacific Ocean Neutrino Experiment (P-ONE). Consequently, penetrators and connectors are not employed within the P-ONE system to mitigate this issue.

The flasher system located next to the cameras has only been enabled manually a few times and has not been integrated into an automated routine yet. Images

of the enabled flasher system in the PMT-Spectrometer 1 and PMT-Spectrometer 2 can be seen in Figure 5.7. Similarly, the LEDs in LiDAR 1 have not been enabled due to the permanently enabled PMTs in the PMT-Spectrometer 1. However, the LEDs have been manually enabled a few times with dim flashes to observe the PMT responses, and a few times with disabled PMTs and brighter LED flashes to track the module position in the camera, as shown in Figure 6.1.

In addition to the scientific LEDs that operate within the visible range of the cameras, certain STRAW-b modules also come equipped with LEDs as part of an art project known as *radioamnion*. As described in Section 2.2, the original objective of the project was to transmit music to the ocean, but the *radiamnion* module lost its connection during the deployment also probably due to a connector failure (Appendix B). As a solution, the *radioamnion* project underwent modifications where four LEDs were incorporated from two different modules, serving as a communication medium. Collaborating artists' compositions are translated into light pulses through an fast fourier transformation (fft) analysis of the audio, where each LED represents a specific frequency spectrum. The flasher's DAQ system also records the events generated by these LEDs.

4.3 Data Acquisition of the Flasher System

Whenever an LED is activated, the DAQ records the time along with the configuration parameters, regardless of whether the LED event is scheduled or manually activated. The LED parameter is stored in the module's general DAQ file with the data product code *SMRD* along with other sensor data like temperature and module orientation. The DAQ system is described in more detail in Section 2.2.3, while Appendix A provides a summary of the file structure.

A manual LED activation is always possible via the command line interface, and the automated measurement routine is robust enough to handle any interference between manual and automatic use. Indeed he DAQ is setup to check for each LED activation if there are PMTs enabled in the same or the neighboring module. If so, it only allows dim flashes to secure the PMTs. If not, also bright configurations are allowed. Furthermore, every LED has a configurable timer disabling the LED automatically. This way, the LED cannot be activated for too

long, threatening to overheat and damage the LED as the heat conduction is limited.

During the commissioning phase, we have observed a period in which the DAQ system failed to track all LED events. Initially, the automated routine enabled all LEDs simultaneously at maximum brightness. At maximum brittleness a single LED consumes 1 A at ~3.3 V leading to ~13 W for all LEDs. Together with the consumption of other components in the module, this exceeds the resistance of the power cable. The voltage drops, leading to a hard shutdown and loss of unsaved data, including the LED configurations. This type of shutdown has the potential to cause severe damage to the instrumentation. Consequently, a procedure was developed to prevent its occurrence during future operations.

While it may seem obvious that exceeding power ratings will result in a power outage, identifying and troubleshooting such issues remotely can be challenging. For instance, the log file of the control software may contain abrupt process crashes due to the critical part not being saved during unexpected shutdowns. To address this issue, logging more system parameters with the module monitoring system, such as the start-up time, can help identify and solve the problem. It is also worth noting that a hard shutdown introduces the risk of corrupting the system, which can lead to further complications. In fact, in the case of the Wavelength-shifting Optical Module (WOM) module, it may not power up after an unexpected shutdown. Therefore, it is crucial to take measures to prevent and mitigate unexpected shutdowns to ensure the system's reliability and integrity.

5 Camera Optics

Capturing images is a fundamental aspect of photography and computer vision. When a camera takes a picture, it projects the 3D scene onto a 2D image plane, resulting in a loss of depth information. This process is known as a projective transformation, which maps the direction space onto a two-dimensional Euclidean space. The camera's field of view (FoV) is defined by the objective lens and image sensor. Fish-eye cameras are a type of camera with a wide FoV, and they can capture a wider scene compared to regular cameras. In this section, we will provide an overview of the basic principles of camera optics, with a particular focus on Fish-eye cameras. Since all of Strings for Absorption Length in Water b (STRAW-b)'s cameras belong to this category

5.1 Basics about Camera Projection

One of the most intuitive ways to represent a camera's projection is through the pin-hole projection model. In this model, the 3D scene is projected onto a 2D image plane through a pin-hole or aperture. The projected image, also known as the real image, is a representation of the object being captured by the camera. Figure 5.1 illustrates the pin-hole projection of an object $\overline{AB}$ and its corresponding real image $\overline{A'B'}$. In the following, primed variables denote variables in the projection space. The mapping function for the pin-hole projection is given by

$$r' = f \tan(\theta) \tag{5.1}$$

where f is the distance between the principal pin-hole and the image plane, θ is the angle of incidence of the projected ray to the camera's optical axis, and

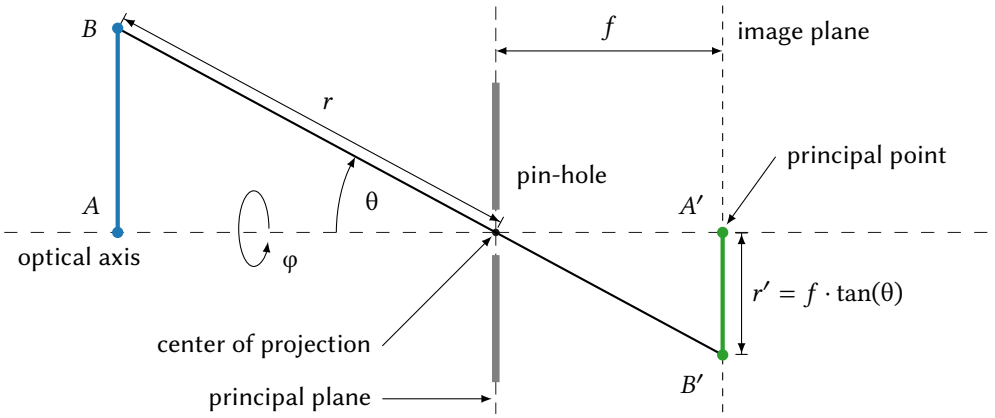

Figure 5.1: Pin-hole projection of an object $\overline{AB}$. The pin-hole projection is rectilinear as it preserves the rectilinearity of the projected scene, i.e., straight lines in the scene are projected as straight lines on the image plane [55, 56].

r' is the projected radial distance from the optical axis. The pin-hole projection is rectilinear, meaning it preserves the rectilinearity of the projected scene. This property is ideal in most applications as it ensures that straight lines in the scene are projected onto straight lines on the image plane [55, 56].

The projection is symmetric around the optical axis, and introducing the rotation angle φ around the optical axis adds the missing dimension for the 3D scene and 2D image plane. Therefore, φ is not affected by the projection, or mathematically $\varphi = \varphi'$. All light arriving at the pin-hole from a direction defined by φ and θ projects to a point

$$\vec{P'}(\varphi, \theta) = \begin{pmatrix} x' \\ y' \end{pmatrix} = r' \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix} = f \tan(\theta) \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix} \tag{5.2}$$

on the image plane with the coordinates x' and y'

Inverting the pin-hole projection function gives the mapping from an image point to its corresponding direction (φ, θ) written as a unit vector

$$\vec{P}\left(\vec{P'}\left(\varphi, r'\right)\right) = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos(\varphi)\sin(\theta) \\ \sin(\varphi)\sin(\theta) \\ \cos(\theta) \end{pmatrix} \quad \text{with: } \tan(\theta) = \frac{r'}{f}. \tag{5.3}$$

Here, $\overrightarrow{P'}(\varphi, r')$ represents the projected point on the image plane, $\overrightarrow{P}(x, y, z)$ represents the direction vector in the 3D scene, and $r' = |\overrightarrow{P'}|$ is the distance to the principal point. The maximum dimensions of the image sensor and f define the FoV for the combination of lens and image sensor [55, 56].

The pin-hole projection is a very simplified description of an objective transformation to illustrate the basic principle. In reality, objective lenses have various types of distortions that require more advanced descriptions to accurately reflect the optics. For example, when characterizing a wide FoV camera using the pin-hole projection, the size of the projected image becomes huge, increasing to infinity at a FoV of 180° in total or 90° from the optical axis. This consideration leads to considerable radial distortion that a fish-eye lens has to introduce, including its description. Furthermore, chromatic aberration or other effects should be considered depending on the scope and level of precision required. [55, 56].

There are several different fish-eye projections. Typically, fish-eye lens manufacturers design a lens so that its distortion follows a projection function. Projections based on spherical geometry with slightly different image representations are named equidistant, equisolid, orthographic, and stereographic. The orthographic projection is limited to a maximum angle of 90° from the optical axis and is therefore not commonly used in fish-eye designs. Other methods, such as polynomial or logarithmic descriptions, have also been proposed. Deviations from the lens design induced by tolerances in the manufacturing process can be modeled with polynomial elements where the model is the first-order parameter.[55, 56].

A digital camera captures the 2D image with the pixels of the image sensor. Each pixel represents an area on the image plane with the center of the area

$$\overrightarrow{P'_{ij}} = \begin{pmatrix} x'_{ij} \\ y'_{ij} \end{pmatrix} = r'_{ij} \begin{pmatrix} \cos(\varphi_{ij}) \\ \sin(\varphi_{ij}) \end{pmatrix} \tag{5.4}$$

and ij are the row and column indices of the pixel grid. Usually, sensors have rectangular pixels all with the same dimensions of $\Delta x'$ by $\Delta y'$, and hence the

positions are

$$\vec{P'_{ij}} = \begin{pmatrix} \Delta x' & 0 \\ 0 & \Delta y' \end{pmatrix} \begin{pmatrix} i \\ j \end{pmatrix} - \vec{c'} = \begin{pmatrix} i \cdot \Delta x' \\ j \cdot \Delta y' \end{pmatrix} - \vec{c'} \quad \text{with: } i, j \in \mathbb{N} \tag{5.5}$$

where $\vec{c'}$ is the position of the principal point on the image sensor. For most sensors, pixels are squared with $\Delta x' = \Delta y'$ and Equation 5.5 can be simplified accordingly.

To accurately model a camera's distortion, it is necessary to know the intrinsic parameters, such as the position of the camera's principal point $\vec{c'}$. Several techniques can be used to obtain these parameters, but one efficient method involves using a chessboard pattern printed on a flat surface and capturing multiple pictures from different camera positions. This method allows for the extraction of both the intrinsic and extrinsic parameters of the camera, including the camera's rotation and translation to 3D camera coordinates, without requiring additional sensors to track the camera's position and rotation.

There are several approaches to this technique, including methods developed by Zhang, Tsai, and Selby [57–59]. Additionally, software implementations, such as OpenCV [48], have been developed to simplify the calibration process, including chessboard recognition. OpenCV is a well-known open-source software implementation that is compatible with multiple languages, including C++ and Python [48].

5.2 Projection of the IceCube Camera

The level of detail provided about individual components can vary depending on the lens and camera used. In cases where parameters are missing or further calibration is necessary, lab measurements can be used, as discussed in Chapter 5. For the deployed lens type, information from the manufacturer is fundamental and is summarized in Table 5.1.

During the detector development process, there was limited time available and a lack of robust readout software for the camera. As a result, a set of pictures with

Table 5.1: Public characteristics of the Arducam M25170H12 lens with M12 mount. The resulting FoV (field of view) depends on the image sensor dimension and is given for image sensors with a 1/2.5" and 1/4" diameter. With the 1/3" sensor used in STRAW-b the FoV is in between. The effective focal length (EFL) is the distance from a principal plane of a lens to its image plane [60].

	FoV for 1/2.5" sensor			FoV for 1/4" sensor
Effective focal length	diagonal	horizontal	vertical	horizontal
1.7 mm	180°	180°	140°	118°

a chessboard pattern, as suggested in the previous section, was not recorded. However, the principal point, or lens center, in pixel coordinates of the image sensor was measured for both PMT-Spectrometer cameras. For the combination of sensor size and lens used, the lens center can be extracted from every well-exposed image, as the lens exceeds its optical limits for high angles, resulting in black areas on the image. The boundary of the black area on the image forms a circle, and the center of the circle represents the center of the lens. Figure 5.2 illustrates the method used to determine the principal point for both modules.

The equisolid projection has been found to "describe the [lens] distortion nearly perfectly" by the IceCube Camera team, with only "slight deviations at the edges" (G. Roellinghoff, personal communication, April 20, 2022). As shown in Figure 5.3, the equisolid projection maps the distance between the principal point and a point on a virtual sphere with radius f to the projected distance from the principal point on the image plane, $r' = |\vec{P'}|$. The equisolid distortion function can be expressed as

$$r' = 2f \sin\left(\frac{\theta}{2}\right) \tag{5.6}$$

where θ is the incident angle of the projected ray to the optical axis. The inverse function is given by

$$\theta = 2 \arcsin\left(\frac{r'}{2f}\right). \tag{5.7}$$

As with the pin-hole projection, there is no distortion of ϕ. Following Equa-

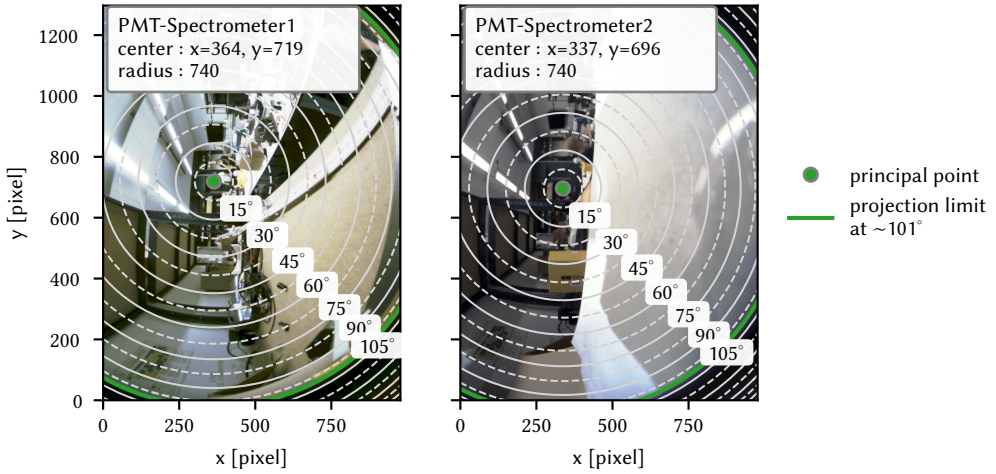

Figure 5.2: Pictures to extract the principal point, i.e., the lens center, on the image sensor. With the combination of lens and image sensor, the image sensor covers areas where the lens is no longer projecting the scene. This results in the black areas on the images. Fitting a circle to the boundaries yields the center of the lens. Based on the equisolid projection, Equation 5.9 yields the incident angle θ of a pixel, which the images illustrate. For comparison, both images show the original orientation and are not rotated according to the scene orientation.

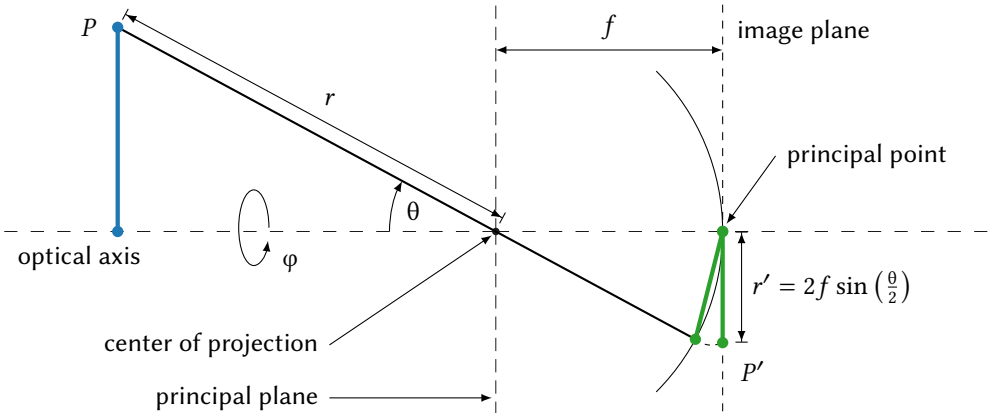

Figure 5.3: Equisolid projection of an point $\vec{P}$ to its representation on the image plane $\vec{P'}$. Equisolid projection does not preserve the rectilinearity of the pin-hole projection. As the name indicates, it is an equal-area projection as the ratio of an incident solid angle and its resulting area in an image is constant [55, 56].

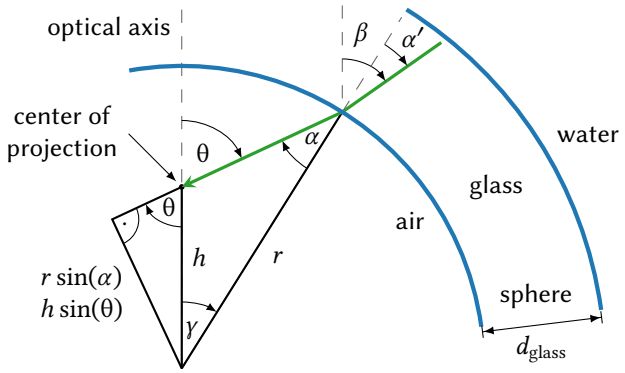

Figure 5.4: Projection of a sphere induced by refraction and where the center of projection is off-centered in the sphere. Like a module with spherical pressure housing, the inner volume is air, and the sphere glass. However, the drawing does not include the refraction from glass to water at the outer boundary of the glass sphere for simplification, and the projection can recursively model it as shown in Section 5.3. Furthermore, the relations of the dimensions do not represent the module with h =~63 mm, r =153.1 mm, and d_{glass} =12.0 mm.

tion 5.2, a ray with an incident direction φ and θ projects to the point on the image plane

$$\overrightarrow{P'}(\varphi,\theta) = 2f \sin\left(\frac{\theta}{2}\right)\begin{pmatrix}\cos(\varphi)\\ \sin(\varphi)\end{pmatrix} \tag{5.8}$$

and the pixel coordinates on the image sensor are

$$\overrightarrow{\hat{I}'}(\varphi,\theta) = \begin{pmatrix}\hat{i}\\ \hat{j}\end{pmatrix} = 2f \sin\left(\frac{\theta}{2}\right)\begin{pmatrix}\frac{1}{\Delta x'} & 0\\ 0 & \frac{1}{\Delta y'}\end{pmatrix}\begin{pmatrix}\cos(\varphi)\\ \sin(\varphi)\end{pmatrix} + \overrightarrow{c'} \tag{5.9}$$

where we have used the inverse of Equation 5.5. The pixel indexes i and j of the image sensor or photo can be obtained by rounding $\hat{i}$ and $\hat{j}$ to the nearest integers.

5.3 Projection of the Spherical Module

The previous Section 5.2 provides a simplified projection to model the camera. This model does not include the refraction induced by the spherical glass pressure housing and the oceanic water, which should be introduced in this sub-

section and can be expressed as an additional distortion. The following derivation presents the spherical refraction analytically in association with the camera model for simplified adoption in other cases.

The pressure housing made of spherical glass creates two spherical surfaces: one facing the water outside and the other facing the air inside the module. The geometry of the refraction at the inner spherical surface is illustrated in Figure 5.4, and the refraction at the outer spherical surface can be applied recursively, as explained later in the text. The camera is positioned off-center inside the glass sphere, at a distance of approximately 63 mm from the center. The inner radius of the glass sphere is 153 mm, and its thickness is 12 mm. Similar to the model of the camera in Figure 5.3, the incident angle from the optical axis is denoted by θ, while the corresponding incident angle after the refraction is denoted by β. The geometry is given by

$$r \sin(\alpha) = h \sin(\theta) \tag{5.10}$$

where α is the incident angle of the refraction at the spherical interface. The refection happens at an angle γ from the optical axis with $\gamma = \theta - \alpha$. Snell's law gives the refracted angle α' with

$$n_{\text{air}} \sin(\alpha) = n_{\text{glass}} \sin(\alpha') \tag{5.11}$$

and the refraction indices. Combining 5.11 with 5.10 yields the dependency between α' and θ

$$h \sin(\theta) = r \frac{n_{\text{glass}}}{n_{\text{air}}} \sin(\alpha'). \tag{5.12}$$

Finally, the incident angle of the sphere β is

$$\begin{aligned} \beta &= \gamma + \alpha' = \theta - \alpha + \alpha' && (5.13) \\ &= \theta - \sin^{-1}\left(\frac{h}{r}\sin(\theta)\right) + \sin^{-1}\left(\frac{n_{\text{air}}}{n_{\text{glass}}}\frac{h}{r}\sin(\theta)\right). && (5.14) \end{aligned}$$

At this point, the second refraction at the outer interface of the sphere has not yet been integrated. However, the geometry explained so far can be used recursively to introduce it. For clarity, all values in the recursive step are labeled with

r, i.e., θ_r. In the recursive step, the refracted angle α' from the inner spherical refraction equals

$$r \sin(\alpha') = h_r \sin(\theta_r) = r_r \sin(\alpha_r) = r_r \frac{n_{\text{water}}}{n_{\text{glass}}} \sin(\alpha'_r) \tag{5.15}$$

where the dimensions adopt to $r_r = r + d_{\text{glass}}$ and $h_r = r$. Hence, the final incident angle after both refractions is given by:

$$\begin{aligned} \beta &= \gamma + \gamma_r + \alpha'_r && (5.16)\\ &= \theta - \alpha + \theta_r - \alpha_r + \alpha'_r && (5.17)\\ &= \theta - \alpha + \alpha' - \alpha_r + \alpha'_r. && (5.18) \end{aligned}$$

This can be written out in full as:

$$\begin{aligned} \beta = \theta &- \sin^{-1}\left(\frac{h}{r}\sin(\theta)\right) + \sin^{-1}\left(\frac{n_{\text{air}}}{n_{\text{glass}}}\frac{h}{r}\sin(\theta)\right) \\ &- \sin^{-1}\left(\frac{n_{\text{air}}}{n_{\text{glass}}}\frac{h}{r + d_{\text{glass}}}\sin(\theta)\right) + \sin^{-1}\left(\frac{n_{\text{air}}}{n_{\text{water}}}\frac{h}{r + d_{\text{glass}}}\sin(\theta)\right) \end{aligned} \tag{5.19}$$

without the same ordering of the single terms of Equation 5.18. The refraction indices are $n_{\text{air}} \approx 1$, $n_{\text{glass}} \approx 1.52$ for BK-7 glass, and $n_{\text{water}} \approx 1.35$ for wavelength around 500 nm [61]. As mentioned before, the dimensions are given by h=~63 mm, r=153 mm, and d_{glass}=12 mm. Hence the individual angles are in the order $\alpha \leq \alpha' \leq \alpha'_r \leq \alpha_r$ which leads to $\beta \leq \theta$ if $0 \leq \theta \leq \pi$. Consequently, the sphere acts as an optical enlargement for the camera because of its position in the spherical module. Due to the symmetry, φ is not affected by the glass. Figure 5.5 shows the different models and the effect of the spherical distortion with the parameters of the cameras in the PMT-Spectrometer modules.

In this section, β and θ were introduced as the incident angles, which measure against the optical axis outside the sphere and at the camera, respectively. For simplification, the angles will be referred to as θ from now on as both angles represent θ typically used in spherical coordinates, and the context will indicate to which it refers.

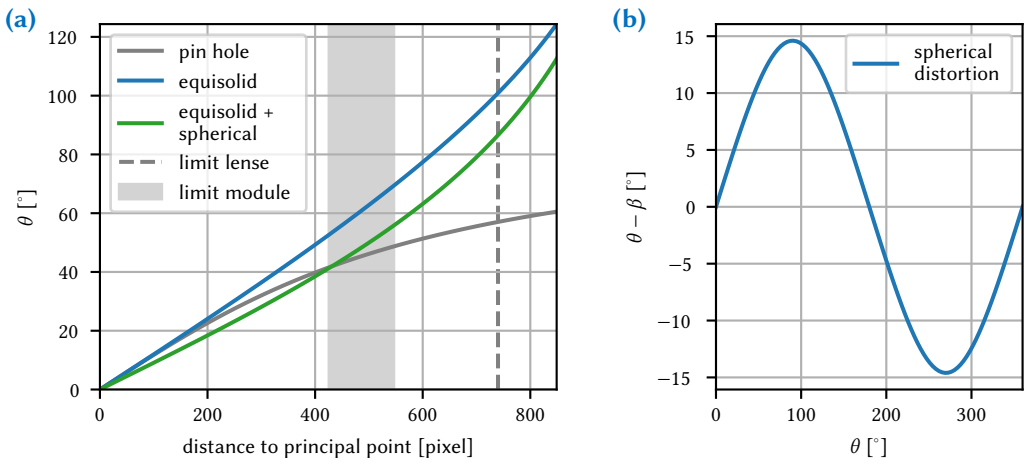

Figure 5.5: A comparison of camera models (a) and the distortion induced by a sphere (b) as the difference of the inner (θ) and outer (β) incident angle. All curves rest upon the camera's parameters within the PMT-Spectrometer modules. The sections 5.1, 5.2, and 5.3 recap the different camera models, pin-hole, equisolid distortion, and spherical distortion, respectively. Furthermore, the FoV limits are indicated in (a) for the lens with 740 pixels and when the camera is integrated into the module (see Section 5.4).

5.4 Field of View of the Deployed Camera

The camera's native FoV or AOV is determined by the focal length of the lens and image sensor size. A shorter focal length or larger sensor dimensions result in a wider FoV. The *35 mm equivalent focal length* parameter provided for cameras or lenses specifies the focal length for a 35 mm sensor with an equivalent FoV.

However, obstacles can limit the FoV. In the PMT-Spectrometer, the camera's placement is surrounded by 12 units of PMTs, lens, and filters, resulting in limited space.Given the space constraints of the pressure housing, the camera had to be placed at a location that does not provide the full native FoV. As a result, the camera's AOV is restricted to an area around the principal point, with the holding structure for the 12 PMT units becoming visible at higher inclination angles. Figure 5.6 shows a sectional view including the dimensions of the limited FoV.

In contrast to the disadvantage of limiting the FoV, the mounting provides dimensions to probe the camera's parameters and if the equisolid projection fits

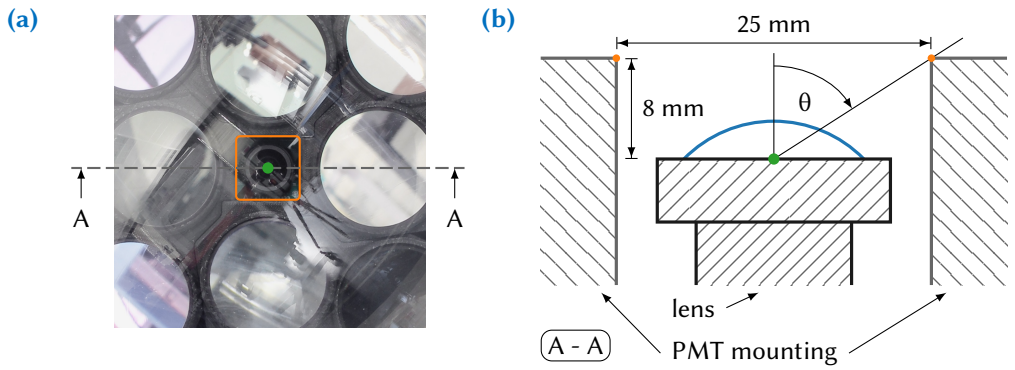

Figure 5.6: Top view (a) and sectional view (b) of the lens position and the photomultiplier tube (PMT) mounting structure. The horizontal line in the top view indicates the cutting edge of the sectional view. The organ line highlights the edge of the PMT mounting, $\theta \approx 57^\circ$ is an estimate for the maximum angle of view (AOV) in the plan of the cross-section, as the center of projection (green point) cannot be measured precisely at the lens.

the lens. The deployed cameras captured images with the two LEDs next to the camera enabled, revealing the mounting boundary on the image. Figure 5.7 shows the pictures from which the mounting boundary on the image is extracted. On one side, the boundary is not visible, so three distances from the principal point yield the mean distance of (441 ± 8) Pixel and (444 ± 8) pixel for the cameras in the PMT-Spectrometer1 and PMT-Spectrometer2, respectively. This results in an angle of $\theta \approx (55 \pm 1)^\circ$ for both modules.

The cross-section of Figure 5.6 represents the same geometry of the measured pixel distance, which is parallel to the y-axis or x-axis in the image and crossing the principal point. The geometry with 8 mm and 25 mm results in $\theta \approx 57^\circ$ from the optical axis. However, as the principle point of the lens is unknown with reference to its geometry, the 8 mm value is an estimate and not a proper measurement. These angles based on the images and camera model are in good agreement with the geometrically calculated angle.

Considering the spherical distortion introduced in Section 5.3, the corresponding incident angle of the FoV limitation outside the module reduces to $\theta \approx (43 \pm 1)^\circ$. However, this is not more than an assessment of the camera model if it fits in a specific domain the reality, and measurements in the lab are necessary for an

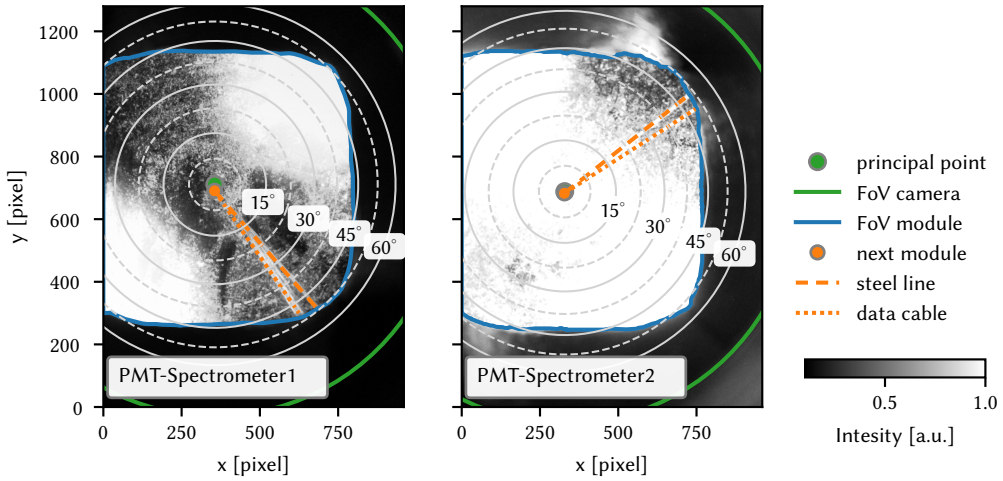

Figure 5.7: Pictures of the deployed modules with enabled LEDs to make the boundary of the PMT mounting structure visible. The LEDs point in the same direction as the camera. Here they flashed with the maximum brightness to increase the contrast of the boundary. Both modules clearly show sedimentation. Also, the data cable and steel cable is visible, and the positions, as marked, are extracted for further analysis. The position of the neighboring module is extracted from other images where their LEDs are enabled.

improved model.

6 Image Processing

Since the cameras started capturing images in late April 2021, the two functioning cameras have taken hundreds of thousands of photos each. As of December 2022, the cameras have captured a total of one million images. More information on the cameras' performance and the number of images captured can be found in Section 3.5.

The cameras are sensitive to two types of light sources: artificial light from LEDs and bioluminescence emitted by animals or microorganisms. The camera can capture the light emitted by LEDs mounted in the neighboring module, as well as the light from the LEDs located next to the camera that make up the flasher system. Figure 6.1 depicts how both types of LEDs appear in the camera's view. For most images, the neighboring modules' LEDs are disabled, and the flash system got activated just a few times. Bioluminescence, visible in a few images each day, is the other light source of interest. In some cases, the remotely operated underwater vehicle (ROV) could also emit artificial light, such as during an inspection of the string. However, no inspection was carried out while the cameras were in operation.

The aim of this chapter is to extract pictures with bioluminescence from the majority of completely dark pictures. Additionally, the technique should be able to measure the position and other parameters such as brightness, enabling a higher-level analysis of the recorded bioluminescence. Figure 6.2 provides a collage of bright images selected based on the parametrization. Speed and robustness are also critical parameters for processing the large number of pictures. Since the pictures have a bit depth of 16 $\frac{\text{bit}}{\text{pixel}}$, the 1.2 megapixels result in a picture size of 2.4 MB, and the nearly 1 million images recorded so far take up around

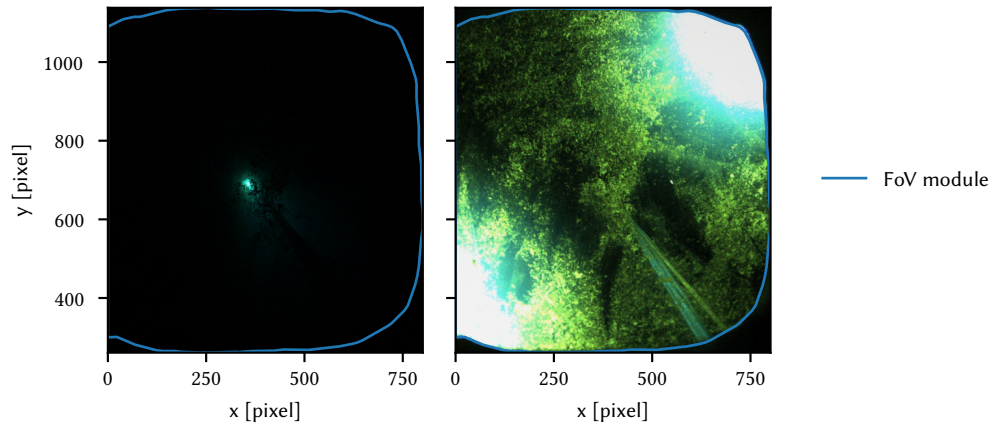

Figure 6.1: Images with enabled LEDs on the neighboring module (left) and the flasher LEDs next to the camera (right). The camera is located in the PMT-Spectrometer 1 module and its neighboring module, LiDAR 1. The biofouling, data cable, and steel line are visible with enabled flasher LEDs. The images are cropped to the module field of view (FoV), and the features on the image are labeled in Figure 5.6.

2.4 TB of storage.

In contrast, extracting pictures with enabled LED does not require special treatment. Every time a LED gets activated, the data acquisition (DAQ) stores the time together with the configuration parameters, regardless of whether it was scheduled or manually triggered. Therefore, pictures with enabled LED can be identified by comparing the image capture times. During commissioning when all the LED have been simultaneously powered with maximal brightness caused module shutdowns, leading to data losses in the DAQ, as described in Section 4.3. The issue was detected and resolved on May 18th, 2021, at around 11:00 UTC. As a result, the period with inconsistent data needs special treatment or can be ignored.

6.1 Pixel Noise

A substantial quantity of the images is the individual noise for each pixel. This noise comes from various sources, including components such as the analog to digital converter (ADC), DC-to-DC converters, and the pixel itself. Typically,

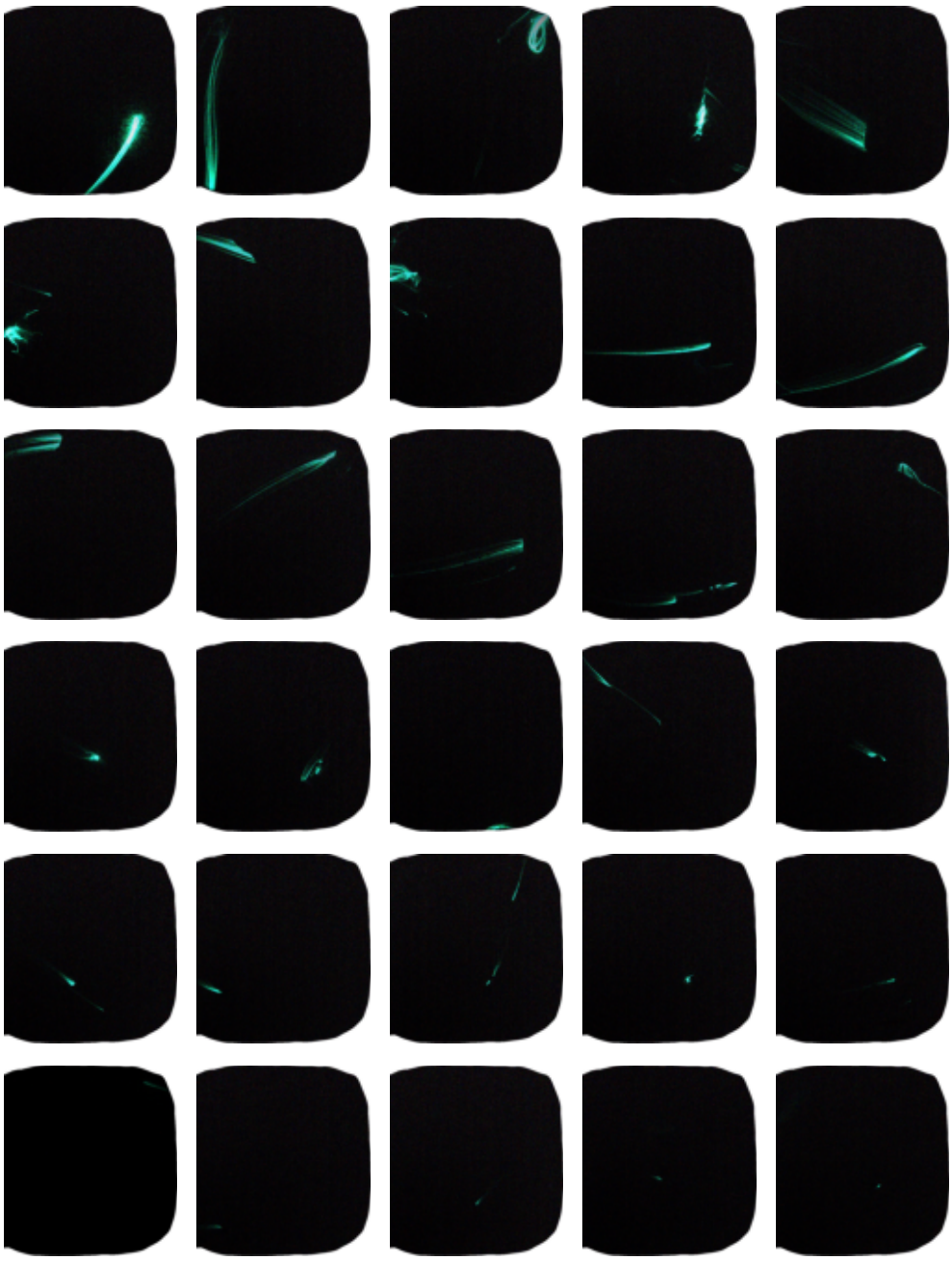

Figure 6.2: A collage showcasing a selection of bioluminescent images captured by the camera in the PMT-Spectrometer 1 module. The images represent a selection of the 1000 brightest pictures recorded by the camera, as determined by the picture parametrization outlined in this chapter. The brightness of a picture is determined by the sum of intensities across all pixels deemed active by the algorithm. All images have been reduced to the field of view (FOV) specified in Section 5.4.

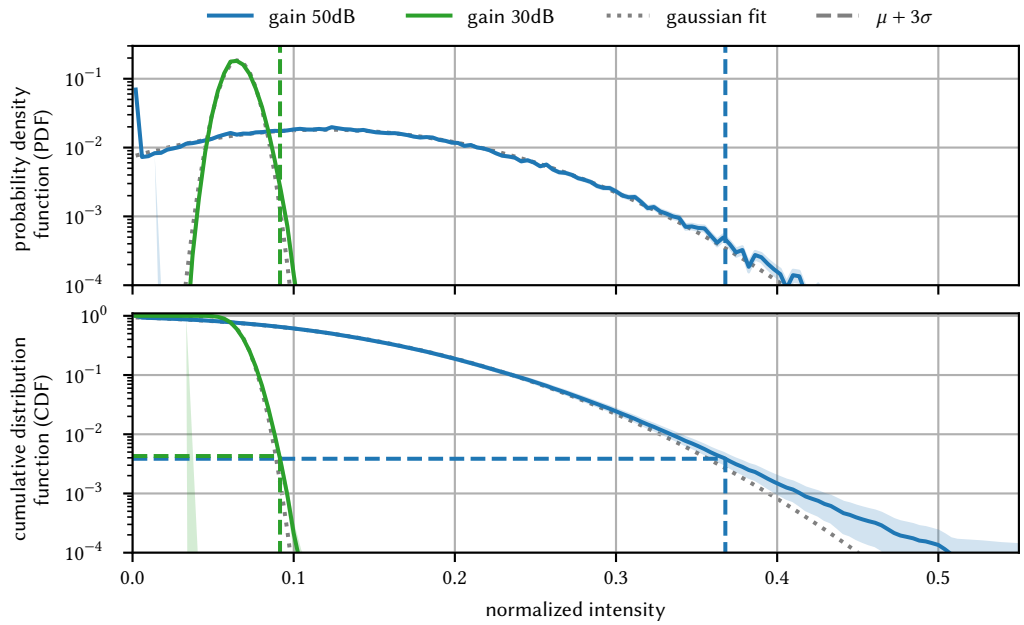

Figure 6.3: Intensity distribution of a single pixel under two different camera gain settings. This figure displays the distribution of intensity values recorded from a single pixel over all images captured until the 31st of October, 2022. The camera gain of 50 dB includes digital amplification, while the gain of 30 dB represents the maximum analog amplification without digital amplification. Similar distributions can be observed for other pixels with a shifted mean (μ) and varying standard deviation (σ). The distributions follow a Gaussian shape, as indicated by the fit. Additionally, the intensity threshold of $\mu + 3\sigma$ is marked, which was used in the cluster detection algorithm described in section 6.2.1.

lower operating temperatures decrease noise, and some cameras have active cooling. However, the cameras in the modules heat up to around ~10 °C when continuously operating due to the pressure housing and the module's internal structure, even though the surrounding Pacific water is ~3 °C.

In the lab, dark noise measurements showed that the noise has a linear dependency on the exposure time and an exponential dependency on the temperature. A noise model was developed to represent the dark noise N_i of each pixel i, given by the equation:

$$N_i(\Delta t, T) = a_{0i} + \left(a_{1i} + a_{2i} \cdot e^{a_{3i} \cdot T}\right) \Delta t. \tag{6.1}$$

Here, Δt is the exposure time, T is the sensor temperature, and a_{ni} are individual

parameters for each pixel. As the camera operates at a constant temperature and with a fixed exposure time, the noise model can be simplified to a single value for each pixel representing the expected dark noise. Moreover, most images are completely dark without bioluminescence or artificial light from the LEDs. By calculating the average value for each pixel across a large number of images, it is possible to obtain an accurate measurement of the dark noise μ_i of individual pixels without the need to remove bioluminescence or LED occurrences.

To illustrate the distribution of the noise, Figure 6.3 shows the intensity distribution of one pixel over all the recorded images until October 31, 2022. The camera operated with a gain of 50 dB initially, which was later changed to 30 dB to disable digital amplification [45]. The distribution is split into the two different gain periods. With a gain of 50 dB, the dark noise spreads up to 40 % of the pixel's intensity range, while with 30 dB, the noise is below 10 %, yielding a broader range for detecting light.

There is significant variation in the dark noise of pixels throughout the entire image, and a recurring pattern can be observed in each image, especially when using a gain of 50 dB. This pattern noise is characterized by an increase towards the upper and left edges of the image in all cameras used, as depicted in Figure 6.4 for the camera utilized in the PMT-Spectrometer 1 module. In some cases, the noise even saturates certain pixels, although these pixels are mainly located outside the module's field of view. By contrast, a gain setting of 30 dB dramatically reduces the pattern noise, and no pixels are saturated solely from noise. However, some pixels may still be broken and reach maximum intensity, referred to as hot pixels.

Overall, The misconfiguration of the camera's digital gain leads to the generation of significant unwanted noise, resulting in the presence of digital artifacts in the intensity distribution shown in Figure 6.4. Nonetheless, the dominant noise component is pattern noise, which can be effectively mitigated by measuring the mean dark noise μ_i and its standard deviation σ_i for each pixel. Since the majority of recorded images are dark and do not contain measurable light sources, a sufficiently large set of images can be used to compute μ_i and σ_i without the need to exclude rare photos with light sources beforehand. As depicted

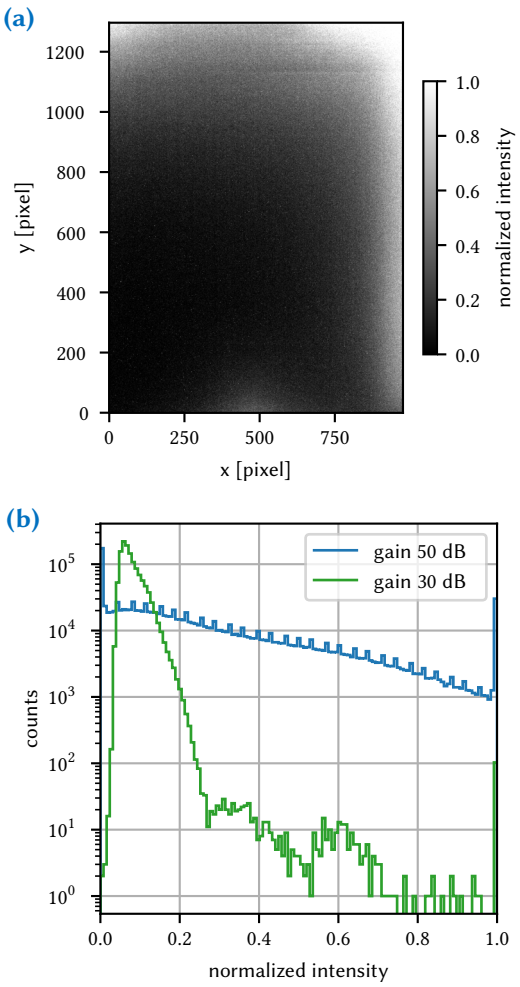

Figure 6.4: *Pattern noise* (a) and pixel dark current distributions (b). *Pattern noise* is the noise that is independent of illumination conditions. Here it is shown for a 50 dB gain setting. The image and histogram were captured during dark scenes with an exposure time of 63 s. For more information on the gain configurations, see Section 3.1.2, which provides a summary of the gain 50 dB with digital amplification and the reduction to maximum analog amplification only, resulting in a gain of 30 dB.

in Figure 6.3, the noise can be modeled with a Gaussian distribution for both gain settings. Thus, a threshold of $I_{th,i} = \mu_i + n_{th} \cdot \sigma_i$, where n_{th} is a constant factor, can be used to filter out noise-induced intensities with similar probabilities, irrespective of the gain setting.

6.2 Cluster Detection Algorithm

As previously discussed, one of the major challenges in the realm of image identification is accurately distinguishing and isolating images that capture even the slightest amount of light from the vast number of completely dark images in the Pacific Ocean's dark environment. When the camera detects a light-emitting object, the pixels receiving the light emit more intensity than the background noise level. As a result, the object appears as a cluster of neighboring pixels with increased intensity. Moreover, the optics' distortion or non-ideal focus may cause objects to expand on the image plane, even if the ideal optical transformation predicts minor expansion. To filter out relevant images, a cluster detection technique is developed, where a cluster is defined as a group of interconnected pixels within a certain proximity of each other.

The first step to detect the cluster is to mark active pixels in an image, which are pixels that measure intensity above a certain threshold. Due to the pixel noise described in the previous Section 6.1, the intensity threshold is set by a factor n_{th} using the equation:

$$I_{th,i} = \mu_i + n_{th} \cdot \sigma_i. \tag{6.2}$$

Here, μ_i and σ_i represent the mean noise level and standard deviation of pixel i, respectively.

The second step of the cluster detection algorithm involves identifying clusters of active pixels within the image and assigning each pixel to its corresponding cluster. A cluster is considered formed when the distance between any two pixels within it does not exceed a specified maximum distance. Figure 6.5 provides an illustration of forming clusters, where active pixels must share at least one corner to be considered part of the same cluster (this corresponds to the later-introduced neighboring order 1).

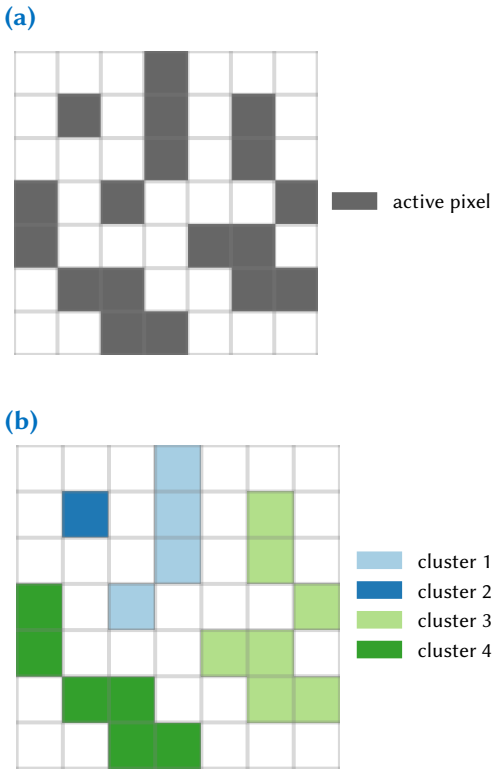

Figure 6.5: Illustration of the cluster detection algorithm and labeling process. In the image, pixels with an intensity greater than $I_{th,i}$ are identified as active pixels (a). In the next step (b)., the clusters of neighboring active pixels are detected and assigned labels starting from 1. All inactive pixels are assigned as label 0.

When determining the distance between pixels, it is important to consider the metric used by the image sensor. Since the image sensor has a Bayer Filter Mosaic (Section 3.2), red and blue pixels are separated by at least two pixels, while green pixels are diagonally adjacent. Consequently, the distance definition between two pixels should incorporate the color mosaic, allowing a cluster to form even if it appears in only one color channel. One possible approach to defining the distance is by considering the nth shell around one pixel in which the other pixel is located, as depicted in Figure 6.6. These shells are referred to as neighboring order or simply order in the subsequent discussion. To perform independent clustering for all colors, an even order is necessary due to the arrangement of red and blue color filters.

Thus, cluster detection involves two parameters: the intensity threshold and the cluster distance, represented by $n_{th} \cdot \sigma$ and the neighboring order, respectively. The ideal cluster detection should be sensitive to weak signals and robust against pixel noise. To achieve this, the threshold and neighboring order must be set to minimize clusters caused by noise without excluding faint events. Lowering the threshold and increasing the neighboring order results in more clusters from noise.

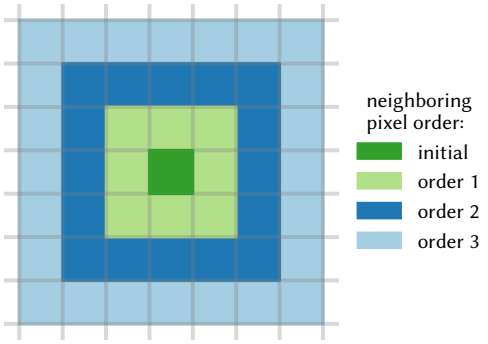

Figure 6.6: Definition of the neighboring pixel order from the initial centered pixel.

6.2.1 Simulation-Based Evaluation of Cluster Detection Parameters

The cluster detection method discussed in Section 6.2 relies on two parameters: the intensity threshold and the neighboring order. Both of these parameters can have a significant impact on the number of fake clusters that are detected purely due to noise. However, by default, the cluster detection method is unable to differentiate between clusters that are the result of an actual signal and those

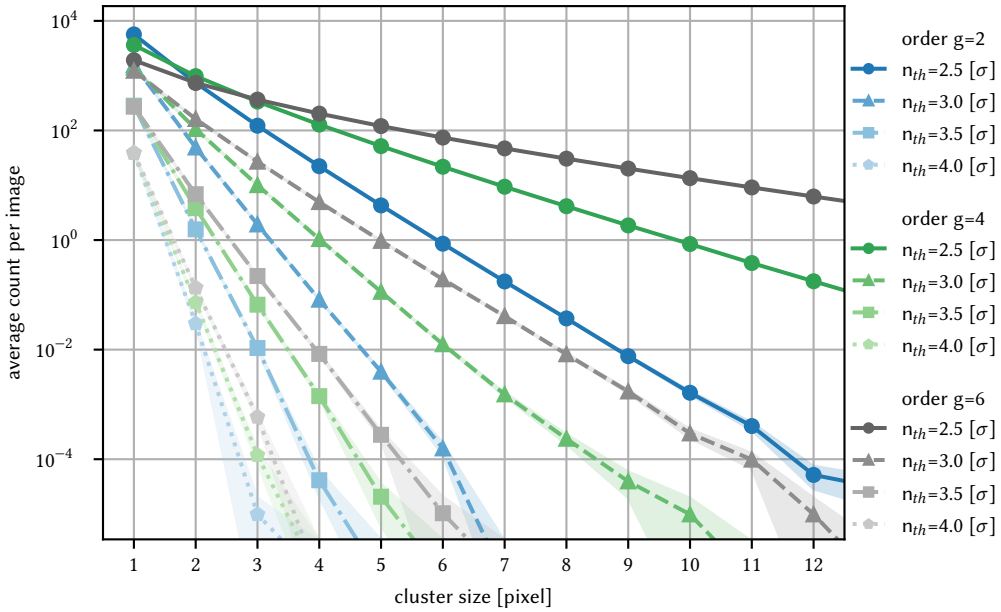

Figure 6.7: Impact of different threshold and neighboring order configurations on random clustering in the cluster detection algorithm. The results highlight that increasing the neighboring order and lowering the threshold lead to an increase in the number of clusters generated by noise. In particular, low thresholds combined with a high neighboring order can result in clusters covering large portions of the image. The crossing lines in the plot indicate that many smaller clusters have been merged into larger clusters with sizes that surpass the x-axis scale.

that are purely noise-induced. This can lead to the detection of fake clusters that are merely statistical fluctuations randomly forming a cluster.

To quantify the impact of noise-induced clusters on the detection process, a Monte Carlo simulation was performed on images that were generated with Gaussian noise that was independent for each pixel. The simulation results, depicted in Figure 6.7, show the average expected number of clusters per image for different combinations of the threshold and neighboring order parameters.

The expected number of clusters per image can be modeled using an effective cluster size n_C and an adopted Bernoulli distribution [62]. The effective cluster size describes the number of pixels that are related to one cluster, including the surrounding inactive pixels. A cluster with only one active pixel requires

$(2g+1)^2 - 1$ inactive pixels, where g is the neighboring order. Therefore, the effective cluster size is $n_C = (2g+1)^2$, including the one active pixel in the center. Clusters with x active pixels require a shape-related effective cluster size. The average effective cluster size, including the image resolution N, can be expressed as:

$$n_C(x, g, n_{th}, N) = (2g+1)^2 \cdot x^{a_0(g,n_{th},N)} \cdot a_1(g, n_{th}, N) \tag{6.3}$$

where $a_i(g, n_{th}, N)$ are variables that depend on n_{th}, g, and N. For simplicity, the variables x, n_{th}, g, and N are denoted as $\vec{X}$ in the following. On average, one image contains

$$C_0(\vec{X}) = \frac{N}{n_C(\vec{X})} \tag{6.4}$$

clusters, and the number of active clusters C is given by

$$C(\vec{X}) = C_0(\vec{X}) \cdot p^x \cdot (1-p)^{n_C(\vec{X})-x} \tag{6.5}$$

$$= \frac{N}{n_C(\vec{X})} \cdot p^x \cdot (1-p)^{n_C(\vec{X})-x}. \tag{6.6}$$

Here, p is a probability that mainly depends on n_{th}, but also on the neighboring order g. It is important to note that this description is empirical, and that a_i and p must be fitted to the simulation data.

The results of the Monte Carlo simulation demonstrate the expected behavior of the cluster detection algorithm. Increasing the neighboring order, g, leads to an increase in the number of clusters originating purely from noise since pixels that exceed the intensity threshold can be separated more and still count as a single cluster. On the other hand, decreasing the intensity threshold, n_{th}, leads to an increase in the number of clusters due to noise since it is more likely that a pixel will exceed the threshold, and clusters can form. However, a combination of a threshold value of $n_{th} = 3$ (corresponding to 3σ) and a neighboring order of $g = 2$ offers a good balance between noise reduction and enhanced cluster detection for weak signals.

It is important to note that image sensors can also be affected by correlated noise between pixels, e.g., due to temperature changes, unstable supply voltage, or fluctuations in the ADCs. These effects can cause higher insensitivity on

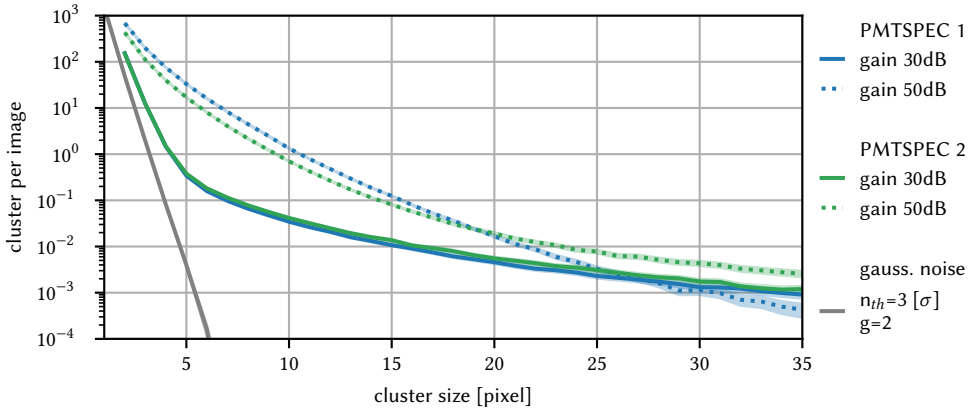

Figure 6.8: Comparison of the average cluster size detected per image for the two cameras and across two gain periods. The clusters were constructed with $n_{th} = 3$ and a neighboring order of $g = 2$. By nature, cluster detection is prone to detecting clusters based on noise, particularly for smaller clusters. To account for this effect, the independent Gaussian pixels noise component is plotted, as shown in Figure 6.7. It is important to note that correlated noise can also affect image sensors. Therefore, the noise shown is an underestimate of the actual noise present.

pixels simultaneously, leading to more clusters. Therefore, the noise contribution estimated from the Monte Carlo simulation is a lower estimation of the actual noise component. Since the detailed noise contribution for a specific camera is unknown, it is not straightforward to model it further. Instead, Figure 6.8 displays the detected clusters for the two different cameras and the two different gain periods along with the simulated contribution of independent Gaussian pixel noise.

6.2.2 Processing and Cluster Parametrization

The cluster detection algorithm is applied to all images captured by the three deployed cameras. For each detected cluster, various parameters are extracted and stored in a database, enabling further analysis. However, some parameters depend on the cluster detection algorithm rather than the cluster itself.

To calculate the mean dark noise μ_i and its standard deviation σ_i per pixel, only a subset of images are used, due to the large number of images and processing requirements. The images are grouped by their capture time into eight-hour

blocks, with approximately 300 images per block. The block period and the number of images used to compute μ_i and σ_i are stored in the database. This approach has two benefits compared to computing μ_i and σ_i over all images and then detecting the clusters in a second step. First, it simplifies parallel processing with multiprocessing, as each image set can be processed independently on a single process without needing to communicate with other processes. Second, importing a single image DAQ file only once significantly speeds up the algorithm, as accessing a file from disk can be a major bottleneck.

Another parameter is the image capture time. All other parameters are observables based on the pixels of the cluster. The number of active pixels, the integrated pixel intensity, and the noise μ_i are calculated for the whole cluster and for each of the three color channels separately. Three different positions for each cluster are also measured, where

$$\vec{R}_{\text{pixel}} = \frac{1}{n} \sum_{i=0}^{n} \vec{r}_i, \tag{6.7}$$

is the center of all involved pixels in the cluster,

$$\vec{R_{\text{int}}} = \frac{1}{\sum_{i=0}^{n} I_i} \sum_{i=0}^{n} I_i \vec{r}_i, \tag{6.8}$$

the center of the pixel intensities weighted by the intensities. $\vec{r}_i$ are the pixel coordinates and I_i is the pixel's intensity without the pixel noise μ_i component. The third position measurement is the center of the minimum rectangle bounding box, which covers all active pixels in a cluster. The rotation and dimensions of the bounding box are also stored. A full list of parameters and observables are summarized in Section C.1.

The minimum cluster size is set to 4 and 10 pixels for the two different gain periods with 30 dB and 50 dB, respectively, to limit the number of detected clusters and speed up the processing. Smaller clusters are ignored in the following.

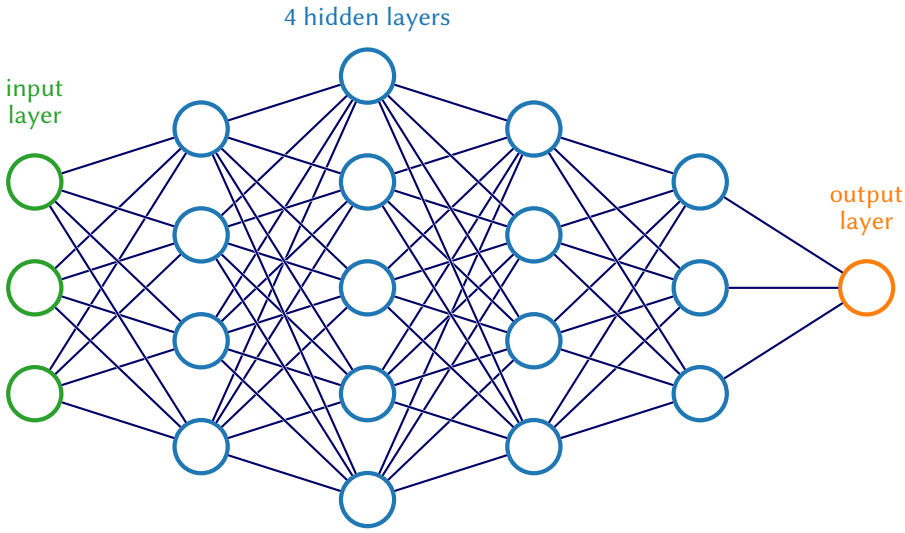

Figure 6.9: Illustration of a fully connected neural network (NN). This figure depicts a representation of a dense network, commonly referred to as a fully connected model. Each neuron in one layer is connected to every neuron in the subsequent layer, enabling information flow from one node to any node in the following layer. Courtesy to Izaak Neutelings.

6.3 Machine Learning Cluster Classification

The cluster detection technique introduced in Section 6.2 is susceptible to noise-induced clusters, particularly for small clusters. Thus, it is necessary to separate the noise. One common approach to classifying data is based on artificial artificial neural networks (NNs), where the input is connected to multiple hidden layers before reaching the output nodes, as shown in Figure 6.9. This architecture enables modern NNs to handle up to several million free parameters, which are adjusted based on a set of training samples. This procedure is referred to as machine learning (ML), in analogy to the human brain. After training, the NN can make predictions on new data not seen by the model before. These predictions can be used for classification. Machine learning can be broadly categorized into three methods: supervised, unsupervised, or reinforcement learning. Supervised learning requires training data where the input-to-output mapping is known, whereas unsupervised learning identifies underlying patterns and reinforcement learning optimizes a reward function to perform the best action in the next step.

The classification outlined in this section works with supervised learning, which rests upon the presents of a bioluminescence hotspot as visible in Figure 6.10. Furthermore, the holding structure limits the camera's FoV. Hence, outside the FoV, clusters must originate from noise with high probability. Consequently, one area is dominated by bioluminescence, and another is nearly entirely noisy. All clusters within the two regions are labeled accordingly and used as the training data for supervised learning so that the model can learn about the features of the two groups. Since there are many more clusters from noise than bioluminescence, a score matrix is introduced to counterbalance the learning process's imbalance. Afterwards, the trained model categorizes all clusters across the entire image.

6.3.1 Hotspots of Clusters

Plotting the cluster pixel center $\vec{R}_{\text{pixel}}$ from Equation 6.7 shows several features visible in Figure 6.10. There are two hotspots dominated by noise and another by bioluminescence.

The noise or pattern noise concentrates in a blob to the lower left part of the image and a band along the upper and right image boundary. Indeed, the blob appears only in the 50 dB gain period and completely disappears for a gain of 30 dB, which is illustrated in Figure 6.11. Contrary does the band, which is nearly gone for 50 dB gain. The band is also visible as a bright area in pictures captured with 50 dB gain, e.g., Figure 6.4. The band does not shop up for 50 dB gain because the pixels are saturated or close to and, therefore, can not exceed the threshold of the cluster detection.

The cluster concentrates in an area around the data cable and steel line. Those clusters are most likely from bioluminescence, as organisms emit light after mechanical stimulation by touching an obstacle or turbulences around obstacles induced by water currents. Furthermore, the module's FoV limit also shows up as a limit of the bioluminescence hotspot. Noticeably the steel line has more clusters close by concerning the data cable. A reason could be that the data cable is not at a fixed position due to the lower tension concerning the steel line. Therefore the data cable has a smeared signature.

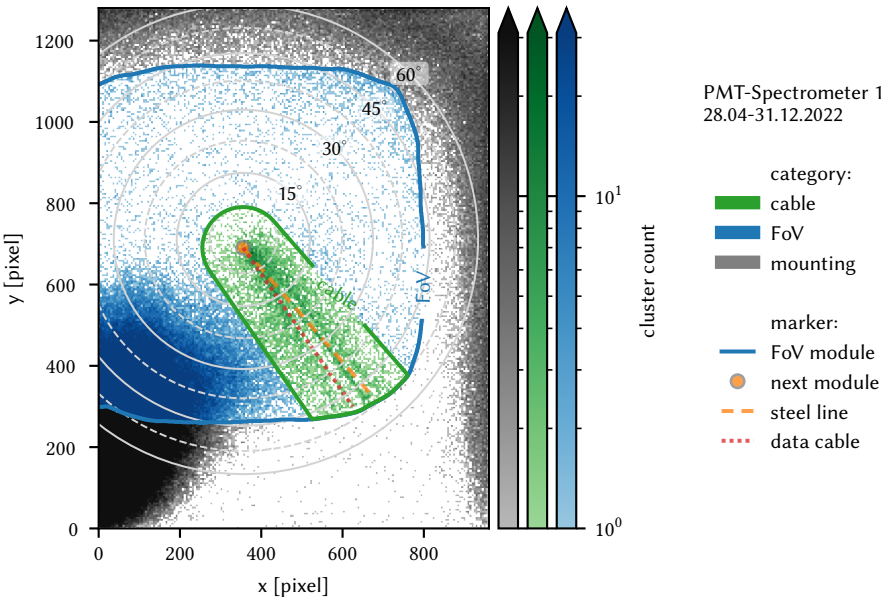

Figure 6.10: Cluster distribution of the cluster's pixel center. Noise is mainly located outside the module's FoV (blue line) where no light is expected. Only the accumulation in the lower left enters the module's FoV. Because the area of the cable is visible and bioluminescence emits light after mechanical stimulation, clusters in this area are from bioluminescence with a high probability. All clusters within a 100 pixel are allocated to the category *bioluminescence*. Supervised ML trains on the camera noise outside the module's FoV (black) and bioluminescence clusters around the cables (green). Afterwards, ML can classify clusters across the entire image including clusters that have not been used for the training, i.e. clusters which are inside the FoV but further away from the cable (blue). Figure C.1 shows the cluster captured by the camera in the PMT-Spectrometer 2 module.

Figure 6.10 shows the cluster distribution exemplary for the camera in the *PMT-Spectrometer 1* module. Fortunately, the other camera in the *PMT-Spectrometer 2* module represents the same characteristics, and only the data cable and steel line position is rotated.

6.3.2 Initial Classification for the Training

Supervised learning needs a data set where the output is known to fit its model. Hence all clusters in the training data need a label to which group they belong, i.e., *noise* or *bioluminescence*. As shown in the previous Section 6.3.1 there is an

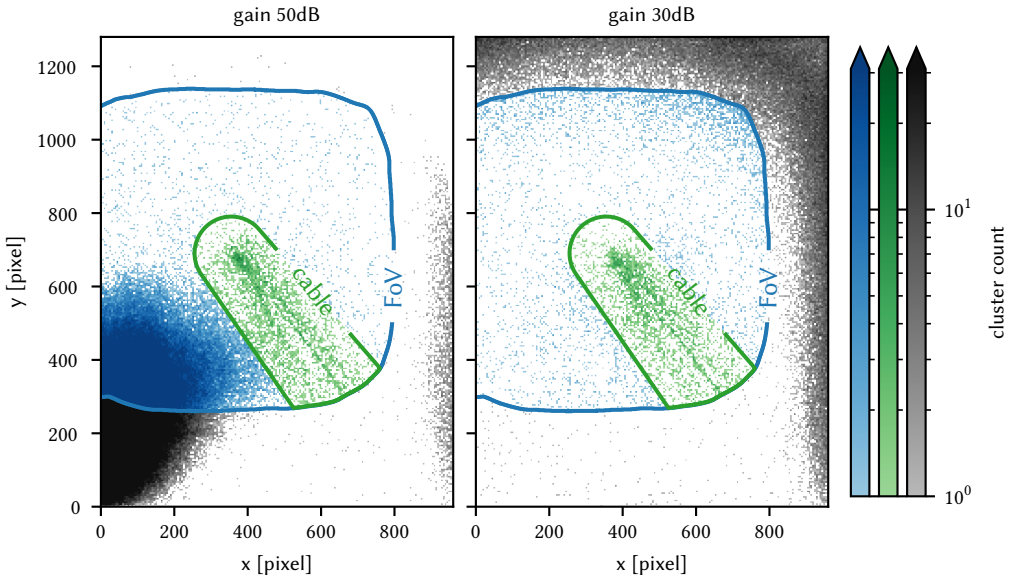

Figure 6.11: Cluster distribution respects the two different gain settings in the camera. Clusters are from the PMT-Spectrometer 1 module's camera, and the pixel center of each cluster is plotted. With a gain of 50 dB, the noise concentrates in the lower left part of the image. In contrast, a gain of 30 dB has the noise along the upper and left image boundary. The reason for the two different noise patterns refer to the text

area in the image dominated by bioluminescence. Therefore, all clusters within a certain distance to the cables are allocated as *bioluminescence*. The maximum distance is set to 100 pixels by visual inspection of Figure 6.10 to separate from the pattern noise in the lower left part of the image. Assigning clusters to the *noise* category utilizes the camera's FoV limitation from the PMT holding structure. In this area, clusters predominantly originate from noise with a minor component of reflections. As the PMT holding structure is made out of black plastic and bioluminescence is not very bright, the chance of reflections is low. Therefore, all clusters outside the camera's FoV are assigned to be *noise*. Clusters that do not belong to one of the two groups are not used for the training.

This initial classification is more probabilistic as specific. For example, an original noise cluster randomly can appear close to the cable and, therefore, is mislabeled as *bioluminescence*. Also, the mentioned reflections at the PMT holding structure could create clusters with a similar signature to *bioluminescence*

but are filtered as *noise* due to the cluster position. However, as ML is also a probabilistic method, the idea is that the model should learn on the majority of correctly labeled clusters without getting disturbed by the minority of falsely labeled ones.

6.3.3 Feature Selection and Normalization

Feature selection and normalization are two essential techniques in the field of ML to reduce the computational cost, especially the training time, and to improve the model's performance and robustness.

Feature normalization is to scale the observables to have similar ranges and distributions. The goal of feature normalization is to ensure that the features are on the same scale so that the model is not affected by the different units of measurement of the features. This is particularly important when using algorithms sensitive to the input data's scale, such as distance-based algorithms and gradient descent. Common normalization techniques include standardization, which scales the features to have a mean of zero and a standard deviation of one, and min-max normalization, which scales the features to a fixed range, such as 0-1. However, most of the ML models apply normalization measures by trainable parameters, and in some cases, normalization even degrades the performance.

For this classification task, the square root is first taken for all charge and noise and pixel observables to compensate for the area they describe. For example, a cluster with doubled radius has approximately four times the pixels. Furthermore, the charge and noise increase by a factor of four as they refer to the sum of pixel parameters. Second, the charge parameters are further logarithmized as they are an intensity that suffers from absorption by the water. And finally, the features are scaled to have a mean of zero and a standard deviation of one.

On the other hand, feature selection is the process of selecting a subset of the most informative and relevant observables. Feature selection is important because it can improve a model's performance by reducing the dataset's dimensionality and removing noise and irrelevant features. This can lead to a more straightforward and interpretable model and faster training and

prediction times.

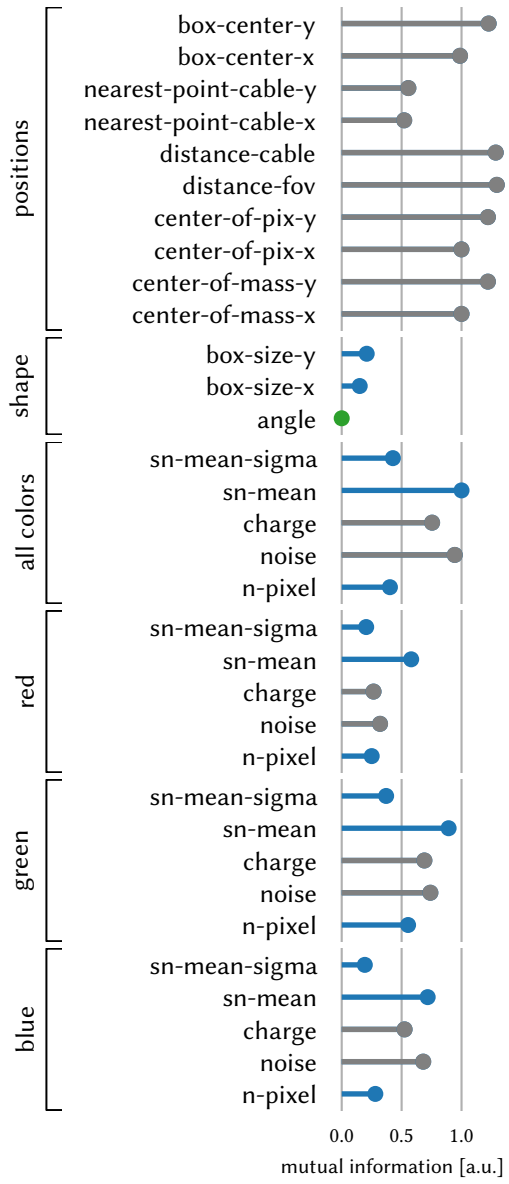

Figure 6.12: Mutual information (MI) of the all observables. *Blue* are the features used for the MI to build its model. *Gray* are all features excluded because of an intrinsic position dependency, and *green* are excluded because of to less MI.

There are several methods for feature selection, including filter methods, which use a criterion to evaluate the importance of each feature, and wrapper methods, which use the model's performance on a subset of features as the criterion. In a classification problem, feature selection is essential to identify the most relevant features for the classification problem, which can improve the model's accuracy. Also, it can help to eliminate noise and irrelevant features.

The dependency between the features and the discrete target class is illustrated in Figure 6.12 using the concept of mutual information (MI). MI is a non-negative value that quantifies the relationship between two variables. A value of zero indicates independence, while higher values indicate stronger dependency. In this case, the function employed utilizes non-parametric techniques that estimate entropy from k-nearest neighbors distances [63, 64].

Obviously, all position-related parameters show the highest MI in Figure 6.12 because it is how the classes are generated. Remarkable is also that the noise parameters show a high relation. A reason for it is that the cluster's noise is dominated by pattern noise which couples with the position of the cluster. The same effect appears in the charge features because the charge is

the sum of all pixels, which exceeds the noise level by 3 σ. Moreover, the σ depends again on the pattern noise yielding a hint to the position. Therefore, those features which are somehow coupled to the place on the image are not used for training as it could lead to wrong estimates. The angle feature is close to zero MI and is excluded, too. All remaining parameters are not connected to the position and show at least some relation also if it is weak. Therefore those are used for the classification.

To verify the selection, a decision tree classifier is used. Decision tree classifiers are another feature selection technique. It yields a comparable selection which is shown in Figure C.2.

6.3.4 Unbalanced Data and the Score Matrix

Unbalanced data refers to a situation where the distribution of classes in the training data is uneven, with some classes having many more examples than others [65]. For example, in the case of the image clusters, the initial class ratio, *bioluminescence* to *noise*, is below 8%. It is called initial classes because the class assignment is based on the cluster location and therefore is not perfectly separated, i.e., there is noise in the *signal* class as outlined in Section 6.3.2. The following describes the measures to mitigate this problem of unbalanced data for the simplified case of two existing classes where *true negative* and *true positive* are correct predictions of the minority and majority class, respectively.

Training a ML model with unbalanced data can make the model biased towards the more prevalent class, resulting in poor performance in the underrepresented classes. One way to address this issue is to use techniques such as oversampling the minority class or undersampling the majority class to balance the data distribution. Another approach is to use cost-sensitive learning, where different misclassification costs are assigned to other classes, to adjust for the imbalance in the data [65].

Balancing the classes helps to equalize the classification performance of both classes. Compared to a model trained without measures against unbalanced data, the class balancing decrease *true negative* and *false negative* while it increase *true positive* and *false positive*. As a result, balancing the classes raise the

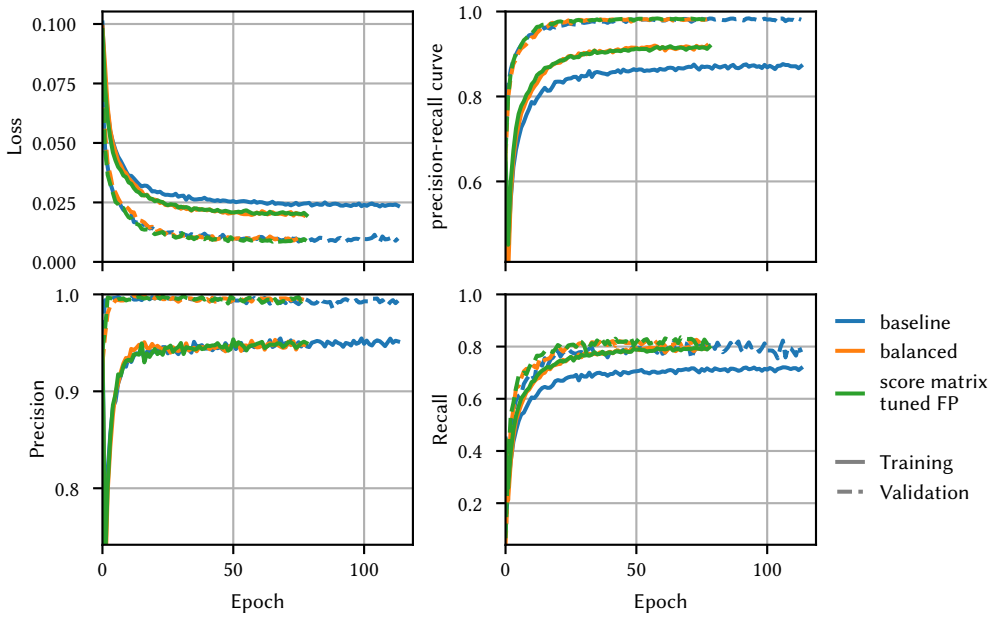

Figure 6.13: Impact of measures handling imbalanced data using the same deep neural network (DNN) architecture. The baseline model does not incorporate any special treatment for the imbalanced data, while the balanced model utilizes class weights to address the imbalance. The score matrix extension of the balanced model results in a decreased false positive (FP) score. Besides, the balanced model, as well as the model with the score matrix, needed fewer training iterations to find their optimum in less time.

recall

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{6.9}$$

while the precision

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{6.10}$$

remains relatively unchanged. This effect can be seen in Figure 6.13 during the training process using the model and optimization techniques described in Section 6.3.5. Nevertheless, it is important to note that this effect may not always be the case, as the model has the ability to adjust its parameters to address the imbalance.

Furthermore, these metrics must be evaluated in the context of each individ-

ual problem, as one type of error may be prioritized over the other. For example, clusters outside the FoV are considered *noise*, so the model should predict a very low number of *false positive*. On the other hand, some clusters around the cable are incorrectly assigned to the *bioluminescence* class and should be *noise* instead. Hence, the model should generate some *false negatives*. Therefore, incorporating cost-sensitive learning is required to weigh misclassifications differently. A simple implementation is scaling the loss function L with a parameter l_s, which accounts for different scores. The scaled loss function is

$$\hat{L} = L \cdot l_s \tag{6.11}$$

where $l_s = 1$ if all predictions $\vec{p}$ match the truth $\vec{c}$ and increase with more misclassifications. Instead of implementing the different losses directly in the loss function, this approach has the advantage that the loss function could be changed easily. The scaling parameter l_s is defined as follows where a scoring matrix is used

$$S = \begin{pmatrix} s_{\mathrm{TN}} & s_{\mathrm{FN}} \\ s_{\mathrm{FP}} & s_{\mathrm{TP}} \end{pmatrix} \tag{6.12}$$

rather than a loss matrix. The score matrix allows different scores s for the different scenarios, true positive (TP), true negative (TN), false positive (FP), and false negative (FN). A score for a correct prediction should be higher or equal to its incorrect counterpart, hence $s_{\mathrm{TN}} \geq s_{\mathrm{FP}}$ and $s_{\mathrm{TP}} \geq s_{\mathrm{TP}}$. The score of a prediction p_i for the sample i is

$$s_i = \langle p_i | S | c_i \rangle \tag{6.13}$$

where c_i is the true value and both, c_i and p_i, are $\binom{1}{0}$ or $\binom{0}{1}$ for indicating *negative* or *positive*, respectively. The scaling parameter l_s is the inverse mean of all scores with

$$l_s = \frac{N}{\sum_i^N \langle p_i | S | c_i \rangle} \tag{6.14}$$

where N is the total number of samples in the training dataset. Normalizing l_s

to 1 equals

$$1 = \frac{\sum_i^N \langle c_i|S|c_i\rangle}{N} \tag{6.15}$$

$$= \frac{n}{N}s_{\mathrm{TN}} + \frac{p}{N}s_{\mathrm{TP}}. \tag{6.16}$$

with n the number of *negative* samples, p the number of *positive* and $N = n+p$. To make both classes contribute equally to the score, the following must hold:

$$\frac{n}{N}s_{\mathrm{TN}} = \frac{p}{N}s_{\mathrm{TP}} = \frac{1}{2N} \tag{6.17}$$

which yields

$$s_{\mathrm{TN}} = \frac{N}{2n}, \quad s_{\mathrm{TP}} = \frac{N}{2p}. \tag{6.18}$$

The score matrix can be simplified to only two free parameters, a and b, by defining $s_{\mathrm{TN}} = bs_{\mathrm{FP}}$ and $s_{\mathrm{TP}} = as_{\mathrm{TP}}$, where a and b are in the range of $(0, 1]$ and determine the lower score assigned for a false negative and a false positive, respectively. The resulting matrix is given by:

$$S = N\begin{pmatrix} \frac{1}{2n} & \frac{a}{2p} \\ \frac{b}{2n} & \frac{1}{2p} \end{pmatrix}. \tag{6.19}$$

This approach automatically accounts for the class imbalance in the training sample and provides the ability to tune the scores for different errors.

As it is known that a cluster outside the FoV must originate from noise, the score of a *false positive* prediction s_{FP} should be minimal or $b << 1$, e.g., $b = 10^{-6}$. *False negative* predictions are expected and should not be treated specially in the score matrix, leading to $a = 1$. Hence the score matrix for this classification task is

$$S = N\begin{pmatrix} \frac{1}{2n} & \frac{1}{2p} \\ \frac{10^{-6}}{2n} & \frac{1}{2p} \end{pmatrix}. \tag{6.20}$$

The impact of measures handling imbalanced data on the Figure 6.13 summarise which shows the precission and recall metric for no balancing (baseline), balancing the classes and a score matrix which is introduced in the following

6.3.5 Building, Training and Optimizing the Model

Building, training, and optimizing a NN model involves several steps. The first step is to define the model's architecture, which includes the number of layers, the type of layers, the number of neurons in each layer, and the activation functions. Then, during training, the model's performance is monitored using accuracy, precision, recall, and loss metrics. Finally, the model's performance is evaluated on the test data to determine its generalization ability. The process of building, training, and optimizing is iterative and requires careful experimentation and tuning to achieve the best results. The following describes the experience of training models for this specific classification task.

The NN models that have been tested in this study all employ a fully connected architecture, also known as a dense network. This deep learning architecture connects every neuron in one layer to all the neurons in the subsequent layer. As a result, information can flow freely between all nodes, allowing the NN to optimize the parameter connections that perform the best.

To prevent overfitting, dropout layers were added after each fully connected layer. Overfitting occurs when a model becomes too specialized to the training data and fails to generalize well to new, unseen data. Dropout is a regularization technique that helps mitigate overfitting by randomly dropping out neurons during training. The dropout rate was determined to be optimal when set between 0.2 and 0.5.

Another common approach to avoiding overfitting is early stopping, which involves monitoring the model's performance on the validation dataset and stopping the training process if the performance breaks improving or begins to deteriorate [66, 67].

The model optimization was performed using cross entropy as the loss function with the score matrix extension explained in Section 6.3.4. Cross entropy is a standard loss function used in classification problems requiring the model to predict class labels. It measures the dissimilarity between the predicted and target probability distributions, helping the model learn to optimize its parameters and reduce the error between the predicted and true outputs [68].

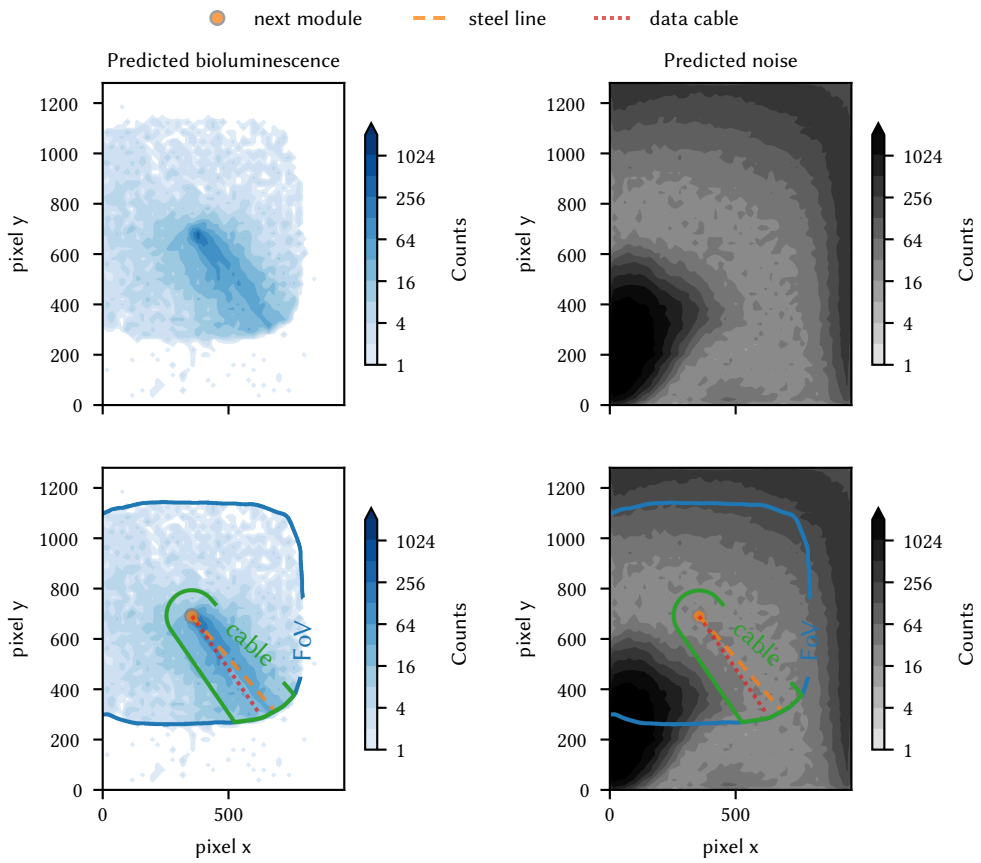

Figure 6.14: Cluster location distribution demonstrating the effectiveness of ML classification for clusters with a size ≥5 pixel. Absolute cluster count is plotted instead of density to highlight the minimum number of clusters in certain areas. Plots in the same color show the same data, where the lower row has marked FoV and cable areas originate from Figure 5.7. Predicted bioluminescence nearly perfectly appears only in the FoV whereas there is a minor increase of the density at the cable in the predicted noise (*false negative*). NN models trained with a different minimum cluster's pixel size of 4 and 6 pixels are shown in Figure C.4 and Figure C.5. This plot shows clusters from the camera in the PMT-Spectrometer 1 module and Figure C.3 from the camera in the PMT-Spectrometer 2 module.

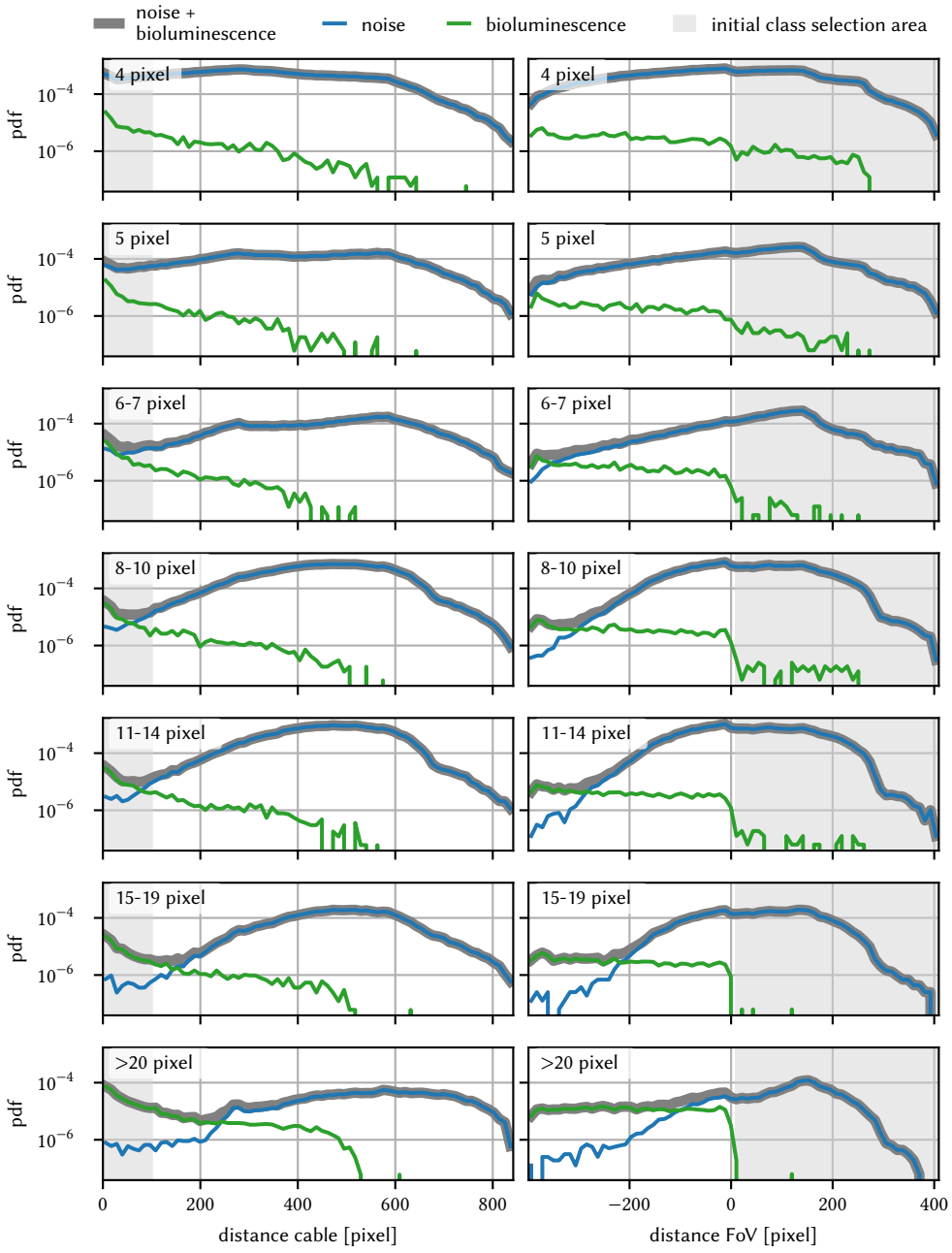

Figure 6.15: Probability density function (PDF)s of the cluster locations depicting the performance of the ML classification. The PDFs are generated for all clusters and their individual class components. The clusters are classified as ”bioluminescence” (*positive*) or ”noise” (*negative*) based on their proximity to the cable or their location within the FoV. Both areas are marked in the distributions, demonstrating the effectiveness of the NN in separating the two classes without prior knowledge of their exact positions.

To evaluate the performance of a model, a commonly used approach is to utilize a dataset that is separate from the training data and whose true classifications are known. However, in this particular case, such a dataset is not available, and the bioluminescence cluster samples contain a certain degree of noise. Nonetheless, it is possible to assess the model's performance by examining the position of the clusters as classified by the NN, as shown in Figure 6.14. This approach allows for a qualitative evaluation of the model's performance, even in the absence of a quantitative benchmark dataset.

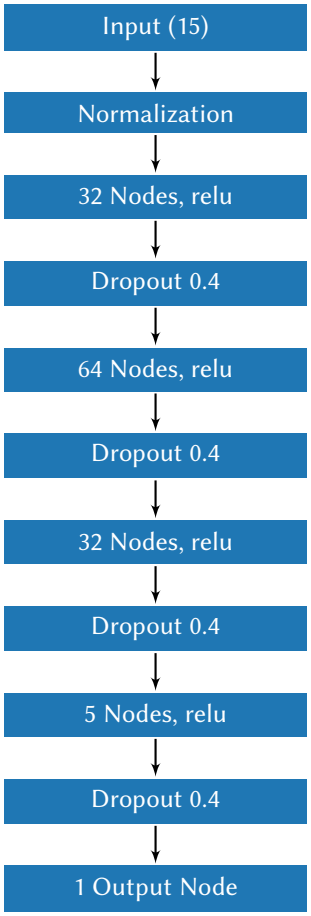

Figure 6.16: The NN classifier architecture.

As previously discussed, bioluminescence clusters should not be predicted outside the FoV, and the density of noise clusters should not show an increase concerning the cable's position. Since both targets cannot be achieved simultaneously, a model selection involves making a trade-off between these two features. Figure 6.15 shows a reasonable classification also for the smallest recorded clusters of 4 pixels. However, the performance on 4 pixels is by at least one magnitude worth on *false positive* as for greater clusters. Therefore, the cluster size limit is set to 5 pixels ignoring smaller clusters. The NN is trained again because it improves the performance when the target classes are known more precisely, which is the case for bigger clusters. Figure C.6 illustrates the improved classification with the retrained NN. Training the NN on all cluster sizes or applying a threshold of 4 and 6 pixels results in Figure C.4 and Figure C.5, respectively.

The results of testing various NN models revealed that a robust architecture capable of delivering reliable results on datasets captured by two cameras and under two different gain settings necessitates using at least three hidden layers. Further analysis uncovered that the last hidden layer should have between 5 to 8 nodes, while the middle layer should have a node count of more than double

the number of nodes in the input layer. Enhancing the size of four hidden layers and increasing the second hidden layer, even more, was also found to contribute to a modest increase in the stability of the NN model. The architecture utilized can be viewed in Figure 6.16.

6.4 Conclusion

The chapter presents a method for detecting bioluminescence in dark scenes through deterministic cluster detection and a NN classifier. First, all clusters are extracted from an image together with a set of observables. The cluster detection mechanism is based on a noise threshold and the maximum distance between pixels to be considered a cluster. Known areas in the images where clusters are more likely to originate from either *noise* (outside the camera's FoV) or *bioluminescence* (around the cables) are used as a training sample for supervised learning. Subsequently, position-related observables are removed from the training sample to train an unbiased NN. After training, the NN can classify all detected clusters and filter out clusters originating from *bioluminescence*. Despite the initial class assignment for the training sample not being 100 % accurate, the filtering performance is remarkable, even for small clusters with as few as 5 pixels.

One of the advantages of this approach is the ability of the NN to handle different camera gain settings and corresponding changes in noise without splitting the data beforehand. Additionally, the method does not need extensive computing resources or long computational time. Detecting all clusters on an image, including extracting all additional observables, takes about 1 second per image on a single CPU core. Furthermore, the small size of the neural network architecture also results in a fast training time, taking less than a minute on a standard CPU for about 5 million detected clusters, which allows for quick tuning of the parameters.

Overall, combining cluster detection and the NN classifier provides an effective mechanism for detecting bioluminescence in dark scenes, which could also be transferred to similar applications.

7 Exploring the Ocean Environment Through Camera Analysis

In this chapter, the environmental analyses performed on the images captured by the cameras integrated into Strings for Absorption Length in Water b (STRAW-b) are discussed. The aim is to gain insights into the mooring movement (Section 7.1) and measure the extent of biofouling (Section 7.2). The analyses presented in this chapter utilize the neural network (NN) aided cluster detection technique, which is discussed in detail in Chapter 6.

7.1 Tracking the Mooring Movement in Images

Mooring lines, like STRAW-b, operate in a constantly changing oceanic environment, where water currents can vary in both speed and direction. Due to this, mooring lines move with the water flow, while the anchor remains the only fixed point.However, the initial classification for the NN presented in Section 6.3.2 assumed a static cable and neighboring module positions in the images. Indeed, the cable and module positions were determined from two images with enabled flashers, and all clusters near the cable and module were assigned as bioluminescence targets for the NN classifier. Tracking the actual module positions can provide insight into the validity of the assumption that relative positions between modules remain static or if a non-static model could improve the initial classification.

Over a span of six months, images were captured by the camera located in the PMT-Spectrometer 2 module with enabled LEDs in the neighboring module (LiDAR 2). Specifically, once per hour, a set of images with varying LED configurations were taken by the camera. Further details regarding the LED flasher

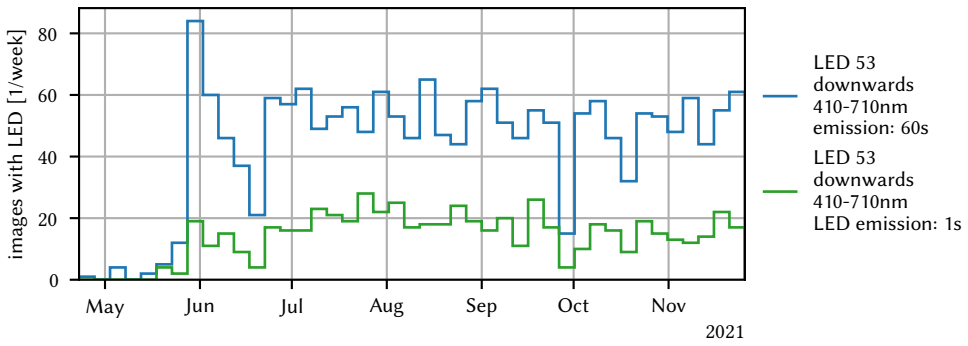

Figure 7.1: Timeline of images with enabled LED in the neighboring module. The measurement stopped end of 2021 because the connection to the module with the LED was lost.

system and the measurement integration can be found in Chapter 4.

Among the various LED configurations used for imaging, the images captured by the white downwards-facing LED were chosen due to their brighter appearance caused by the attenuation spectrum. This improved the stability of the module tracking. Moreover, the symmetrical shape of the downwards-facing LED and its direct alignment with the upwards-facing camera increased the image brightness and facilitated the extraction of the LED position. The selection of the desired LED configuration was based on the recorded timestamps, as the data acquisition (DAQ) system registered the activation of each LED with its corresponding configuration.

After filtering the images, the detection of the cluster introduced in Section 6.2 provided a set of parameters, including different location calculations listed in Table C.3. The LED emission cone projects as a symmetrical cone in the image, and the cone's center represents the module's location. Among the different position definitions, the one that considers the pixel intensity (Equation 6.8) provides a robust tracking of the symmetrical cone's center, where the intensity is maximized in the center.

Despite the hourly image capture schedule, the number of successful measurements is lower due to several factors, including camera malfunctions, inter-module communication issues or delays, and the DAQ system's ongoing de-

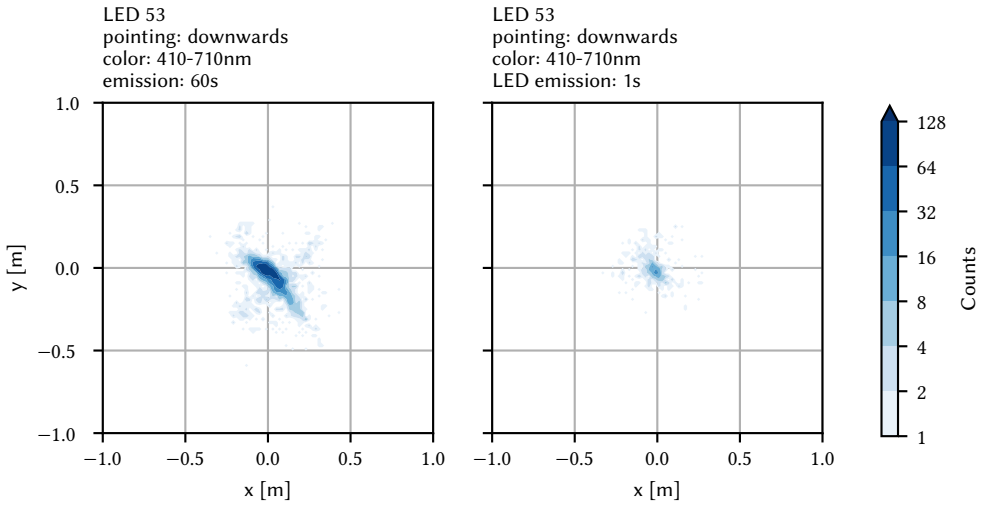

Figure 7.2: Tracking the movement of the neighboring module. The origin (0, 0) is the initial module position based on a single image. The x and y coordinate in meters calculate from the known distance of 24 m, and the camera projection is introduced in Chapter 5. The two plots show two different LED configurations, where the left is less bright but with a longer emission time than the right. The plots signifies that the modules are aligned synchronously within the currents, because no rotation with the doubled distance light-emitting diode (LED) to cable of ~0.6 m (2·0.295 m) is visible.

velopment during the study period. Additionally, biofouling affected the light's path and disrupted the symmetry of the light cone, leading to less accurate position calculations. Figure 7.1 illustrates the image success rate after applying the filters as mentioned earlier.

The cluster with the shortest distance to the initial module position for each image was selected. The pixel coordinates were then projected onto cartesian coordinates using the 24 m distance between the modules and the camera projection discussed in Chapter 5. This resulted in a heat map of all tracked positions, as shown in Figure 7.2, with most positions within 0.2 m radius. The left plot in Figure 7.2 exhibits a clear eigenvector of movement following the plot's diagonal, indicating that the modules are aligned synchronously within the currents, despite the current direction changing every 6 h. Additionally, only minimal induced rotation along the steel cable where modules are mounted can be observed.

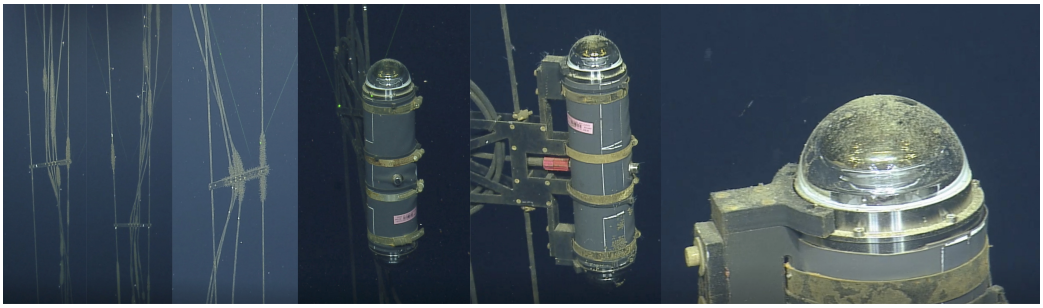

Figure 7.3: Collage of the STRAW inspection from September 2020 operated by Ocean Networks Canada (ONC) [27]. Biofouling appears at the modules and along the cables.

While the assumption that relative module positions remain almost static is valid, a deep-sea neutrino detector requires absolute positioning of all modules with high precision, within a few centimeters, over a distance of more than 1 km. Until absolute location measurements are available, the data presented in this section can serve as a basis for crosschecking hydrodynamic detector simulations in combination with the actual currents at the site.

7.2 Biofouling on the Detector

Objects in the ocean undergo three different processes where their surfaces become covered by various substances: marine snow, sedimentation, and biofouling (MSB). Over time, the layer of accumulated material on these objects can attenuate light in detectors, becoming increasingly dense and affecting their performance. Further investigation is required, but an inspection dive revealed that biofouling contributes the most to detector contamination, as shown in Figure 7.3. The inspection also suggests that contamination increases towards the sea floor. This section first summarizes the underlying process of biofouling before it outlines an indirect measurement of the contamination with the camera. Then, for future detector arrays, a possible direct measurement is presented and measures against biofouling, called anti-fouling, are summarized.

7.2.1 Underlying process of Biofouling

Biofouling begins with *pioneer bacteria* gradually covering the surface with a biofilm. Once established, other organisms build up on it, strengthening the adhesion, which makes it difficult to remove it again. Anywhere where water is present, biofouling can occur and degrade the device's primary purpose to which the surface belongs. For example, on marine vessels, the accumulation increases both the hydrodynamic volume of a vessel and the hydrodynamic friction, leading to increased drag which can increase the fuel consumption by up to 40% [69]. Although, it is not an aspect of a neutrino telescope; fouling on ships can transport and introduce non-indigenous, invasive species, which pose a significant risk in various facets [70]. This subsection first covers a technique to measure biofouling before it summarizes different options to prevent biofouling within the scope of an application in a deep sea neutrino telescope.

7.2.2 Indirect Measurement of Biofouling with the Camera

The brightness levels of the clusters are considered to measure the effect of marine snow, sedimentation, and bio-fouling (MSB) with the camera. The cluster brightness is defined as

$$b_c = \frac{1}{N} \sum_{i=0}^{N} \frac{p_i - \mu_i}{\sigma_i} \tag{7.1}$$

over all active pixels i with its individual pixel noise level μ_i and the noise standard deviation σ_i, to compensate for the effects of fluctuations in the electronic noise baseline. Figure 7.4 shows the mean brightness of all recorded images in periods of 27 days. Selecting 27 days over a monthly period reduces interference effects. Because bioluminescence is strongly related to the water current, which is again related to the tides that are induced by the moon. Therefore, the moon period of 27 days appears in the currents and the bioluminescence.

The errors displayed represent the statistical error, including the cluster detection error. When multiple pixels close to the threshold are present, it can cause a significant error in the cluster brightness. A linear fit analysis reveals an annual loss of 34.0%±24.5% and 34.3%±19.9% for the PMT-Spectrometer 1 and 2 cameras, respectively. Remarkably, both cameras show a similar loss over time

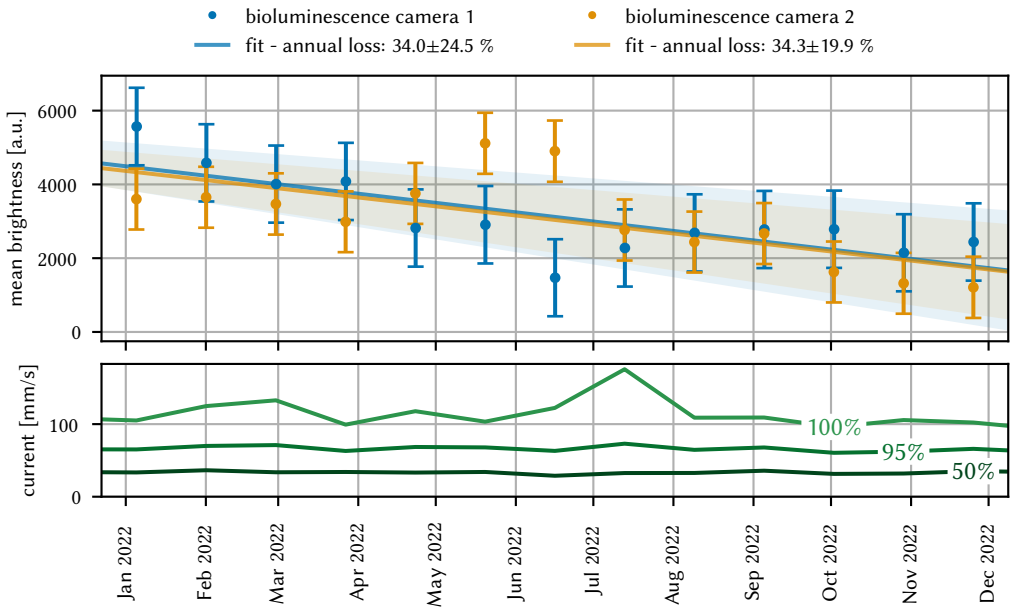

Figure 7.4: Cluster brightness loss in the images. The data show the average brightness calculated in the cluster detection (Equation 7.1) for periods of 27 days to match the moon period, which strongly influences the currents and again influences bioluminescence. The current percentiles are plotted in the lower plot to avoid being biased by current changes.

despite being separated by a depth of 260 m. In order to crosscheck if the currents could cause the loss, the current data is also plotted in Figure 7.3. However, the average currents remain constant during the observed period, indicating that the currents are unlikely to be the bias causing the loss in average brightness. Although the measurements indicate a loss, the error margins are high, including only a minimal loss of a few percent per year.

7.2.3 A Possible Direct Measurement of Biofouling

Location and material have a strong impact on the speed of biofouling. Predictions on the speed are difficult and show to be inaccurate, leading to measure the stemming in situ where information about the contamination is required. A technique to monitor the biofilms stemming base on the stimulated fluorescence of microorganisms constituting the biofilm. More precisely, it predicates on the amino acid tryptophan, an organic molecule that dominates the intrin-

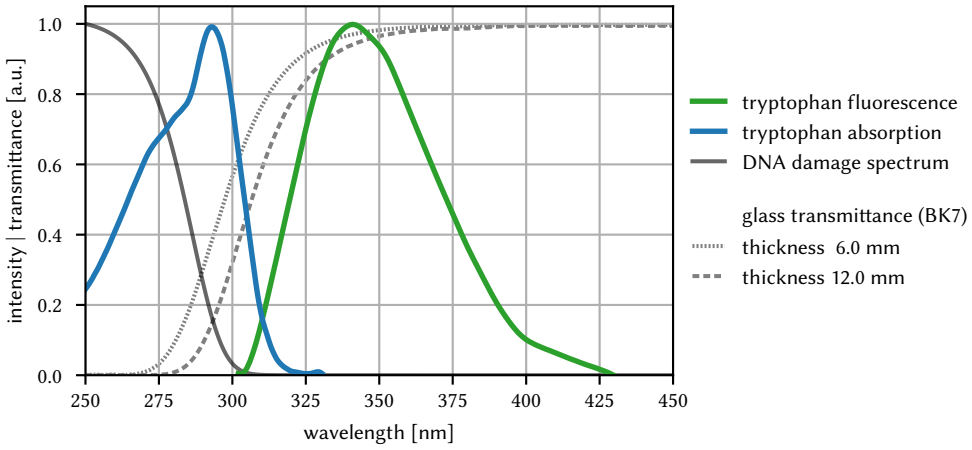

Figure 7.5: Tryptophan absorption and intrinsic fluorescence spectrum. Stimulated with a UVC-LED, the amino acid tryptophan fluoresces in the near-UV. A light detector such as a spectrometer can measure the fluorescence raveling the presence or even measuring the degree of contamination of biofouling [71, 72]. In addition, the plot shows the absorption of BK7 glass for different thicknesses and the DNA damage spectrum [54, 61, 73].

sic fluorescence when present. After stimulating amino acid tryptophan with ultraviolet (UV) light at 280 nm, it emits UV light between 308 nm and 350 nm depending on the polarity of the direct surrounding [71, 72].

The absorption spectra of the BK7 glass used for the pressure housings cut off wavelengths smaller 280 nm, impeding the measurement. Typically, the spherical pressure housings have a thickness around 10 mm. For reference, STRAW-b has spheres with 12 mm. As indicated in Figure 7.5, reducing the thickness does not help too much to increase the tryptophan stimulation. BK7 glass is common practice for deep sea pressure housings because of the reduced cost compared to quartz glass which therefore is transparent in UV up to 170 nm. However, stimulating tryptophan should be possible with BK7 glass, and the tryptophan fluorescence is well below the cutoff. Figure 7.5 shows both the absorption and fluorescence spectrum of tryptophan together with the transmittance spectra of the BK7 glass for different thicknesses [71, 72].

Besides measuring the contamination, UV light can prevent the biofilm from growing in an early stage by damaging the deoxyribonucleic acid (DNA) of the

microorganisms. Without biofilm, bigger organisms can not attach to the glass preventing biofouling if the system operates regularly. The DNA damaging spectrum proposed by the network for detection of stratospheric change (NDSC) steering committee is

$$A_{\mathrm{DNA}} = \begin{cases} \frac{1}{a_0} \exp\left[a_1 \left(\frac{1}{1+\exp[(\lambda-\lambda_0)/a_2]} - 1\right)\right] & \text{for} \quad \lambda \leq 370nm \\ 0 & \text{for} \quad \lambda > 370nm \end{cases} \tag{7.2}$$

with the parameters a_0=0.0326, a_1=13.82, a_2=9 and λ_0=310 [73, 74]. Figure 7.5 shows the spectrum normalized to 250 nm. To assess the biological effect E_{bio}, the radiation spectrum $E(\lambda)$ of the light source convolves with the DNA damaging spectrum to

$$E_{\mathrm{bio}} = \int_{\lambda} E(\lambda) \cdot A_{\mathrm{DNA}}(\lambda)\, d\lambda. \tag{7.3}$$

With respect to the tryptophan absorption, the DNA damaging spectrum rises at an even lower wavelength. This allows measuring the tryptophan stemming non-destructively for the microorganisms [71]. On the other side, damaging spectrum rises below the BK7 glass cutoff, and a DNA damaging light source should not be encapsulated with BK7 glass. Contrary, a BK7 pressure housing has the advantage of effectively shielding its interior from external UV radiation, such as protecting sensible PMTs. As the tryptophan fluorescence decays within some nano-seconds, there should be no direct afterglow effect which compromise the photomultiplier tubes (PMTs) [75].

7.2.4 Antifouling

More generically, processes to prevent biofouling from forming summarize under the name anti-fouling and group in toxic or non-toxic coatings and active systems by injecting energy like thermal, acoustic, or electromagnetic treatments.

Toxic coatings induced harmful environmental effects on marine organisms, and since 2008 a convention of the international maritime community (IMO) prohibit the use of harmful organotin compounds [76]. Non-toxic coatings prevent the attachments of microorganisms with hydrophobic or oppositional hydrophilic

surfaces. Hydrophobic base on low surface energy and therefore results in low friction especially preventing the attachment of more significant microorganisms. In contrast, hydrophilic coatings have, by definition, a strong affinity for water and increase the energetic penalty of removing water for proteins and microorganisms to attach [77]. Another option is building up a nanoscale layer on the surface with minimal attachment points similar to the skin of sharks and dolphins.

Active systems can also treat fouling as the previously mentioned DNA destruction with UV light. One option is to heat up the surface to not extreme temperatures of $40\,^{\circ}\mathrm{C}$ for some minutes with the drawback that heating needs much energy [78]. An alternative with lower energy consumption is ultrasonic vibrations dissolving the microorganisms from the surface and preventing biofilm creation.

Concluding the final assessment, maritime engineering developed several methods to prevent biofouling. For a neutrino, telescope biofouling can reduce the photoelectric sensitive area resulting in lower detector sensitivity and statistics. Furthermore, the telescope should operate for years or even decades to generate enough statistics for sensitive studies. Therefore, a combination of different passive and active systems is also promising because an in-situ evaluation is limited to shorter timescales. An outstanding position has UV light as it prevents biofouling and can measure the biofilm contamination within one system.

8 Bioluminescence at Cascadia Basin

When the wings [of a firefly] were tied together, abandoning their constant movement, the light faded. In the same way, the glitter of the sea, of fish and putrid wood depends on motion.

Domenico Bottoni, 1692 [79]

This chapter aims to provide a deeper understanding of bioluminescence using the images captured by the cameras in Strings for Absorption Length in Water b (STRAW-b). The neural network (NN) aided cluster detection technique outlined in Chapter 6 is utilized to track the bioluminescent features in the images and analyze their properties. Specifically, this chapter covers the periodic variations in the bioluminescence baseline rates, the spatial distribution of bioluminescence emissions in the images, the population of the emitting organisms, and the processes that trigger bioluminescence and relate the emission rate to the current speed and concentration of bioluminescent organisms.

8.1 Periodic Variations in Bioluminescence Rates

This section explores the variability of bioluminescence rates over time at the future site of Pacific Ocean Neutrino Experiment (P-ONE). It aims to address two main questions: whether bioluminescence occurrence is affected by seasonal changes and whether it varies with depth. To answer these questions, Figure 8.1 presents the bioluminescence image and cluster rates over a period of 558 days, from 21.06.2021 to 31.12.2022. Although the cameras have been recording images

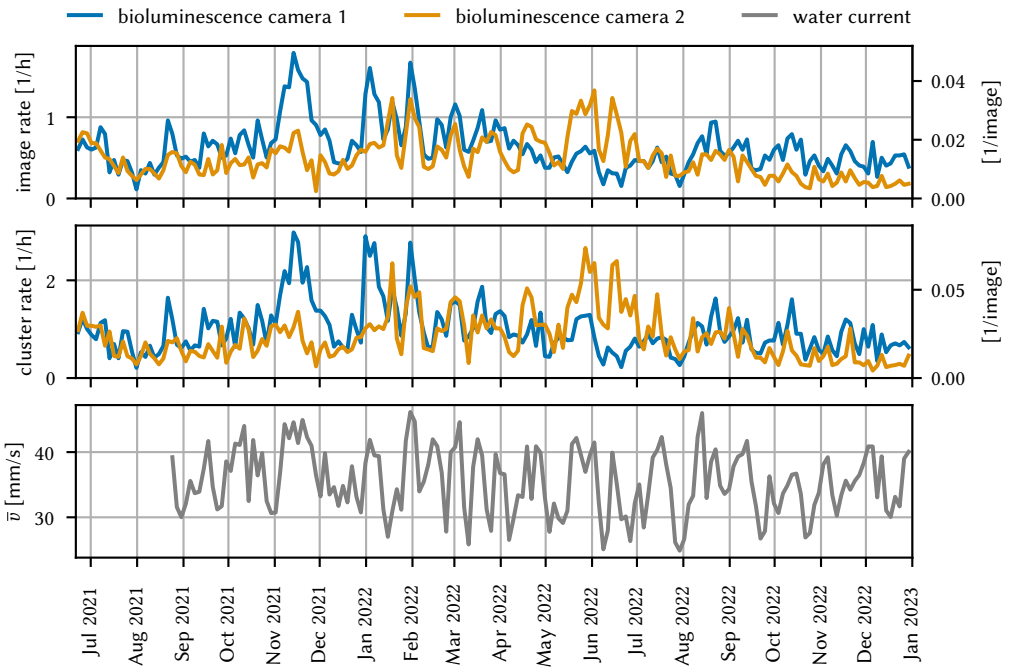

Figure 8.1: Rates of images and clusters showing bioluminescence normalized to the total number of recorded images. The lower plot displays the mean current velocity measured by a current meter located 2.1 km northwest of the STRAW-b position. The data is binned in three-day intervals for smoothing, and the rates are normalized to 99.5 s per image to compensate for downtimes. The absolute counts and rates are presented in Table 8.1.

since the end of April 2021, the initial few days of operation were problematic due to camera and data acquisition (DAQ) issues. Consequently, that period is excluded from the analysis.

The rates of bioluminescence images and clusters, which respectively count images with at least one bioluminescence cluster and individual bioluminescence clusters, are presented in Figure 8.1. Notably, these rates vary by almost an order of magnitude, also the rates are accumulated for three days. Most of the time, both cameras show a simultaneous increase or decrease in the rates, with comparable rates between them. Although both cameras show generally synchronous increases or decreases in their rates, with similar rates between them, some periods of up to a month exist where one camera records a substantially higher rate than the other before the curves converge again. While the exact rea-

Table 8.1: Summary of the detected bioluminescence in images. Camera 1 and 2 are located in the modules PMT-Spectrometer 1 and PMT-Spectrometer 2, respectively. The numbers in the table are for the period from 2021-06-21 to 2022-12-31, which is 558 days. Figure 8.1 shows the timeline of the rates. As outlined in the text, Camera 1 which is lower to the seafloor, captures around 20% more bioluminescence as Camera 2.

		Bioluminescence images		Bioluminescence cluster	
Camera	Images	Counts	Rate [1/h]	Counts	Rate [1/h]
Camera 1	476714	8356	0.634	13094	0.993
Camera 2	447790	6160	0.497	10564	0.853

son for this variation remains uncertain, it could potentially stem from different concentrations of bioluminescent organisms or variations in current speed along the water column. Regarding the depths of the cameras, the lower camera, positioned at a depth of 144 m, detected roughly 20 % more bioluminescence images and clusters compared to the higher camera located at 408 m (Table 8.1).

It is known that bioluminescence happens when the organism faces forces either by colliding with an object or caused by shear forces from turbulences in the water. Therefore Figure 8.1 includes the mean current, too. The relation between current speeds and bioluminescence rates can be found in Figure 8.15.

Another way to express the relationship is to consider the frequency space of both signals. The tides have different periodic components, mainly from the Earth (E), moon (M), and sun (S) rotation mechanics. The Earth rotates in T_E =24 h, but simultaneously, the moon rotates in the same direction. Therefore a point on the Earth's surface requires a period

$$t_{EM} = \frac{2\pi}{\omega_E - \omega_M} = \frac{1}{\frac{1}{T_E} - \frac{1}{T_M}} \tag{8.1}$$

between equal alignments with the moon, where ω denotes the angular velocity, and T the rotation period. This formula can be generalized to any two periods from the two rotating objects X and Y to t_{XY}. Harmonic resonances, which occur at frequencies that are integer multiples of a fundamental frequency f_0, can lead to the appearance of other periodicities in a system. Another type of period is

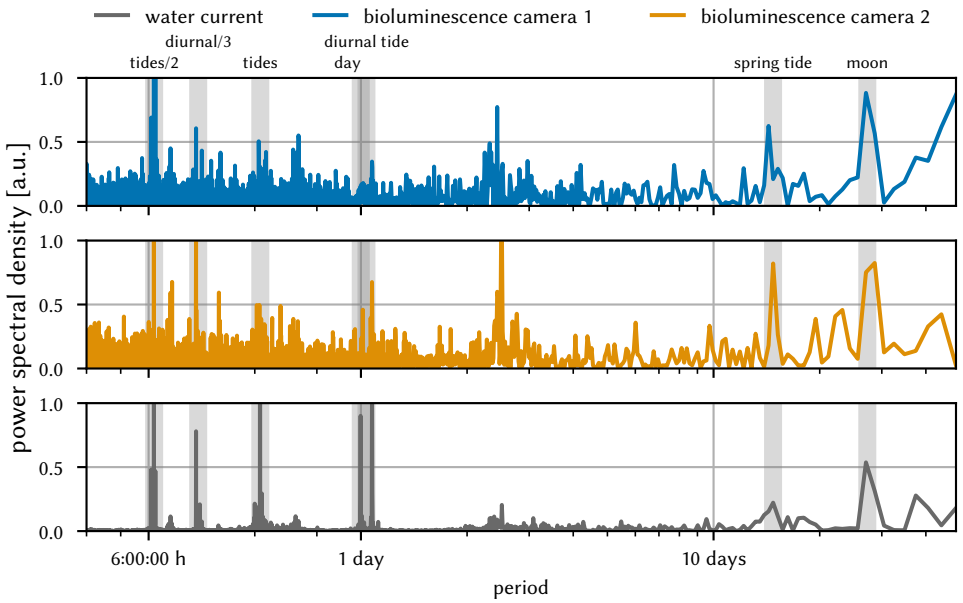

Figure 8.2: Periodogram of the water currents and the bioluminescence appearance in images. Important periods are marked. The x-axis is in logarithmic scale to cover the total range of periods and labeled with time classical nomenclature for readability.

when the moon, Earth, and sun line up, called spring tide, which is half the period ($t_{MS}/2$) of the sun and moon.

Figure 8.2 shows the periodograms for the two camera and measured current. All three presented periodograms are comparable with peaks located at the expected period from the mechanics. The most significant period in the three datasets is the 6-hour (6:13:41 h) period, which originates from the velocity change and therefore is half of the semidiurnal tidal period (12:27:21 h). Around 8 h (8:18:14 h) there is a harmonic resonance of the diurnal tide (24:54:42 h). Also the moon period of 27 days and the spring tide are clearly visible. What cause the period around 2.5 days remains unclear.

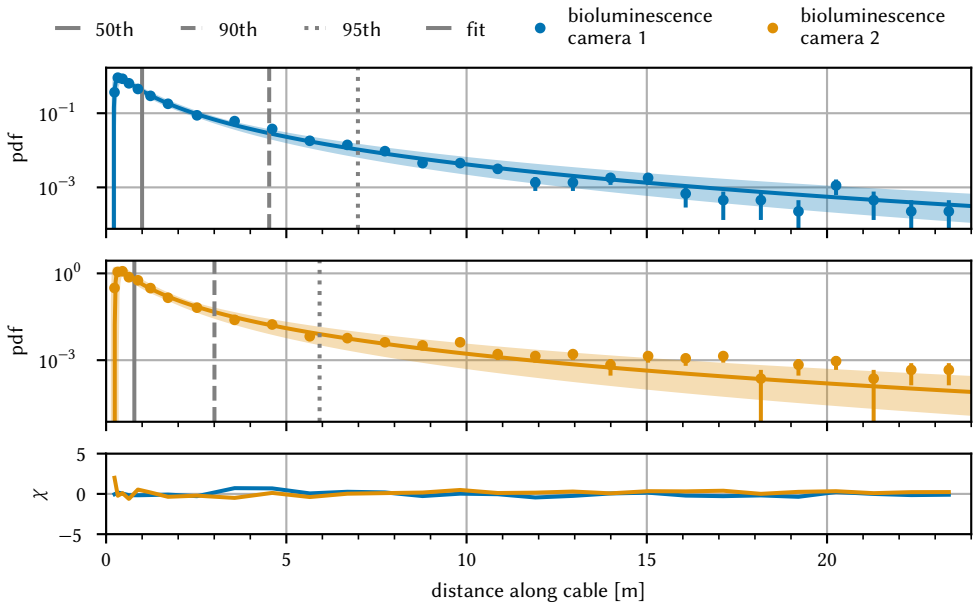

Figure 8.3: Camera's distance distribution of detected bioluminescence along the cable. The distance calculation use bioluminescence clusters close to the cable position and combines optical projection with the known cable position to access the missing 3D information from the 2D image. The data can be modeled with a log-normal distribution. The plots includes the 50th, 95th, and 99th percentiles.

8.2 Spatial Distribution of Bioluminescence Emissions in the Images

Bioluminescence is triggered mainly by mechanical forces, with only a minority occurring spontaneously [80–82]. Consequently, organisms emit light during collisions or turbulence encounters. In the context of a neutrino telescope like P-ONE, understanding how these emissions are distributed around an obstacle, such as the detector cable or module, is crucial, since bioluminescence emissions can create background noise for neutrino reconstruction. Cameras, such as the one installed in STRAW-b, can provide insights into this distribution, as they detect the direction of an emission. However, a single camera cannot fully measure the distribution, as the distance of an emission remains unresolved.

To obtain distance information from the camera image, the known position of

Table 8.2: Parameters of the fitted distance distribution of Equation 8.2. Figure 8.3 shows the distribution and the fits including χ^2.

Camera | Module	α	μ	σ
Camera 1 | PMT-Spectrometer 1	1.247±0.021	0.201±0.005	0.780±0.018
Camera 2 | PMT-Spectrometer 2	1.177±0.027	0.236±0.014	0.538±0.014

the cable relative to the camera can be used as a reference. The horizontal distance between the camera and the cable (295 mm) and the distance between two neighboring modules (24 m) are the two known positions. These positions are also identifiable in the image, as illustrated in Figure 6.1. To utilize the camera's optical projection (Chapter 5), a coordinate system is established along the cable with the z-axis and x-axis pointing towards the camera. The origin of this coordinate system is placed at the same height as the camera, so the camera's position can be represented as (0.295, 0, 0) [m], the neighboring module as (0.295, 0, 24) [m], and the cable between the two modules extending from (0, 0, 0) [m] to (0, 0, 24) [m]. To align the camera coordinate system with the outlined coordinate system, a translation of 0.295 m on the x-axis and a rotation matrix are required to be applied to the camera coordinate system. By establishing this coordinate transformation and mapping the pixel locations in the image to the corresponding distances, it is possible to extract distance information from the image along the cable.

To associate each bioluminescence emission with its closest point on the cable, a simple pixel-based geometry approach is used, without considering the optical projection. Each bioluminescence emission is assigned a distance value based on the closest point on the cable. However, this method is only suitable for organisms that emit light when triggered by the cable. If an organism's position in the image is significantly distant from the cable in space, this technique may yield inaccurate distance measurements. Nevertheless, the cable's position is clearly visible in the bioluminescence cluster's pixel positions (Figure 6.10 or Figure 6.14), making this method feasible for the majority of clusters that are located within a considerable distance from the cable.

As seen in Figure 8.3, the distributions of the calculated distances indicate that

both cameras follow a log-normal distribution

$$f_X = \frac{1}{(x-\mu)\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\log^2\left(\frac{x-\mu}{\alpha}\right)} \tag{8.2}$$

where μ is a location parameter, σ is a scale parameter and α is a shape parameter. Table 8.2 summarize the best fit parameters. The histogram considers clusters that are within a 50 pixel distance to the cable. Furthermore, the histogram shows at which distance the camera detects bioluminescence. Both cameras detect 50% of the clusters within the first meter and 90% below 5 m. There is no rise towards the 24 m distance where the neighboring module is located. Therefore, the camera is not sensitive enough for bioluminescence induced by the neighboring module. Indeed, the module should cause about the same bioluminescence rate as the 24 m string segment between the two modules. This is because the bioluminescence rate is proportional to the cross-section towards the water flow (Section 8.4.1) and the 24 m long cable with a diameter of 4 cm, including the steel line and data cable has an area of 0.1 m^2 which equals the area of the 13 inch module sphere.

Additionally, the distribution of bioluminescence emissions orthogonal to the cable can be analyzed. As previously discussed in Section 7.1, the modules are aligned with the current flow, resulting in the current's velocity having only an x-axis component based on the defined coordinate system. To extract 3D information, it is assumed that bioluminescence emissions are orthogonal to the current flow, i.e., along the y-axis. Therefore, the distance from the cable in pixels can be converted to a distance on the y-axis using optical projection. While this may not be an entirely accurate assumption, it can provide an estimate of the radius around an object orthogonal to the currents. The distribution for both cameras is shown in Figure 8.4. Around 50 % of the clusters are triggered within 10 cm of the cable, while 90 % are located within a radius of 1.5 m. These findings are consistent with the water flow simulation presented in Section 8.4. Camera 2 displays a flatter slope, which may be attributed to the higher expected currents for the upper camera, leading to an extended radius around the cable.

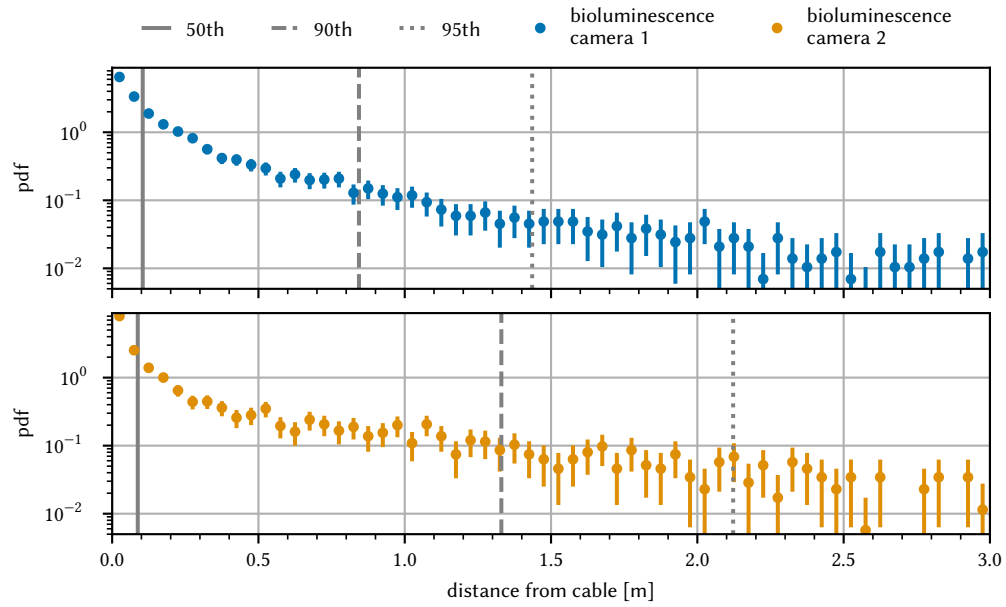

Figure 8.4: Orthogonal distribution of bioluminescence around the cable. As the text outlines, the distance is assumed to reflect the distance around the cable in an orthogonal direction to the currents.

8.3 Spectral Population of Bioluminescent Organisms

Bioluminescence emission is biased towards blue light due to its better propagation in water. However, the emission spectra can vary depending on the emitting species and the underlying biochemical reaction. For many organisms, the emission spectra are known and can be expressed as a spectral catalog of bioluminescence [80]. The spectral catalog is visually represented in Figure 8.5. Since the camera captures images in three different color channels (RGB), obtaining spectral information of an emission requires special treatment, similar to white balance in photography. This section provides an overview of the individual components involved in this spectral analysis before presenting the measured spectral population of bioluminescent organisms.

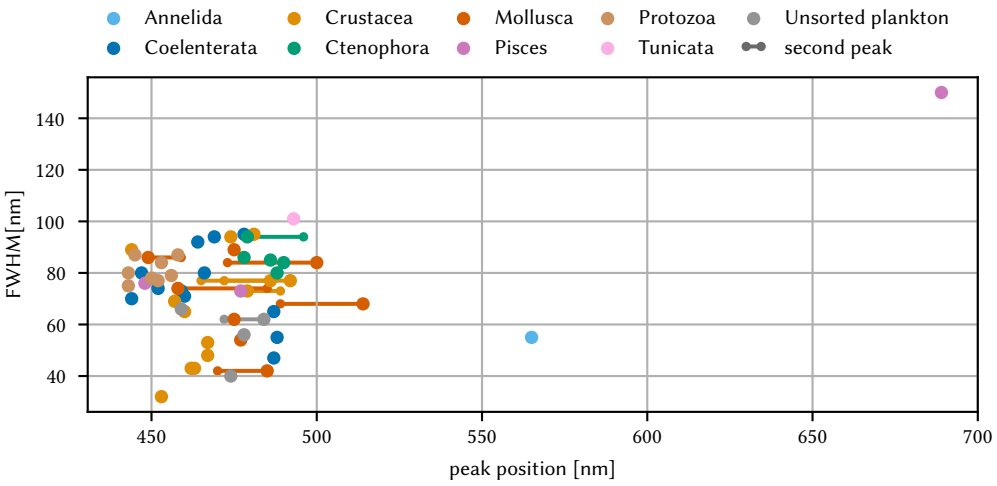

Figure 8.5: Spectral catalog of bioluminescence emissions. Each point represents a species, and the colors indicate their respective groups. Some species have a secondary peak in their spectra which is illustrated as a horizontal line connecting the mean and secondary peak. Data are from [80] and listed in Table D.1.

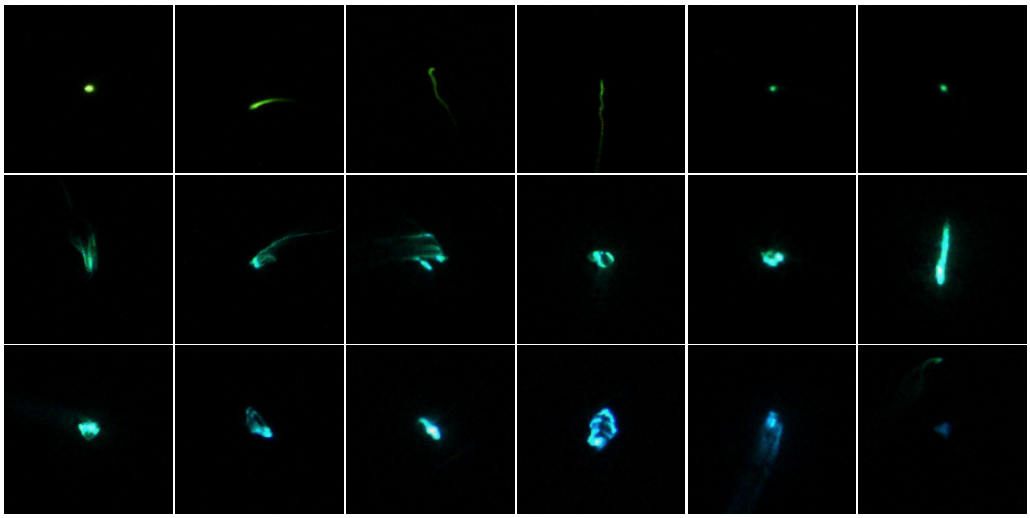

Figure 8.6: Hues of the deep-sea. A selection of bioluminescence emissions, sorted by their hue (top left to bottom right). All images have the same zoom level. Remarkable is the last image with two different emissions, including their hue. Hues from the first row predominantly appear in the lower module (PMT-Spectrometer 1) as shown in Figure 8.8.

8.3.1 Hue of the Detected Bioluminescence

A common technique for interpreting RGB color is transforming the values into a different color space, such as HSV (hue, saturation, and value). The hue component of HSV represents an angle on a color wheel, ranging from 0 to 360°, and can be calculated from the RGB values using the following formula:

$$\text{hue}(R,G,B) = 60^\circ \begin{cases} \mod\left(\frac{G-B}{\Delta}, 6\right) & \text{if } R \geq G \wedge R \geq B \\ 2 + \frac{G-B}{\Delta} & \text{if } G > R \wedge G \geq B \\ 4 + \frac{G-B}{\Delta} & \text{if } B > R \wedge B > G \end{cases} \tag{8.3}$$

where $\Delta = \max(R,G,B) - \min(R,G,B)$.

Each bioluminescence event could be mapped to a hue value using the RGB intensity parameters measured by the cluster detection. Figure 8.6 shows a selection of bioluminescence events sorted by hue, covering the full range of recorded hues. The hue distribution in both cameras is shown in Figure 8.8. It is noteworthy that both distributions show a similar prominent peak around 170° hue and that camera 1 shows a peak below 135° hue that camera 2 does not. However, it is essential to note that hues cannot be directly translated into a spectral signature.

8.3.2 Spectral Camera Response

The camera response must be considered to calculate the hue of an emission spectrum. If an emission has a specific spectrum $I(\lambda)$, the spectrum convolute with the acceptance spectra $a_c(\lambda)$ of the individual color channel c and the recorded intensity is

$$I_c = (I * a_c)(\lambda) = \int_{-\infty}^{\infty} I(\lambda)\, a_c(\lambda)\, d\lambda. \tag{8.4}$$

Hence, a hue value can be calculated for each emission spectra, but the inverse is not directly possible. Furthermore, the transmittance spectra of the water

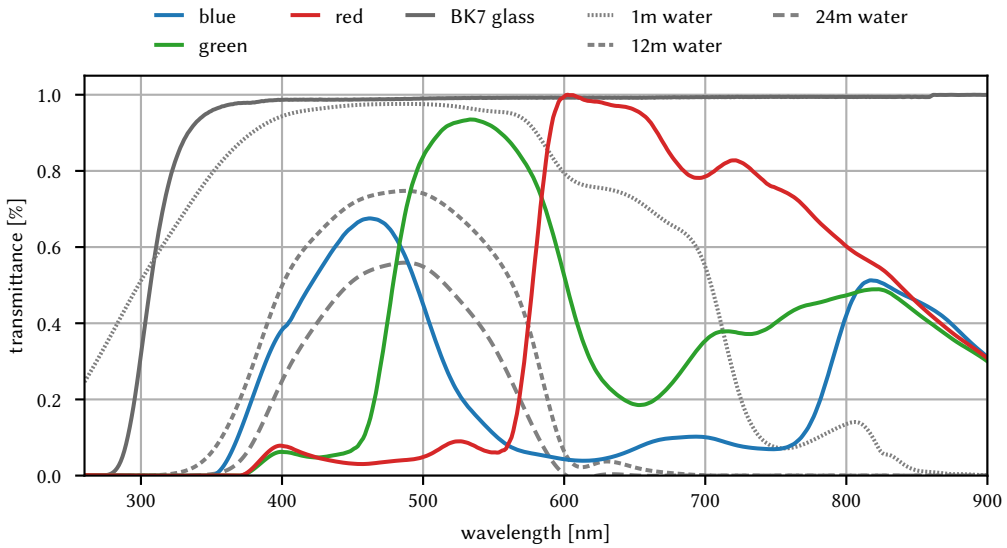

Figure 8.7: Summary of the transmittance distributions for the simulation. The water model $t_w(\lambda, r)$ uses the values reported from [83] as it is in good agreement with the attenuation measurement of STRAW but contains finer wavelength resolution [13]. Glass transmittance and camera acceptance base on the values listed in their datasheets [44, 54, 61].

$t_w(\lambda, r)$ and the glass housing $t_g(\lambda)$ need to be integrated with

$$\begin{aligned} I_c &= (I * a_c * t_w(r) * t_g)(\lambda) \\ &= \int_{-\infty}^{\infty} I(\lambda)\, a_c(\lambda)\, t_w(\lambda, r)\, t_g(\lambda)\, d\lambda \end{aligned} \tag{8.5}$$

where r is the distance between the bioluminescence emitter and the camera. The transmittance spectra of the different components are shown in Figure 8.7.

Fortunately, a variety of bioluminescence spectra $I(\lambda)$ are known by their peak position and width. In a simplified scenario, based on the bioluminescence catalog, an individual bioluminescence spectrum can be approximated with a Gaussian function of standard deviation 35 nm [80]. In addition, the camera acceptance curve can be obtained from its datasheet. However, the distance r between the emitter and the detector remains unknown, as no 3D information is available from a 2D image. As a result, the spectral analysis of a particular emission can

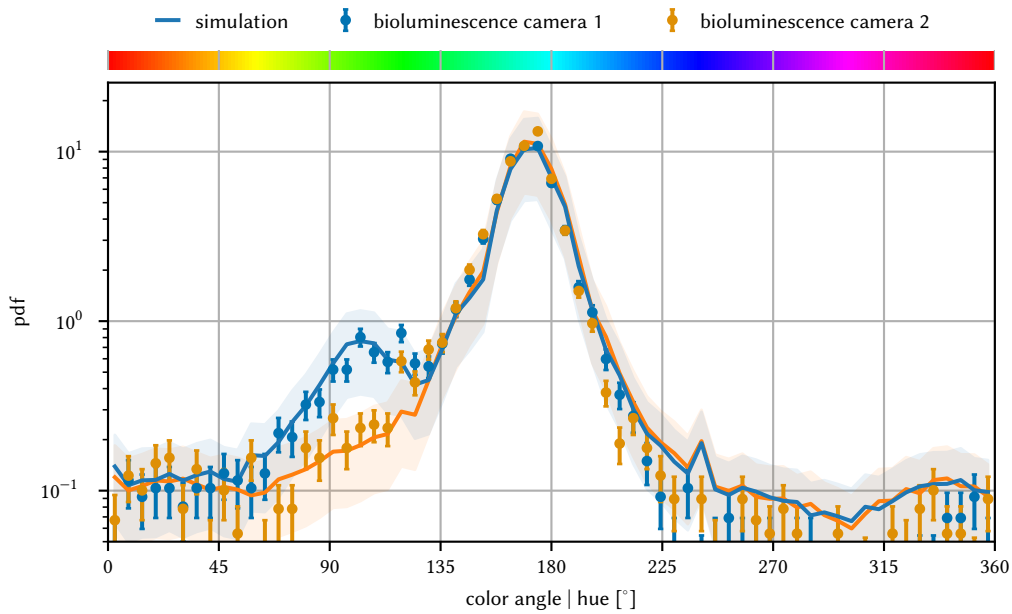

Figure 8.8: Comparison of the detected and simulated hue distribution. The data represents all detected bioluminescence in the period from 22.12.2021 - 31.12.2022. The simulation requires a spectral population of emitters which is shown in Figure 8.9. The color scale represents the monochromatic color for a specific hue angle as reference, but the hue angle gets ambiguous for non-monochromatic spectra, as outlined in the text.

only provide limited information, as the water absorption properties are highly dependent on the color, which can lead to a shift in the peak wavelength of the spectrum depending on the distance between the emitter and the detector.

8.3.3 Modeling the Spectral Population of Bioluminescent Organisms

By analyzing a population of bioluminescent emitters (spectral population), determining the distance for a single bioluminescence event, as in Section 8.3.2, is not necessary. Instead, the distance can be described by a distribution. A set of emitters is simulated where the peak wavelength (μ) for each emitter is drawn from the spectral population, and the distance (r) is drawn from the distance distribution measured in Figure 8.3. The resulting hue distribution is then compared to the hue distribution from all images, as shown in Figure 8.8.

However, the simulation requires some more components to be accurate. The water model, $t_w(\lambda, r)$, uses values reported by [83], which are in good agreement with the attenuation measurement of Strings for Absorption Length in Water (STRAW) and contain finer wavelength resolution [13]. The transmittance distributions for all components listed in Equation 8.5 are summarized in Figure 8.7. The other two components are modeled with data taken from their datasheets [44, 54, 61]. A camera model then adds normally distributed noise to the simulated intensity readings. Finally, a camera DAQ model applies upper and lower intensity thresholds to integrate saturation and its opposite. The range between the upper $th_{\max}$ and lower $th_{\min}$ intensity thresholds is 16 bit, hence $th_{\max} = 2^{16}\, th_{\min}$.

Matching the simulation to the data was done manually by comparing different configurations per iteration. Naive minimizers did not perform well, mainly because generating a single hue probability density function required several seconds with the current simulation. Before integrating a more sophisticated minimizer, the simulation needs speed improvements, which are beyond the scope of this thesis. Therefore, the reported values are a rough estimate.

To accurately reproduce the measured hue distribution(Figure 8.8), the simulation needs a distribution for the emitters population, referred to as the spectral population hereafter. Each emitter in the spectral population is simulated with a Gaussian function centered at its peak wavelength and a full width at half maximum (FWHM) of 82 nm (σ=FWHM/$\sqrt{8\ln(2)}$=35 nm), which agrees well with the values reported in Figure 8.5. The spectral population, H, is modeled as a sum of Gaussian distributions

$$\begin{aligned} H(\lambda \,|\, \vec{\mu}, \vec{\sigma}, \vec{s}\,) &= \sum_i s_i \, \mathcal{N}(\lambda \,|\, \mu_i,\, \sigma_i^2) \\ &= \sum_i \frac{s_i}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\lambda_i - \mu_i}{\sigma_i}\right)^2} \end{aligned} \tag{8.6}$$

where $\vec{mu}$, $\vec{\sigma}$, and $\vec{s}$ define the peak positions, width, and scale (amplitude). Moreover, the camera noise model must be incorporated, whereby the added noise in green and red colors is 20% higher than in blue. Thermal effects may be the reason for this behavior; also, the camera sensor tries to match the accep-

Table 8.3: Parameter of 12 Gaussians to model the spectral population. The simulation uses the spectral population to describe the measured hue distribution. The scale_1 and scale_2 values, measured in permille, represent the scaling factors for the two cameras. For completeness, the Gaussians at 689 nm and 443 nm are listed even, so they do not appear in the final population (scale=0 ‰). Figure 8.9 illustrates the resulting spectral population probability density function (PDF).

μ [nm]	689	565	514	500	492	486	482	478	471	466	458	443
σ [nm]	–	14	7	7	4	3	3	3	3	3	3	–
scale_1 [‰]	0	69	27	27	275	302	165	110	14	8	3	0
scale_2 [‰]	0	0	29	29	295	324	177	118	15	9	3	0

tance in the infrared (Figure 8.7).

Figure 8.9 illustrates the resulting spectral population to reproduce the detected hue distribution (Figure 8.8). Furthermore, it includes a representation of the bioluminescence catalog (Table D.1 [80]). Surprisingly, the spectral populations align perfectly with the bioluminescence catalog. The primary hue peak at 170° in both cameras is expressed by the same sub-part of the spectral populations. The minor hue peak, which only appears in Camera 1 (located closer to the ground), requires a population around 565 nm and a standard deviation of 14 nm. The publication of the bioluminescence catalog reports a standard deviation of 14 nm for some organisms; also, it remains unclear to which organism this parameter refers [80, 84].

8.3.4 Spectral Classification of Emitting Organisms

Assuming that the bioluminescence catalog covers all the species present in the Cascadia Basin, the peak at 565 nm can be attributed to *Tomopteris nisseni*. This species belongs to the segmented worm group (*Annelida*) and is reported to be planktonic, indicating that it cannot move against water currents [80, 84]. Its size ranges from 3 cm–40 cm, and it weighs between 0.5 g–50 g [84]. Figure 8.10 displays an image of *Tomopteris nisseni*. Some of the yellowish images shown in Figure 8.6 reveal a structure that matches that of a worm. Therefore, all evidence suggests the presence of *Tomopteris nisseni*. It is also notable that this species only appears in the lower camera, with a rate of 69‰ among all bio-

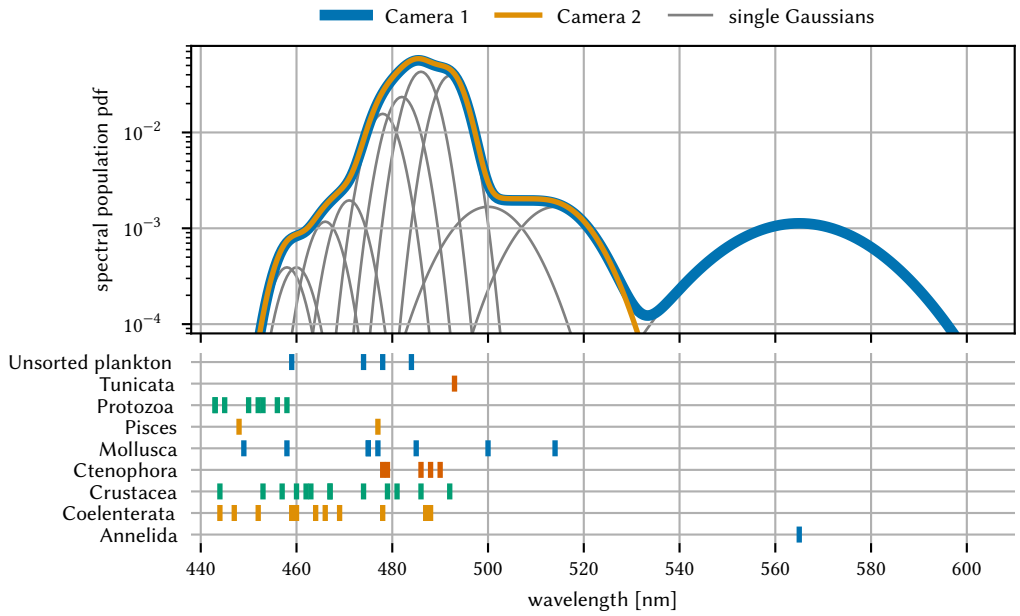

Figure 8.9: Spectral bioluminescence population t reproduce the hues distribution measured with the cameras in Figure 8.8. The distribution describes the population of Gaussian emitters, each with a FWHM of 82 nm. Table 8.3 provides the parameters modeling the spectral population. The lower plot shows the peak positions of listed organisms in the bioluminescence catalog summarized in Table D.1 [80].

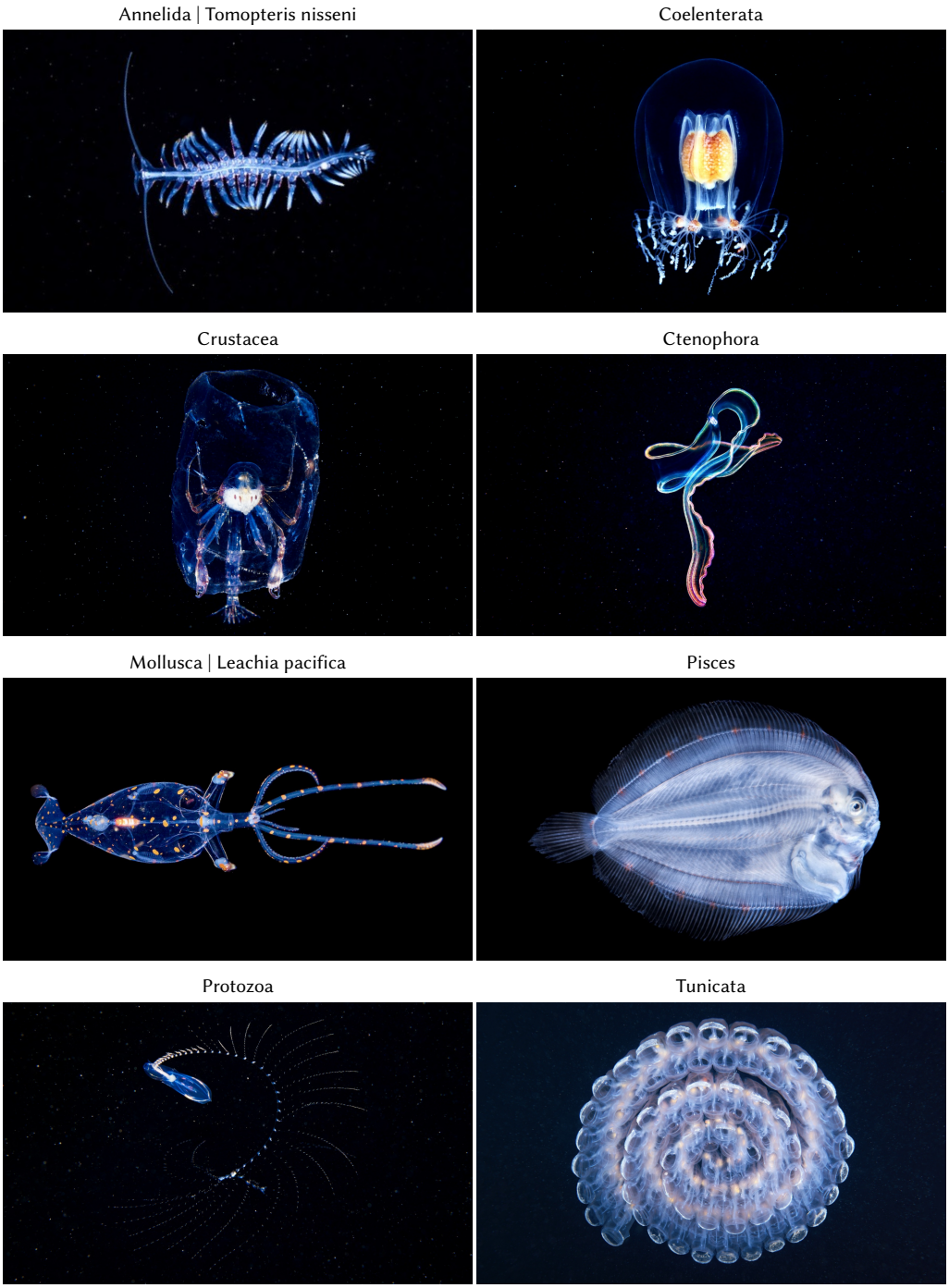

Figure 8.10: Images of species with their biological classifications. The listed classifications are all part of the bioluminescence catalog Table D.1 [80]. *Tomopteris nisseni* and *Leachia pacifica* are the two species detected by the spectral analysis outlined in Section 8.3. Images by courtesy of [85–87].

luminescence. Organism classification based on remotely operated underwater vehicle (ROV) videos from dives close to San Francisco indicates that *Tomopteris nisseni* is present in the lowest 200 m layer from the ground [88]. However, this data is not represented in a distance-to-ground distribution.

Following the same principle, Figure 8.9 reveals the presence of two different emissions from the *Leachia pacifica*. *Leachia pacifica* is a genus of glass squids (*Mollusca*) and the species *Leachia pacifica* is about 15 mm–25 mm long. Females of this species develop light organs on the ends of their third arm pairs as they mature, which are thought to be used in mating displays to attract males [89, 90]. The two species emitting at 514 nm and 500 nm have an equal fraction of the total spectral population of 29‰ and 27‰ in the cameras 1 and 2, respectively.

Other species can not be explicitly identified, as various species emit light below 500 nm. It is likely that multiple organisms contribute to the prominent peak in the spectral population at 470 nm–500 nm. However, the spectral population reveals that there are very few organisms emitting light below 450 nm, despite the camera still being sensitive in this range, and the water being more transparent compared to the 565 nm emitted by *Tomopteris nisseni*.

Based on previous mentioned dives off the coast of San Francisco, the prominent peak could be explained by the presence of *Ctenophora* organisms (also known as comb jellies). Reports suggest that *Ctenophora* and *Annelida* [1] are the main contributors to bioluminescence in the deep sea. The emission wavelength of *Ctenophora* organisms aligns with the prominent peak, while *Annelida* emits light at a different wavelength. This suggests that *Ctenophora* are likely responsible for the majority of bioluminescence emissions around the modules [88].

In conclusion, this section demonstrates that spectral analysis based on camera images can reveal the spectral population of emitting organisms. In some cases, single species can be identified by matching the spectral population to a bioluminescence catalog that reports only one species for a specific wavelength range. However, it is important to note that the identification of species depends on the completeness of the bioluminescence catalog.

[1] *Annelida* is refereed to its class *Polychaeta* in [88]

8.4 Bioluminescence Triggers

Bioluminescence is predominantly triggered by mechanical forces, with only a small proportion occurring spontaneously [80–82]. Mechanical forces are caused by the flow of water around objects, including the components of the detector. Understanding the relationship between water flow speed and the emission rate is crucial for predicting the probability of bioluminescence emissions in future neutrino telescopes like P-ONE.

This section will cover the mathematical dependency between the two variables and provide a hydrodynamic description. The flow of water around objects can create vortex street oscillations, which can even be measured by the magnetic compass of the modules. Finally, the section will present the distribution between the flow speed and emission rate, from which the concentration of bioluminescent organisms can be estimated.

8.4.1 Bioluminescence Emission Rate and Water Flow

In principle, there are two sources, which are contact forces when the organisms collide with an object and shear forces from the induced turbulences around an object. Spontaneous emissions are excluded in the following due to the rare probability [80–82]. Therefore the bioluminescence emission rate induced around an object is

$$R = R_{\text{contact}} + R_{\text{shear}}. \tag{8.7}$$

The contact emission rate

$$R_{\text{contact}} = c_{\text{bio}}\, A_{\perp}\, p_{\text{contact}}\, |\vec{u}| \tag{8.8}$$

can be expressed by the cross-section of the object $A_{\perp}$ which measures orthogonal to the flow velocity $\vec{v}$ and the concentration of bioluminescent organisms c_{bio} in the water. p_{contact} denotes the emission probability and is approximately constant [91]. The emission probability due to shear stress is more complex as it is linearly connected to the water current's gradient with

$$p_{\text{shear}}\left(\vec{u}\right) \sim \alpha\, \nabla\vec{u} \tag{8.9}$$

where α is a proportionality factor [92–94]. In order to calculate the emission rate, the volume V around the object that gets disturbed must be considered with

$$R_{\text{shear}} = c_{\text{bio}}\,\alpha \iiint_V \nabla\vec{u}\,dV \tag{8.10}$$

where α is chosen that c_{bio} is identical to Equation 8.8. By assuming an organism emits only once, each emission can be seen as a divergence, and the volume integral can be expressed by the divergence theorem to

$$R_{\text{shear}} = c_{\text{bio}}\,\alpha \oiint_S \vec{u}\,\hat{n}\,dS. \tag{8.11}$$

with the closed surface S around the volume V and the outward pointing unit normal $\hat{n}$ at each point on the surface. This can be further simplified by choosing a box around the object that is big enough to cover the whole flow disturbance by the module. Furthermore, the box should align with the current so that two opposite surfaces are orthogonal to the current and all other surfaces are parallel to the current. The contribution of the parallel surfaces to the integral is 0 as the $\hat{n}$ is orthogonal to $\vec{u}$ by the definition of the surfaces. Hence, the rate of emissions is proportional to the bioluminescent organisms flow through the surface facing the current

$$R_{\text{shear}} = c_{\text{bio}}\,\alpha\,A_{\perp}^{*}\left(\vec{u}\right)\,|\vec{u}|. \tag{8.12}$$

The surface is the effective cross-section covering the entire flow disturbance downstream. Therefore, the surface size depends on the flow vector $\vec{u}$.

As shown in the following Section 8.4.2, $A_{\perp}^{*}$ can be assumed constant for the expected flow speeds. Hence, the rate of bioluminescence emissions is proportional to the flow speed and the concentration of bioluminescent organisms with

$$\begin{aligned} R &= c_{\text{bio}}\,A_{\perp}\,p_{\text{contact}}\,|\vec{u}| + c_{\text{bio}}\,\alpha\,A_{\perp}^{*}\,|\vec{u}| \\ &= c_{\text{bio}}\,\left(A_{\perp}p_{\text{contact}} + \alpha\,A_{\perp}^{*}\right)\,|\vec{u}|. \end{aligned} \tag{8.13}$$

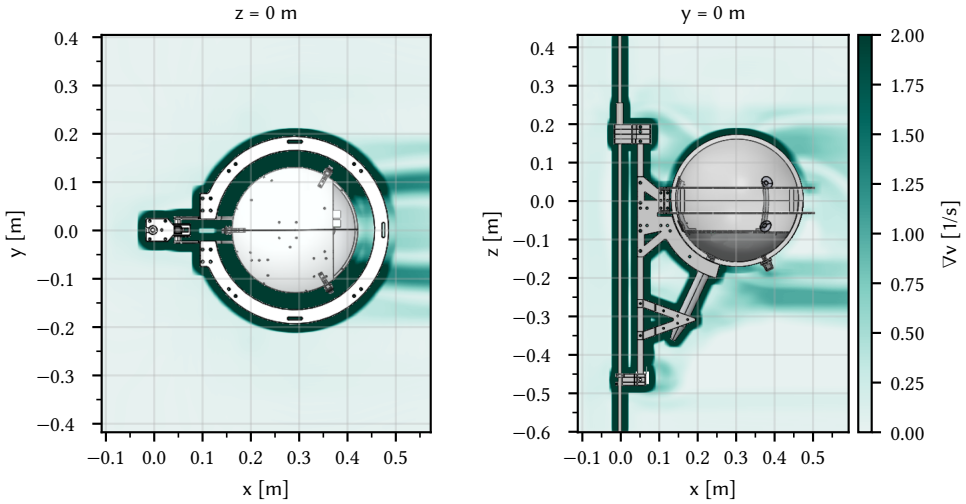

Figure 8.11: Water flow gradient around module and cable for a velocity of 10 $\frac{\text{cm}}{\text{s}}$. The underlying simulations were performed in Solidworks by Christian Spannfellner. Other simulation visualizations are listed in Section E.1.

8.4.2 Flow Around the Detector

> *Although in most ways the exact manner in which water moves is difficult to perceive and still more difficult to define, as are the forces attending such motion, certain general features both of the forces and motions stand prominently forth, as if to invite or defy theoretical treatment.*
>
> Osborn Reynolds, 1883 [91]

In order to derive the vector filed $\vec{u}$, water current simulations are required. Typically, Navier-Stokes equations are used in simulations. Because $0 = \nabla \cdot \vec{u}$ holds for an incompressible medium, the incompressible Navier-Stokes equation is

$$0 = \frac{\partial u}{\partial t} + \left(\vec{u} \cdot \nabla\right) \vec{u} + \nabla^2 \vec{u} - \nu \nabla \vec{u} + \nabla \frac{P}{\rho} \tag{8.14}$$

where ρ is the density, P the pressure, and ν is the kinematic viscosity. The kinematic viscosity connects to the viscosity μ with $\nu = \mu/\rho$ [95]. Figure 8.11 shows a simulation of a STRAW-b module and the mooring cable for a water

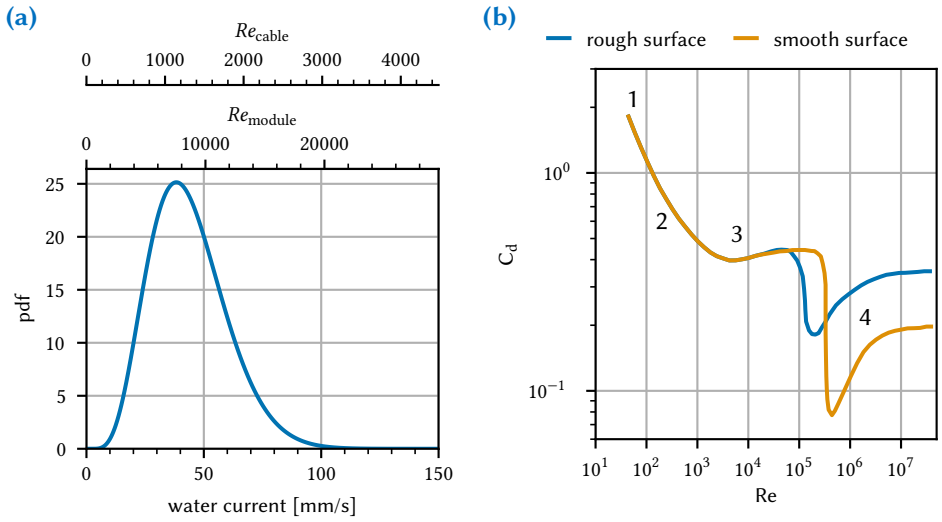

Figure 8.12: Reynolds number distribution for the module and cable (a) and drag coefficient of a spheres with different surfaces (b). The numbers along the drag coefficient indicate several flow regimes. While the Reynolds number increases, the attached laminar flow (1.) starts to separate downstream, producing a vortex street (2.). The vortex street transitions to a chaotic, turbulent wake (3.), and beyond the post-critical flow, the boundary layer is fully turbulent (4.) [96].

flow of 10 $\frac{\text{cm}}{\text{s}}$. Running simulations is computationally intensive and requires an accurate description of an object, including the surfaces exposed to the flow, and is difficult to calibrate to reality, especially for remote objects.

A common way to compare and predict fluid flow patterns besides simulations, is the Reynolds number

$$Re = \frac{|\vec{u}|L}{\nu}. \tag{8.15}$$

where ν is the previously mentioned kinematic viscosity and L is a characteristic linear dimension. Water has a kinematic viscosity of ν=1.6737 $\times$ 10^{-6} $\frac{\text{m}^2}{\text{s}}$ with a density of 999.9 $\frac{\text{kg}}{\text{m}^3}$ at the water temperature of 2° in Cascadia Basin [97].

To estimate the upper limit of the expected Reynolds numbers, the following overvalues the different parameters. The flow can be up to 20 $\frac{\text{cm}}{\text{s}}$. The characteristic linear dimension is chosen to be the diameter of the 13” sphere and 5 cm cable. Hence, Reynolds numbers can increase up to 10^5 and 10^4 for the cable and module, respectively. Figure 8.12 shows the expected flow distribution

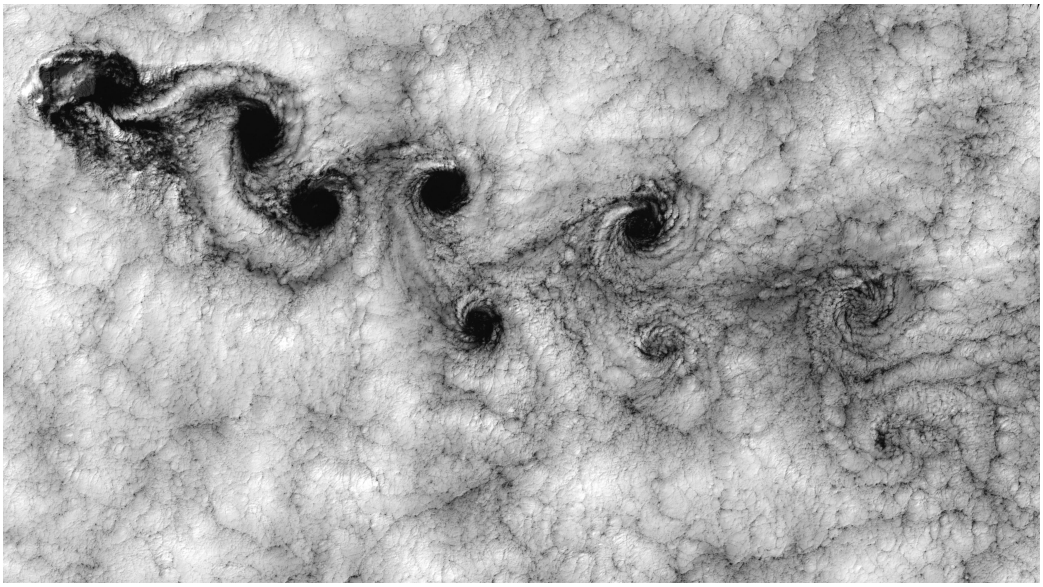

Figure 8.13: Kármán vortex street in the wake of Juan Fernández Islands off the Chilean coast [100]. To guarantee the stability of the vortices street, neighboring swirls rotate in opposite directions with a fixed ratio of cross-street to along-street distance between vortices [99]. The ratio depends on the object's shape but is independent of the flow speed.

and its translation to Reynolds numbers. In this range, the flow around an object is laminar upstream and starts to separate and gets turbulent at around 80° from the front stagnation point - the point which faces the flow [96, 98]. In the wake, a so-called Kármán vortex street (Figure 8.13) develops, which gradually decreases in regularity for higher Reynolds numbers as the wake gets turbulent [96, 98]. The vortex street is a repeating pattern of swirling vortices. Neighboring swirls rotate in opposite directions with a fixed ratio of cross-street to along-street distance between vortices. The ratio or angle depends on the size but is necessary to guarantee the stability of the vortices street [99]. Hence, $A_{\perp}^{*}$ can be assumed constant, and the rate of bioluminescence emissions is proportional to the flow speed and the concentration of bioluminescent organisms as summarized in Section 8.4.1.

8.4.3 Induced Oscillations from Kármán Vortex Streets

As previously mentioned, the flow around the module and cable is in the range where Kármán vortex streets appear in the wake of the objects. Figure 8.13 shows a Kármán vortex streets in clouds downstream of Juan Fernández Islands. Vortices generate with a frequency defined by the Strouhal number (St) and

$$f = \frac{St\,|\vec{u}|}{L}. \tag{8.16}$$

In the range of Reynolds numbers from 250 to 2×10^5, the Strouhal number measures

$$St \approx 0.21 \left(1 - \frac{19.7}{Re}\right) \tag{8.17}$$

for a sphere [101] and the frequency is

$$\begin{aligned} f &= 0.21 \left(\frac{|\vec{u}|}{L} - \frac{19.7\,\nu}{L^2}\right) \\ &= \frac{0.21}{L}\,|\vec{u}| - \frac{4.137\,\nu}{L^2}. \end{aligned} \tag{8.18}$$

As the vortices couple to the object causing the vortex street, the object starts to oscillate. This provides a unique feature to verify the previously made flow assumptions around the cable and module based on Reynolds numbers. The magnetic sensor, which provides a heading every 10 s, measures the module's oscillations. The water current data origins from a current meter measuring close to the ground and located ~2.5 km northeast of the module. The magnetic heading aligns well with the current data as shown in Figure E.5 for an exemplary period.

By filtering periods of constant water flow in speed and direction, the frequency space of the heading data reveals a prominent oscillation peak (8.14a) that disappears for periods with changing currents. The amplitude of the strongest oscillations is approximately 5°, as shown in Figure E.6, while the average is around 1.5°. The correlation between vortex street frequency and flow speed is shown in Figure 8.14, where the best-fit parameters of the slope (i.e. $0.21/L$) and offset (i.e. $4.137, \nu/L^2$) are $(0.59 \pm 0.10)\,\frac{1}{\mathrm{m}}$ and $(-0.02 \pm 0.01)\,\frac{1}{\mathrm{s}}$, respectively. This result is in good agreement with the expected values for a module of diameter 13

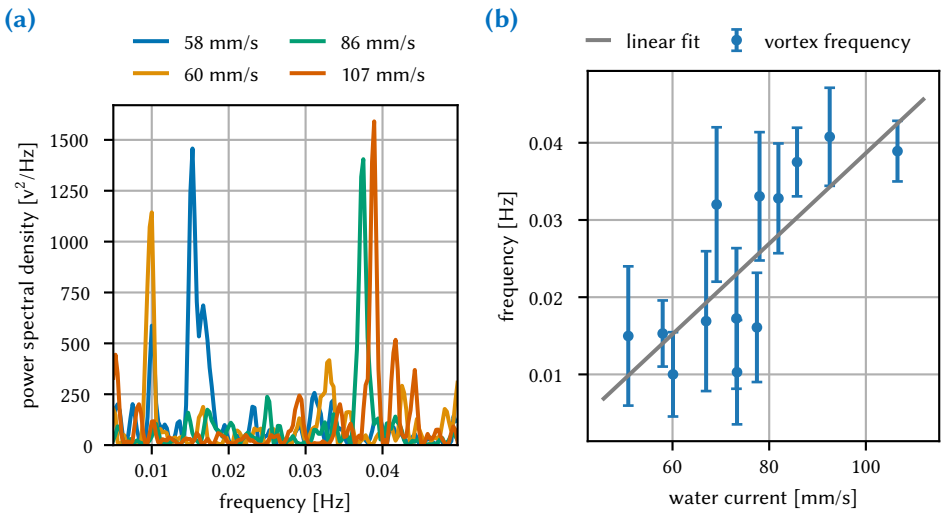

Figure 8.14: Kármán vortex street oscillations of the module. The oscillations are measured with the magnetic heading sensor of the module, and the water current data origins from a currentmeter measuring the flow close to the ground and located ~2.5 km northeast of the module. During periods of stable currents, the vortex street induces oscillations of the module's heading (a). The linear dependency of oscillations frequency and current speed (b) aligns linearly as expected from theory.

inches, where equation 8.18 yields a slope of $0.64\,\frac{1}{\mathrm{m}}$ and an offset of $6 \times 10^{-5}\,\frac{1}{\mathrm{s}}$. However, the measured oscillation frequency is lower than predicted by equation 8.18, which can be explained by the fact that the module rotates around the cable, adding momentum and allowing it to rotate in the direction of the new generating vortex, thereby damping the initialization of the next vortex on the opposite side. In other words, the object has a larger effective size than its physical dimensions with respect to Equation 8.18. Additionally, the tension in the string has a damping effect on the oscillation.

8.4.4 Concentration of Emitting Organisms

As indicated in Section 8.4.1, the relationship between the expected bioluminescence emission rate and the current speed at Cascadia Basin is linear. To establish this relationship in the data, it is necessary to establish a connection between the measured current speeds and the detected bioluminescence in the captured images. The current speeds at Cascadia Basin are recorded using a cur-

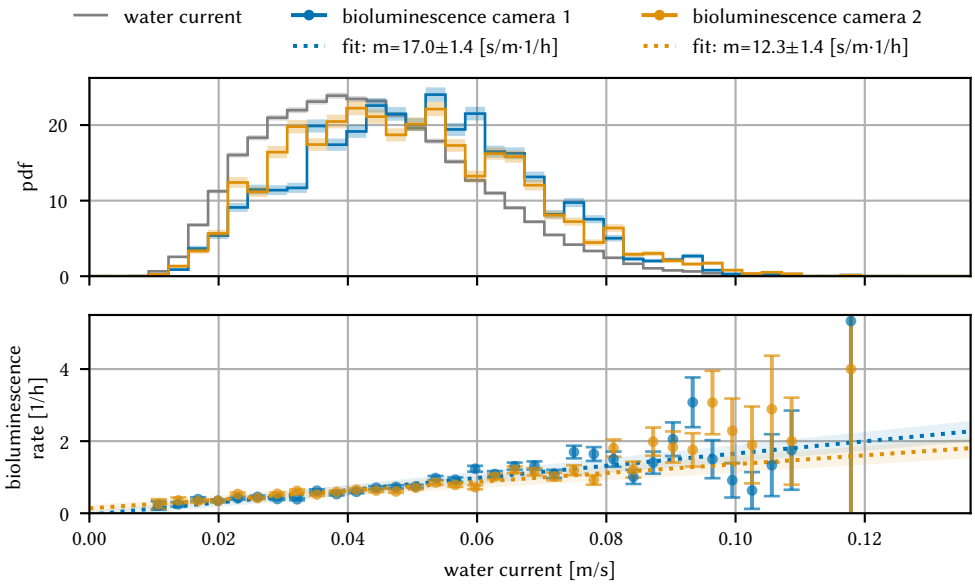

Figure 8.15: The upper plot displays the PDFs of the measured flow speeds $f_V(|\vec{u}|)$ (shown in gray) and the flow speeds during which bioluminescence occurrences took place $f_{V\cap B}(|\vec{u}|)$. These PDFs are utilized to establish the relationship between current speed and bioluminescence emission rate, as described in the text. The lower plot demonstrates the expected linear dependence between the emission rate and flow speed, which was derived in Section 8.4.1.

rentmeter positioned 2.1 km northwest of STRAW-b, as indicated in Figure 2.9. The process of detecting bioluminescence in the images is described in Chapter 6.

The bioluminescence emission rate can be expressed as

$$R(t) = \frac{d}{dt} n(t) \tag{8.19}$$

where n represents the total number of emissions at time t. The time can then be used to establish the link between the emission rate and current speeds. However, directly accessing the bioluminescence emission rate at a specific moment in time poses challenges due to the low frequency of images displaying bioluminescence, which is approximately $1\,\frac{1}{\mathrm{h}}$ as shown in Figure 8.1. Additionally, the current speed undergoes continuous changes, as depicted in Figure 8.2 and Figure E.5. To overcome this limitations, the emissions rate can be calculated

over all periods at a specific current speed. Let T be the total duration, N is total number of emissions and $|\vec{u}|$ represent a specific current speed. The total duration at a specific current speed can be calculated with

$$t(|\vec{u}|) = T\, f_V(|\vec{u}|) \tag{8.20}$$

where f_V is the PDF of the current speeds. Similarly, the total number of emissions during periods with a specific current speed can be expressed as

$$n(|\vec{u}|) = N\, f_{V\cap B}(|\vec{u}|) \tag{8.21}$$

where $f_{V\cap B}(|\vec{u}|, B)$ is the PDF of current speeds during which the condition B of an emission occurrence holds. The emission rate can then be calculated as follows:

$$R(|\vec{u}|) = \frac{n(|\vec{u}|)}{t(|\vec{u}|)} \tag{8.22}$$

$$= \frac{N}{T}\frac{f_{V\cap B}(|\vec{u}|)}{f_V(|\vec{u}|)} \tag{8.23}$$

$$= \frac{N}{T} f_{B|V}(|\vec{u}|) \tag{8.24}$$

where $f_{B|V}(|\vec{u}|)$ represents the conditional density of an emission occurrence given a specific current speed [102]. The probability density functions $f_V(|\vec{u}|)$ and $f_{V\cap B}(|\vec{u}|)$ are displayed in the upper plot of Figure 8.15, while the lower plot illustrates the emission rate $R(|\vec{u}|)$. Linear fits of the calculated emission rate

$$R(|\vec{u}|) = m|\vec{u}| + R_0 \tag{8.25}$$

result in best-fit parameters, which are summarized in Table 8.4. Using Equation 8.13, the slope translates to the concentration of emitting organisms with

$$c_{\text{bio}} \approx \frac{m}{A^*_\perp}. \tag{8.26}$$

The offset parameter R_0 is not considered, as it is close to 0 for both cameras. The effective area ($A^*_\perp$) is here a combination of field of view (FoV) and the effective area of the flow used in Equation 8.13. This effective area can be calculated

Table 8.4: Parameters to obtain the concentration of emitting organisms. The presented errors reflect the statistical error without the systematic error, which are discussed in Section 8.4.4.

	$m\ [\frac{\mathrm{s}}{\mathrm{m\,h}}]$	$R_0\ [\frac{1}{\mathrm{s}}]$	y^* [m]	z^* [m]	$A^*_\perp\ [\frac{1}{\mathrm{m}^2}]$	$c_{\mathrm{bio}}\ [\frac{1}{\mathrm{m}^3}]$
Camera 1	17.0±1.4	0.0±0.1	0.154	0.982	0.151	0.031±0.003
Camera 2	12.3±1.4	0.1±0.1	0.124	0.794	0.096	0.035±0.004

from the bioluminescence distributions along (Figure 8.3) and orthogonal to the line (Figure 8.4). The orthogonal distributions can be directly used, as the module aligns with the currents, and therefore the distributions show the required distance orthogonal to the currents. Each distribution has an effective distance defined by

$$d^* = 1/\max(f_X) \tag{8.27}$$

where f_X is the PDF. The effective area is the product of both effective distances.

The calculation lacks the emission probabilities for contact (p_{contact}) and shear (α) from Equation 8.13, which cannot be measured with the existing equipment in situ. Incorporating laboratory measurement values is also not preferable due to differences in the environments, resulting in more inaccuracy and reduced comparability to other measurements. Without considering the emission probability, the presented concentration can be expressed as an adequate concentration of emitting organisms that includes those probabilities.

Apart from emission probabilities, other uncertainties contribute to this measurement, which can only be roughly estimated. For instance, the effective area depends on the 3D position information, but the camera provides only 2D information, as explained in Section 8.2, which particularly affects the orthogonal distance distribution. Furthermore, the vertical distance between the currentmeter and the module adds uncertainty. As flow speed is measured close to the sea floor, the flow should be increased at the module. Ocean Networks Canada (ONC) also provides an acoustic Doppler current profiler (ADCP) that profiles the flow speed up to 500 m from the ground. However, the ADCP data has much noise due to the measurement principle. Therefore, the ADCP data are

not used to decrease the complexity of the analysis. Lastly, the cluster detection requires a minimum intensity underestimating the actual event rate.

In conclusion, the systematic error in the concentration is substantial. Assuming a 100 % systematic error on the effective area, which includes the uncertainty in the bioluminescence cluster detection because it is linearly connected, and a 20 % systematic error on the current speed, the overall systematic error accumulates to 120 %. Table 8.4 presents the parameters and the resulting concentration of emitting organisms with their statistical error. Both cameras yield a similar concentration of $(0.03 \pm 0.03)\,\frac{1}{\mathrm{m}^3}$, including the systematic error, which agrees with the range of $0.622\,\frac{1}{\mathrm{m}^3}$–$0.018\,\frac{1}{\mathrm{m}^3}$ measured in ANTARES [103].

9 Conclusion

This thesis offers a comprehensive study of deep-sea bioluminescence using cameras, focusing specifically on its relevance to the future Pacific Ocean Neutrino Experiment (P-ONE) located in the Cascadia Basin off the west coast of Vancouver Island, BC. The thesis encompasses a broad range of topics, including a detailed description of hardware components, the diverse aspects of the data acquisition (DAQ) system, bioluminescence detection in images, and various analyses of bioluminescence properties.

Despite the extreme depths of over 1.5 kilometers in the ocean abyss, the future neutrino telescopes P-ONE encounter life through bioluminescence. Bioluminescence is the emission of light by living organisms, including bacteria, plankton, and fish, found in the deep sea. Although bioluminescence presents a challenge for neutrino telescopes, requiring an understanding of its physical and biological processes and advanced modeling techniques to account for its effects, it also provides a unique opportunity for biologists to study the vast and unexplored habitat of the deep sea, which is the largest on Earth.

At the prospective location of P-ONE, two pathfinder missions, Strings for Absorption Length in Water (STRAW) and STRAW-b (Strings for Absorption Length in Water b), were carried out to examine the environmental conditions and assess the potential for a large-scale neutrino detector. STRAW focused on measuring water's optical properties in the Cascadia Basin, while STRAW-b employed cameras and spectrometers to study bioluminescence in the deep sea. The obtained results support the feasibility of a long-term neutrino telescope based at the Cascadia Basin.

The initial part of this thesis centers around the crucial elements of the STRAW-b pathfinder mission, encompassing low-level module operations, the DAQ system, and hardware components. Notably, particular attention is dedicated to the integration of the camera and flasher system within specific STRAW-b modules. Subsequently, the thesis delves into camera optics, presenting an analytically derived model that incorporates lens projection and the distortion of the spherical pressure housing.

The subsequent section of this thesis focuses on image processing techniques employed to detect bioluminescence emissions in the images. The detection algorithm follows a two-folded approach. Firstly, clusters of bright pixels are formed and parameterized, taking into account the characterization of pixel noise. This characterization enables the establishment of an intensity threshold for each pixel, determining its classification as *bright*. Secondly, a novel machine learning (ML) technique is developed to filter out clusters originating from bioluminescence, distinguishing them from non-bioluminescent clusters. The neural network (NN) is trained using regions in the images where high bioluminescence density is visually apparent. Ultimately, the trained NN accurately identifies bioluminescence throughout the entire image. Results demonstrate the exceptional performance of the NN classification, which only requires a minimum cluster size of five pixels for bioluminescence filtering. This capability enables a multifaceted analysis of the bioluminescence phenomenon, summarized in the following.

The utilization of the flasher system allows the camera to track the position of neighboring modules. The study demonstrates that the neighboring modules align synchronously with the currents without any observed rotation between them. Additionally, the relative module position has a vertical displacement of approximately 20 cm. This information is crucial as it simplifies the cluster detection process, as the training of the ML algorithm relies on the assumption of a constant mooring line and module positions.

The camera enables an indirect assessment of module contamination caused by biofouling or sedimentation. This measurement relies on monitoring the decrease in brightness of bioluminescence over time. The results demonstrate a

brightness annual loss of $(34.0 \pm 24.5)\,\%$ and $(34.3 \pm 19.9)\,\%$ for the two cameras. Visual inspection of the STRAW by a remotely operated underwater vehicle (ROV) suggests that biofouling is the primary contributing factor, although further analysis is required to validate this observation.

Moreover, the cameras measure bioluminescence rates of approximately $1\,\frac{1}{\mathrm{h}}$. Analysis of the periodic variations in bioluminescence rates reveals a strong correlation with water currents, where the dominant periods of both variables align. These periods can be attributed to the astronomical influences of the Earth, moon, and sun, which drive the ocean tides.

Additionally, it is observed that bioluminescence emissions exhibit a high concentration around the mooring cable of the detector. Approximately $50\,\%$ of the emissions occur within a $10\,\mathrm{cm}$ radius around the cable, while $90\,\%$ are concentrated within a $1.5\,\mathrm{m}$ radius. This suggests that bioluminescence is predominantly triggered by collisions or shear stress.

The camera's color information enables the measurement of bioluminescence emission colors. Through comprehensive detector simulations, this thesis presents, for the first time, a measurement of the spectral population of bioluminescent organisms in the deep sea. The study demonstrates how the spectral population and a catalog of bioluminescent organisms [80] can be utilized to identify the contributions of individual species. Notably, the two identified species are: *tomopteris nisseni*, a worm known for its yellow emission at $(565 \pm 14)\,\mathrm{nm}$, and *leachia pacifica*, a glass squid emitting light at $(514 \pm 7)\,\mathrm{nm}$. It is noteworthy that *tomopteris nisseni* accounts for $6.9\,\%$ of the emissions in the lower camera, situated $144\,\mathrm{m}$ above the sea floor, while its yellow emission is undetectable in the upper camera positioned $408\,\mathrm{m}$ from the sea floor. Indications in the literature have been found that *tomopteris nisseni* predominantly resides in the $200\,\mathrm{m}$ layer above the sea floor, although this information is not directly reported [88]. Furthermore, the dominant peak at $486\,\mathrm{nm}$ in the spectral population, which covers $50\,\%$ of the emissions, is likely caused by species of *Ctenophora* (comb jellies).

Subsequently, a comprehensive analysis is conducted to derive the relationship between the bioluminescence emission rate, the current speed, and the

concentration of emitting organisms. This investigation necessitates a precise comprehension of the water flows encompassing the detector. The findings reveal that the water flow surrounding the detector operates within the range that generates vortex streets. These vortex streets induce oscillations of the modules, which are effectively measured using the magnetic compass. By correlating this data with the measurements from the currentmeter, the anticipated linear dependence between oscillation frequency and flow speed is confirmed. The observed oscillations exhibit an average amplitude of 1.5°, with the strongest peak reaching 5°, and frequencies ranging from $0.01\,\mathrm{Hz}$–$0.04\,\mathrm{Hz}$. Importantly, this technique can be reversed, enabling the measurement of the current speed around a module based on the oscillations detected by the magnetic compass.

In this thesis's concluding analysis, the concentration of emitting organisms is estimated by examining the likelihood of bioluminescence emissions at various flow speeds. The data collected from both cameras results in a concentration of $(0.03 \pm 0.03)\,\frac{1}{\mathrm{m}^3}$. This measurement falls in the range of other sites instrumented with neutrino telescopes.

Acknowledgment

Undoubtedly, this thesis represents one of the most remarkable milestones in my scientific journey. Managing a project of this magnitude has come with its fair share of challenges, and I am immensely grateful to the outstanding team of individuals who have provided their unwavering support.

First of all, I want to express my heartfelt appreciation to Elisa Resconi for giving me the opportunity to lead my own project and for her invaluable guidance and support throughout the years. Your advice has been instrumental in shaping this thesis.

The contributions of numerous individuals have significantly influenced the outcome of this thesis. I want to express my sincere appreciation to Christian Spannfellner and Christian Fruck for your exemplary leadership in the hardware development of STRAW-b and to Li Ruohan for your unwavering support throughout the module development and testing process. Special thanks also go to Lisa Schumacher and Sophie Loipolder for your pivotal role in establishing the foundation of the neural network image classifier. Moreover, I want to extend my gratitude to Patrick Hatch and Braeden Veenstra for your collaborative work on the deep sea data.

Furthermore, I am deeply grateful to Stephan Meighen-Berger, Christian Haack, and Felix Henningsen for your enlightening discussions and invaluable proofreading contributions to this thesis. Your insights have been instrumental in refining its quality and coherence.

My deep gratitude also to my colleagues. I've always appreciated the time we have spent over several coffee breaks, group meetings, as well as private discussions. Special thanks go to Theo; sharing an office with you has always been a pleasure.

I am deeply grateful to Jol Thomson and Diogo da Cruz for providing me with invaluable interdisciplinary collaboration opportunities that have resulted in var-

ious art projects over the past years.

Thanks specifically to Martina Karl, Linus Lotz, and the entire *ito* team for their exceptional dedication and remarkable efforts in advancing Bluetooth and data privacy for tracing the COVID-19 pandemic.

My appreciation also goes to the entire ONC team for their exceptional deep-sea operations and unwavering support throughout the process of integrating the detector.

Furthermore, I want to extend special thanks to the team behind the IceCube camera, particularly Carsten Rott, Christoph Tönnis, and Kerrit Roellinghoff, for their invaluable contribution of five cameras, the control software, and their unwavering assistance in addressing any queries that arose.

I am immensely grateful to my family, whose persistent presence and support have propelled me forward at every step, bringing me to where I stand today. Finally, I express my most profound appreciation to Lena and Frida, who have stood by my side throughout this entire journey. Your steadfast support, inspirational influence, and the joyful moments we shared during the arduous writing phases have been invaluable to me. I am genuinely thankful for your unwavering faith and support, which have consistently provided me with boundless motivation and strength.

A Data Acquisition File Structure

STRAW-b's data acquisition system (DAQ) is responsible for storing measurement data in HDF5 files. Each HDF5 file contains sensor data as well as related settings and status parameters, which are organized according to the structure outlined in Section 2.2.3. For single measurements that are not continuous, such as LiDAR scans, a single file is generated. For continuous measurements, a new file is generated after a fixed rollover period to control the file size.

The file names follow the ONC data naming scheme and are structured as

<deviceCode>_<dateFrom>-SDAQ-<label>.hdf5.

The *deviceCode* is a unique identifier for each module (e.g., *TUMPMTSPECTROMETER001*), the *dateFrom* is the starting date and time of the period to which the file belongs, and the *label* corresponds to a Data Product Code (DPC) in the ONC database (Oceans 2.0 or Oceans 3.0). Refer to Table A.1 for information on which sensor data is stored in each file, along with the rollover interval and DPC.

Each HDF5 group corresponds to a timeseries of data, and the *time* dataset in each group stores the timestamps for that timeseries. Since the first dimension of all datasets is time, all datasets within a HDF5 group have the same size along the first dimension. For more information on the internal HDF5 file structure for each DPC, please refer to the tables A.2 to A.7.

Table A.1: List of the most important data acquisition (DAQ) files in Strings for Absorption Length in Water b (STRAW-b) including Data Product Code (DPC) and the period when a the DAQ starts a new file (rollover) to manage the maximum file size.

Sensor	Modules	rollover[a]	DPC[b]	label
3-Axis Accelerometer	All Modules	24 h	SMRD	MODULE
3-Axis Magnetometer	All Modules	24 h	SMRD	MODULE
PTH[c]	All Modules	24 h	SMRD	MODULE
6 Powermeter[d]	All Modules	24 h	SMRD	MODULE
3 Temperature	All Modules	24 h	SMRD	MODULE
Camera	Mini- & PMT-Spectrometer	1 h	MSSCD	CAMERA
Mini-Spectrometer	Mini- & PMT-Spectrometer	24 h	MSSD	MINISPEC
Mini-Spectrometer	LiDAR	per run	MSSD	MINISPEC
16 PMT (TOT counts)	PMT-Spec.	1 h	PMTSD	PMTSPEC
16 SiPM (TOT events)	Muon-Tracker	24 h	MTSD	MUON
LiDAR (TOT events)	LiDAR	per run	LIDARSD	LIDAR

[a]Period after which a new file is started. The periods normalize to full hours, midnight, etc.
[b]Data Product Code
[c]Pressure-Temperature-Humidity sensor
[d]The Powermeter measures for each channel current and voltage separately

Table A.2: HDF5 file structure of basic module DAQ file recorded by all modules. The Ocean Networks Canada (ONC)'s Data Product Code (O-DPC) is used in the data base (DB) of ONC to effectively filter the files.

O-DPC	HDF5 Group	Dataset	Shape
SMRD	accel	time	(t,)
SMRD	accel	x	(t,)
SMRD	accel	y	(t,)
SMRD	accel	z	(t,)
SMRD	magneto	time	(t,)
SMRD	magneto	x	(t,)
SMRD	magneto	y	(t,)
SMRD	magneto	z	(t,)
SMRD	pth	humidity	(t,)
SMRD	pth	pressure	(t,)
SMRD	pth	temperature	(t,)

Table continued on next page

Table A.2 – continued from previous page

O-DPC	HDF5 Group	Dataset	Shape
SMRD	pth	time	(t,)
SMRD	pwrmoni	c2_current	(t,)
SMRD	pwrmoni	c2_voltage	(t,)
SMRD	pwrmoni	laser_current	(t,)
SMRD	pwrmoni	laser_voltage	(t,)
SMRD	pwrmoni	motor_current	(t,)
SMRD	pwrmoni	motor_voltage	(t,)
SMRD	pwrmoni	padiwa_current	(t,)
SMRD	pwrmoni	padiwa_voltage	(t,)
SMRD	pwrmoni	switch_current	(t,)
SMRD	pwrmoni	switch_voltage	(t,)
SMRD	pwrmoni	time	(t,)
SMRD	pwrmoni	trb3sc_current	(t,)
SMRD	pwrmoni	trb3sc_voltage	(t,)
SMRD	temperatures	temp1	(t,)
SMRD	temperatures	temp2	(t,)
SMRD	temperatures	temp3	(t,)
SMRD	temperatures	time	(t,)

Table A.3: HDF5 file structure of the Mini-Spectrometer DAQ file. The ONC's Data Product Code (O-DPC) is used in the DB of ONC to effectively filter the files. <ID> refers to the Mini-Spectrometer which can be between 1 and 5. The LiDAR and PMT-Spectrometer modules only hosts one Mini-Spectrometer (<ID>=1). The Mini-Spectrometer module has five Mini-Spectrometers (<ID>=1..5) and therefore the HDF5 file has five HDF5 Group.

O-DPC	HDF5 Group	Dataset	Shape
MSSD	<ID>	ADCcounts	(t, 288)
MSSD	<ID>	exposure_time	(t,)
MSSD	<ID>	temperature_after	(t,)
MSSD	<ID>	temperature_before	(t,)
MSSD	<ID>	time	(t,)
MSSD	<ID>	trigger_counter	(t,)

Table A.4: HDF5 file structure of the camera DAQ file. The ONC's Data Product Code (O-DPC) is used in the DB of ONC to effectively filter the files.

O-DPC	HDF5 Group	Dataset	Shape
MSSCD	camera	exposure_time	(t,)
MSSCD	camera	exposure_time_cmd_setting	(t,)
MSSCD	camera	gain	(t,)

Table continued on next page

Table A.4 – continued from previous page

O-DPC	HDF5 Group	Dataset	Shape
MSSCD	camera	lucifer_options	(t, 4)
MSSCD	camera	measured_capture_time	(t,)
MSSCD	camera	measured_download_time	(t,)
MSSCD	camera	measured_exposure_time	(t,)
MSSCD	camera	raw	(t, 1297, 977)
MSSCD	camera	reported_resolution	(t, 2)
MSSCD	camera	time	(t,)

Table A.5: HDF5 file structure of the PMT-Spectrometer DAQ file. The ONC's Data Product Code (O-DPC) is used in the DB of ONC to effectively filter the files.

O-DPC	HDF5 Group	Dataset	Shape
PMTSD	counts	ch0	(t,)
PMTSD	counts	ch1	(t,)
		...	
PMTSD	counts	ch15	(t,)
PMTSD	counts	time	(t,)
PMTSD	daq	frequency_readout	(t,)
PMTSD	daq	state	(t,)
PMTSD	daq	time	(t,)
PMTSD	daq	trb	(t,)
PMTSD	hv	ch0	(t,)
PMTSD	hv	ch1	(t,)
		...	
PMTSD	hv	ch15	(t,)
PMTSD	hv	power	(t,)
PMTSD	hv	time	(t,)
PMTSD	padiwa	offset	(t,)
PMTSD	padiwa	power	(t,)
PMTSD	padiwa	th1	(t,)
PMTSD	padiwa	th2	(t,)
		...	
PMTSD	padiwa	th16	(t,)
PMTSD	padiwa	time	(t,)

Table A.6: HDF5 file structure of the Muon-Tracker DAQ file. The ONC's Data Product Code (O-DPC) is used in the DB of ONC to effectively filter the files.

O-DPC	HDF5 Group	Dataset	Shape
MTSD	counts	ch0	(t,)
MTSD	counts	ch1	(t,)
		...	

Table continued on next page

Table A.6 – continued from previous page

O-DPC	HDF5 Group	Dataset	Shape
MTSD	counts	time	(t,)
MTSD	daq	frequency_readout	(t,)
MTSD	daq	state	(t,)
MTSD	daq	time	(t,)
MTSD	daq	trb	(t,)
MTSD	hv	hv_value	(t,)
MTSD	hv	time	(t,)
MTSD	padiwa	offset	(t,)
MTSD	padiwa	power	(t,)
MTSD	padiwa	th1	(t,)
MTSD	padiwa	th2	(t,)
		...	
MTSD	padiwa	th16	(t,)
MTSD	padiwa	time	(t,)

Table A.7: HDF5 file structure of the LiDAR DAQ file. The ONC's Data Product Code (O-DPC) is used in the DB of ONC to effectively filter the files.

O-DPC	HDF5 Group	Dataset	Shape
LIDARSD	counts	ch0	(t,)
LIDARSD	counts	ch17	(t,)
LIDARSD	counts	ch18	(t,)
LIDARSD	counts	time	(t,)
LIDARSD	daq	frequency_readout	(t,)
LIDARSD	daq	frequency_trigger	(t,)
LIDARSD	daq	pmt	(t,)
LIDARSD	daq	state	(t,)
LIDARSD	daq	time	(t,)
LIDARSD	daq	trb	(t,)
LIDARSD	gimbal	delay	(t,)
LIDARSD	gimbal	pos_x	(t,)
LIDARSD	gimbal	pos_y	(t,)
LIDARSD	gimbal	power	(t,)
LIDARSD	gimbal	time	(t,)
LIDARSD	hld	file_end	(t,)
LIDARSD	hld	file_start	(t,)
LIDARSD	hld	time	(t,)
LIDARSD	laser	diode	(t,)
LIDARSD	laser	frequency	(t,)
LIDARSD	laser	power	(t,)
LIDARSD	laser	pulsewidth	(t,)
LIDARSD	laser	set_adjust_x	(t,)
LIDARSD	laser	set_adjust_x_offset	(t,)
LIDARSD	laser	set_adjust_y	(t,)

Table continued on next page

Table A.7 – continued from previous page

O-DPC	HDF5 Group	Dataset	Shape
LIDARSD	laser	set_adjust_y_offset	(t,)
LIDARSD	laser	time	(t,)
LIDARSD	measurement	step	(t,)
LIDARSD	measurement	time	(t,)
LIDARSD	tot	hld_start_time	(t,)
LIDARSD	tot	time	(t,)
LIDARSD	tot	time_ns	(t,)
LIDARSD	tot	tot	(t,)

B Sensor and Module Malfunctions in STRAW-b

Exploration and research in the deep sea can be incredibly challenging due to the harsh environment. Equipment used in this context must be designed and constructed to withstand the extreme pressures and temperatures found at great depths. This can be a significant engineering challenge, as the materials and components used must be able to function reliably despite the difficult conditions. Additionally, because repair and maintenance are often not possible once equipment is deployed in the deep sea, reliability is critical and a failure of even a single piece of equipment could jeopardize an entire mission.

STRAW-b faced some issues with single components or entire modules. In the following, the known malfunctions are discussed. As stated in Table B.1, a total of four modules experienced a loss of communication since the deployment of STRAW-b. Standard Module 3, in contrast to the others, had no network connection from the initial deployment. However, the remaining modules encountered a loss of network connection after several months of operation. Determining the exact cause of the connection issues while STRAW-b is deployed in the deep sea remains challenging. However, there are indications that the problem could be related to network connection issues, such as broken fibers or connector problems.

B.1 Analyzing Module Failures in STRAW-b

An interesting observation is that the electronics of Standard Module 3 successfully endured the deployment, as evidenced by the flashing LED response during the remotely operated underwater vehicle (ROV) inspection. The flashing LED

Table B.1: Sensor and Module Malfunctions in STRAW-b. Further investigation into the causes of lost communication with certain modules is discussed in Section B.1. Additionally, Section B.2 summarizes the issues encountered with the broken FPGA in the PMT-Spectrometer 2 module.

Module	Malfunctions	Date
Standard Module 3[a]	no communication to module	since deployment
PMT-Spectrometer 2	Padiwa: broken or short circuit FPGA which is required for the PMT readout	2021-10-11 11:11
WOM	booting fails due to a broken SD card hosting the OS (operating system)	2021-10-30
Mini-Spectrometer	lost communication to module	2021-06-14 02:30
LiDAR 2	lost communication to module	2021-11-28 10:09

[a]Standard Module 3 hosts the art installation *radioamnion*

serves as an indicator that the electronics survived the deployment. When the module detects the light emitted by the ROV, it triggers the LED to flash, which remains active for a few minutes before being permanently disabled to avoid interference with subsequent operations. Another indication pointing towards network connection issues is the power consumption of the modules, which is continuously monitored by the mini-junction box (mJB). Comparing the power consumption during booting sequences before and after module failures, a similar chronological power consumption pattern can be observed, as depicted in Figure B.1 for the WOM. This similarity suggests that the modules are booting properly, and the issue lies specifically with the network connection. Furthermore, the statistic of ONC about connector failures is approximate 10 %, something considered too high for P-ONE. Consequently, penetrators and connectors are not employed within the P-ONE system to mitigate this issue.

The failure of the WOM is anticipated to be connected to a broken SD-card that hosted its operating system. This failure occurred subsequent to an unforeseen power outage experienced by the entire NEPTUN[1] infrastructure due to a

[1]NEPTUNE is a ocean observatory off the west coast of Vancouver Island. Further information are summarized in Chapter 2.

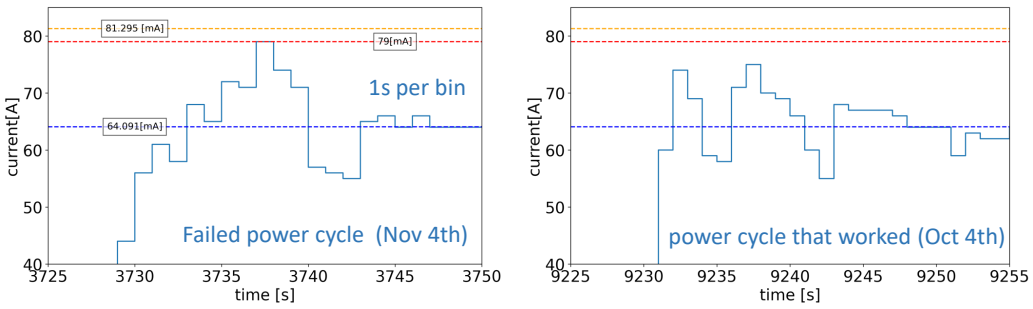

Figure B.1: Current consumption of the Wavelength-shifting Optical Module (WOM) during boot sequence. The left plot shows a period of normal module operation, while the right plot depicts the power up sequence after an unexpected power outage. Since the incident, the WOM does not respond to network communication. The interpretation is discussed in the text. Graphs courtesy of Lutz Köpke, University of Mainz.

storm in close proximity to the NEPTUN shore station. Notably, it is important to highlight that only the WOM utilized an SD-card for hosting its operating system, whereas the other modules employed eMMC (embedded MultiMediaCard), which are more resilient to data corruption compared to SD-cards.

B.2 FPGA Short Circuit Incident in the PMT-Spectrometer 2

On October 11, 2021, at 11:11, a significant incident occurred in the PMT-Spectrometer 2 module of the system. The incident involved a short circuit in the Field-Programmable Gate Array (FPGA) located on the Padiwa board. Unfortunately, the FPGA could not be reprogrammed following the short circuit, rendering it non-functional.

The functionality of the PMT-Spectrometer 2 module heavily relies on the FPGA integrated into the Padiwa board. Its crucial role involves applying thresholds to the signals received from the PMTs, enabling the accurate measurement of time-over-threshold (ToT) as detailed in Section 2.2.4. However, due to the FPGA malfunction, the signal processing capability of the PMT-Spectrometer 2 module was compromised, resulting in the inability to record the signals from the PMTs.

It is important to note that while the FPGA issue affected the PMT signals, other sensors present in the PMT-Spectrometer 2 module, such as the camera and mini-spectrometer, remained unaffected. Their functionality and data acquisition capabilities remained intact, allowing for continued operation and data collection from these sensors despite the FPGA malfunction.

C Cluster Detection and Classification

C.1 Cluster Data Base Parameters

Description of the parameters that the detection algorithm stores for each cluster. There are parameters which belong to the period used to calculate the noise level (μ_i) and its standard deviation (σ_i). Other parameters refer to the image in which the cluster occurs. And finally, parameters calculated on the clusters pixel location, color and charge values. The parameters are separated in this three groups in the following tables, Table C.1, Table C.2, and Table C.3, respectively. It should be mentioned, that the python package *pandas* is used and therefore multidimensional parameters are stored in different columns as *pandas* has increased performance for dimensional data per column. E.g. a 2D position vector is stored in two different columns with the suffices *_x* and *_y*.

Table C.1: Parameter about the period over which the pixels mean noise level (μ_i) and its standard deviation (σ_i) is calculated.

Parameter	Description	Unit
mean_std_n	Number of images in the period	count
mean_std_start	Date of periods first image	date
mean_std_stop	Date of periods last image	date

Table C.2: Parameter of the image to which the cluster belongs.

Parameter	Description	Unit
time	Capture time	date
label	A unique label for each cluster in one image. Labels can occur more than once in a database if the clusters they refer to belong to different images. Label 0 describes the background, i.e. all pixels whose charge is lower than the specified threshold.	count

Table C.3: Parameters calculated on the cluster's pixels.

Parameter	Description	Unit
center_of_mass[1]	Coordinates of the center of mass of the pixels belonging to the cluster with respect to the charge. $$\vec{R_{\text{int}}} = \frac{1}{\sum_{i=0}^{n} I_i} \sum_{i=0}^{n} I_i \vec{r_i}$$	pixel
center_of_pix[1]	Coordinates of the center of the pixels belonging to the cluster without considering the charge. $$\vec{R}_{\text{pixel}} = \frac{1}{n} \sum_{i=0}^{n} \vec{r_i}$$	pixel
box_center[1]	Coordinates of the center of the minimum bounding box around the cluster's pixels.	pixel
box_size[1]	Width and height of the minimum bounding box around the cluster's pixels.	pixel

Table continued on next page

Table C.3 – continued from previous page

Parameter	Description	Unit
angle	The angle by which the minimum bounding box of the cluster is rotated to the x-axis. The side of the box that encloses the angle with the x-axis is the one that connects the corner with the smallest y-value to the next corner in the counterclockwise direction.	rad
n_pixel[2]	The number the cluster's pixels.	pixel
charge[2]	The integrated charges of the cluster's pixels. The noise level (μ_i) is subtracted.	intensity
noise[2]	The integrated noise (μ_i) of the cluster's pixels.	intensity
sn_mean[2]	The mean signal (pixels charge) to noise (μ_i) ratio in multiples of the standard deviation (σ_i) over the cluster's pixels.	factor
sn_sigma[2]	The standard deviation signal (pixels charge) to noise (μ_i) ratio in multiples of the standard deviation (σ_i) over the cluster's pixels.	factor
distance_fov	shortest distance of the center_of_pix to the modules FoV boundary, i.e. where the PMT mounting cuts the FoV of the camera.	pixel
distance_cable	shortest distance of the center_of_pix to either the steel line or data cable	pixel
nearest_point_cable[1]	closest point of the center_of_pix on either the steel line or data cable	pixel

[1] This parameter exists for x and y separately, i.e.: *center_of_mass* are 2 parameters.
[2] This parameter is parameter exits for red, blue, green and all colors separately.

C.2 Hotspots of the Detected Cluster in the Images

The distribution of clusters in the PMT-Spectrometer 2 module (Figure C.1) exhibits similar hotspots as the PMT-Spectrometer 1 module (Figure 6.10), with the exception that the cable is in a different location.

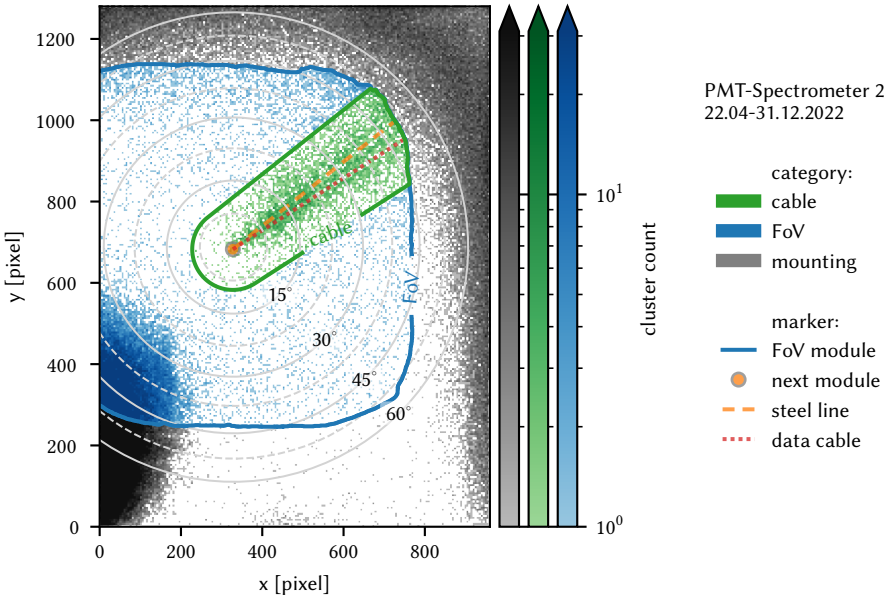

Figure C.1: Cluster distribution of the cluster's pixel center. Noise is mainly located outside the module's FoV where no light is expected. Only the accumulation in the lower left enters the module's FoV. Because the area of the cable is visible and bioluminescence emits light after mechanical stimulation, clusters in this area are from bioluminescence with a high probability. All clusters within a 100 pixel are allocated to the category *bioluminescence*. Supervised machine learning (ML) can train on the camera noise outside the module's FoV and bioluminescence clusters around the cables to classify clusters across the entire image. Figure 6.10 shows the cluster captured by the camera in the PMT-Spectrometer 2 module.

C.3 Feature Importance for the Machine Learning

Feature importance in a decision tree classifier is determined by the information gain achieved by splitting on each feature. The higher the information gain, the more important the feature is for classification [104]. Another technique for

assessing feature importance is mutual information, as shown in Figure 6.12. Both techniques identify position-biased features that should be excluded from the machine learning process.

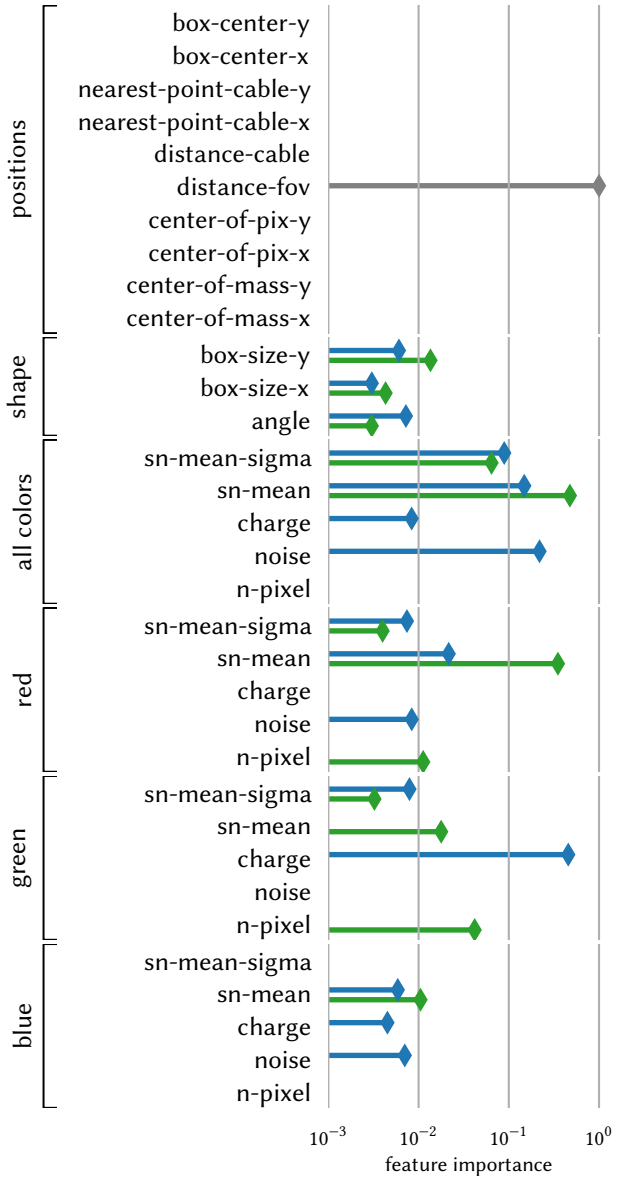

Figure C.2: Feature importance from a decision tree classifier [104]. The colors represent the three different sets of features provided to the decision tree classifiers. *Gray* takes all listed features. *Blue* excludes all features within the group *position*. And *green* excludes the *position*, *noise* and *charge features*.

C.4 Locations and Size of Predicted Clusters

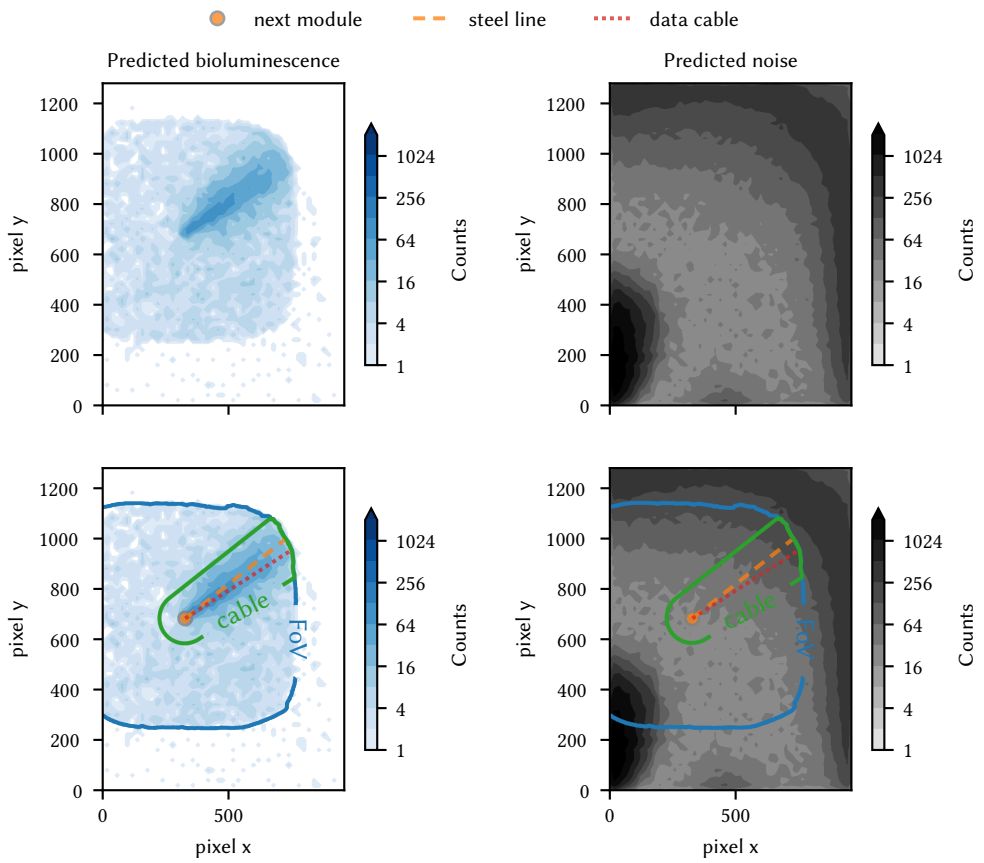

Figure C.3: Cluster location for clusters with a size ≥5 pixel and camera in the PMT-Spectrometer 2 module, similar to Figure 6.15, which shows camera in the PMT-Spectrometer 1 module. Absolute cluster count is plotted instead of density to highlight the minimum number of clusters in certain areas. Plots in the same color show the same data, where the lower row has marked FoV and cable areas which originate from Figure 5.7.

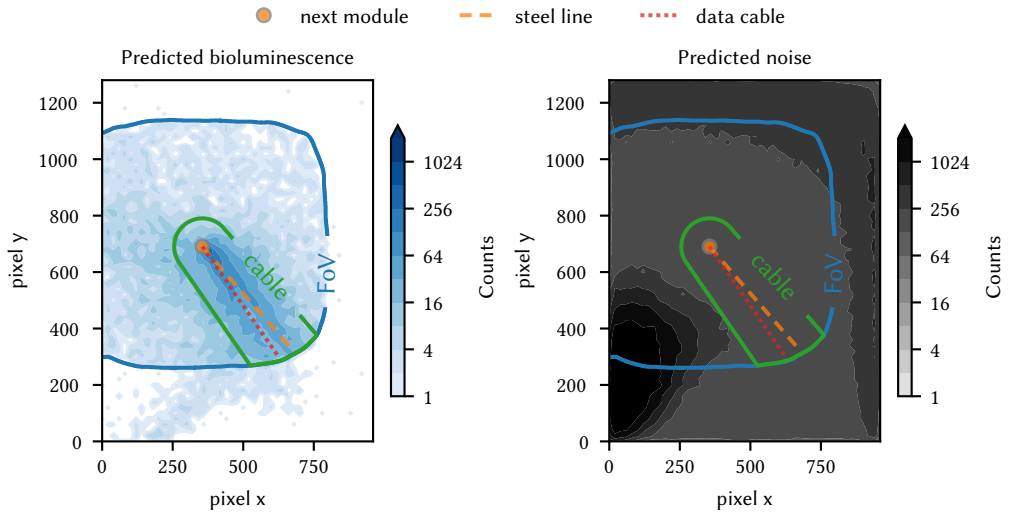

Figure C.4: 2D histogram of the cluster location for clusters with a size ≥4 pixel, similar to Figure C.5 and Figure 6.15, which use 6 or 5 pixel as threshold to train the neural network (NN). Absolute cluster count is plotted instead of density to highlight the minimum number of clusters in certain areas. Plots in the same color show the same data, where the lower row has marked FoV and cable areas which originate from Figure 5.7.

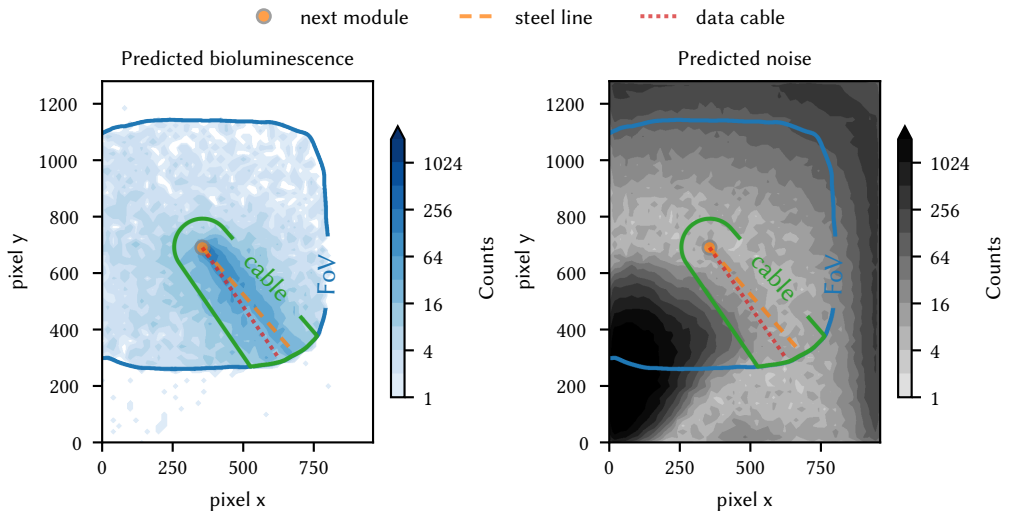

Figure C.5: 2D histogram of the cluster location for clusters with a size ≥6 pixel, similar to Figure C.4 and Figure 6.15, which use 4 or 5 pixel as threshold to train the NN. Absolute cluster count is plotted instead of density to highlight the minimum number of clusters in certain areas. Plots in the same color show the same data, where the lower row has marked FoV and cable areas which originate from Figure 5.7.

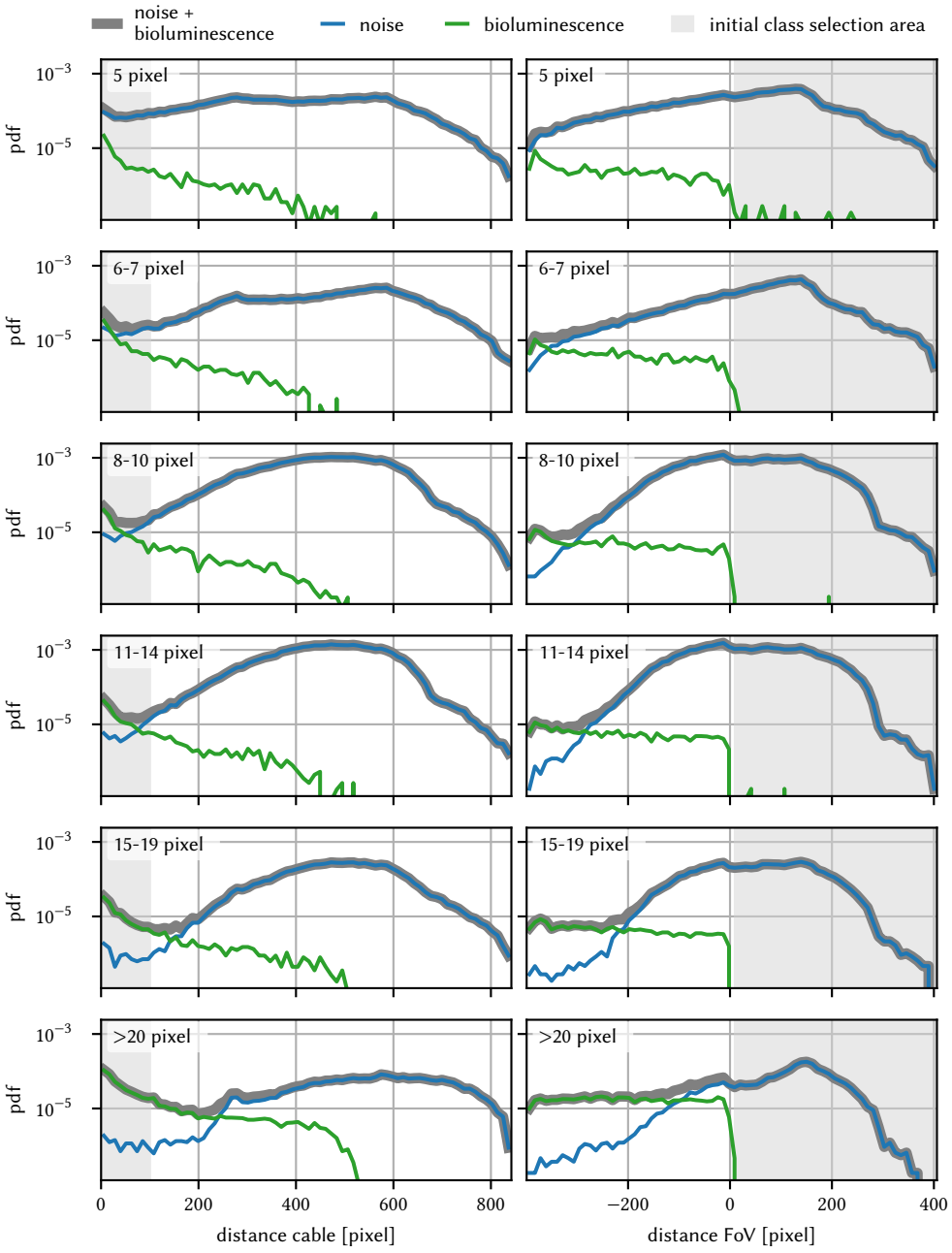

Figure C.6: Probability density function (PDF)s of the cluster locations depicting the performance of the ML classification. The plot is similar to Figure 6.15 but the NN is trained with clusters >4 pixel in size which improves the performance slightly.

D Bioluminescence Catalog

The bioluminescence catalog from Latz, Frank, and Case (1988)[80] is a comprehensive reference for bioluminescent organisms in the ocean. During the April 1985 Biowatt cruise, the spectral characteristics of bioluminescence were measured using an OMA system in the western Sargasso Sea. The highest intensity of bioluminescence emissions was observed in the blue region of the visible spectrum, specifically within the wavelength range of 440 nm–500 nm. Intraspecific variability in spectra was observed in several species, and some species exhibited dynamic color shifts. Measurements from independent samples of unsorted plankton indicate that the spectral emissions of bioluminescence vary depending on the species assemblage [80]. Table D.1 lists the bioluminescence catalog and Figure D.1 shows a visual representations of the catalog.

Table D.1: Spectral catalog for bioluminescence emissions. Data are from [80]. Visualizations of the catalog are in Figure 8.5 and Figure 8.9

Identification	peak [nm]	FWHM [nm]	SN	2nd peak [nm]
Unsorted plankton				
mixed plankton 0 (333 μm^{1})	459	66	50	
mixed plankton 1 (25 μm^{1})	474	40	66	
mixed plankton 2 (25 μm^{1})	478	56	18	
mixed plankton 3 (25 μm^{1})	484	62	24	472.0
Protozoa				
Rhaphidozoum acuferum	458	87	28	
A crosphaera murrayana	443	80	36	
Siphonosphaera tenera	450	78	34	
Myxosphaera coerulea	453	84	37	
Collosphaera huxleyi	456	79	31	

Table continued on next page

Table D.1 – continued from previous page

Identification	peak [nm]	FWHM [nm]	SN	2nd peak [nm]
Collosphaera TCM[2] 1	452	77	35	
Collosphaera TCM[2] 2	445	87	71	
Collosphaera SCM[3]	443	75	63	
Coelenterata				
Chrysaora hysosceles	478	95	67	
Bougainvillia carolinensis	452	74	38	
Pelagia noctiluca	469	94	32	
Aeginea citrea	459	73	149	
Pegantha clara	460	71	103	
Pandea sp. nov.	466	80	94	
Hippopodius hippopus	447	80	112	
A galma okeni	444	70	94	
Amphicaryon ernesti	487	47	144	
Amphicaryon acaula	487	65	78	
Diphyes dispar	464	92	25	
Rosacea larva	488	55	90	
Ctenophora				
Cestum veneris	490	84	72	
Tinerfe lactae	486	85	57	
Beroe cucumis	479	94	52	496.0
Beroe ovata	478	86	51	
Bolinopsis sp.	488	80	28	
Mollusca				
Phyllirrhoe sp. Leachia lemur	475	89	44	
Leachia lemur 0	500	84	55	473.0
Leachia lemur 1	514	68	75	489.0
Leachia lemur 2	458	74	58	485.0
Leachia lemur 3	449	86	40	459.0
Pyroteuthis margaritifera 0	477	54	24	
Pyroteuthis margaritifera 1	475	62	81	
Pyroteuthis margaritifera 2	485	42	12	470.0
Annelida				
Tomopteris nisseni	565	55	92	
Crustacea				
Conchoecia imbricata	474	94	47	
Conchoecia secernenda	481	95	24	
Scina sp.	444	89	38	
Pleuromamma xiphias	492	77	117	472.0
Pleuromamma abdominalis	486	77	78	465.0
Gaussiaprinceps	479	73	38	489.0

Table continued on next page

Table D.1 – continued from previous page

Identification	peak [nm]	FWHM [nm]	SN	2nd peak [nm]
Nematoscelis microps	463	43	77	
Nematobrachionflexipes	453	32	32	
Euphausia brevis	462	43	64	
Euphausia gibboides	467	53	42	
Oplophorus spinosus sp.[4]	457	69	133	
Systellaspis debilis sp.[4]	460	65	73	
Systellaspis debilis sp.[5]	467	48	38	
Tunicata				
Pyrosoma atlanticum	493	101	39	471.0
Pisces				
Ultrastomias mirabilis	477	73	325	
Hygophum hygomi	448	76	58	
Stomias brevibarbatus	689	150	105	

[1] Mesh size to collect the plankton
[2] TCM: Toroid colony morphology
[3] SCM: Spherical colony morphology
[4] Luminescent secretion
[5] Photphore emission

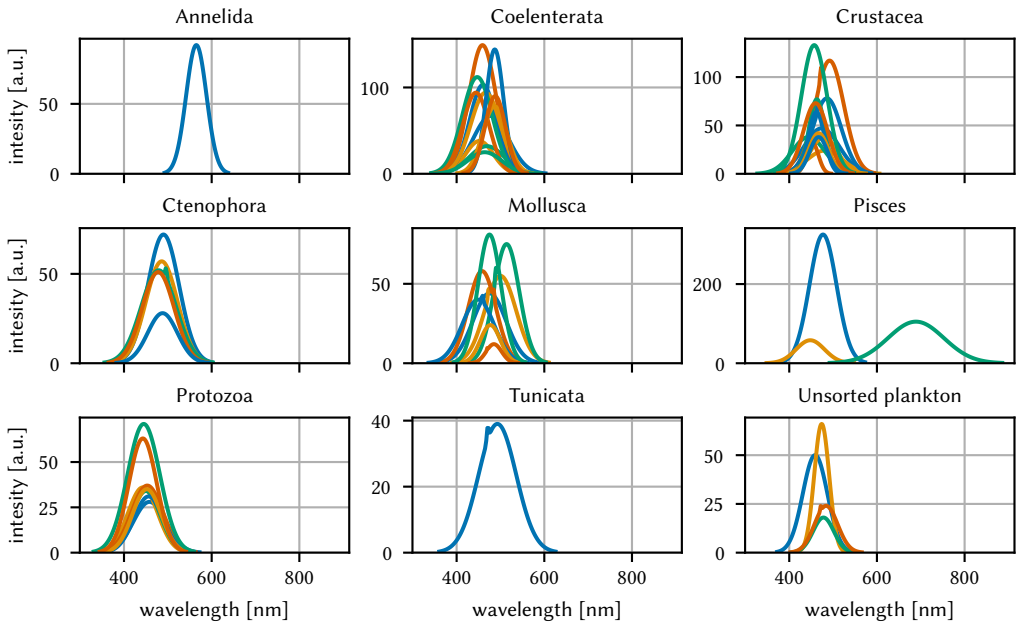

Figure D.1: Visualization of the bioluminescence spectral catalog. Data are from [80] and the emission spectra is modeled by a normal distribution. If a secondary peaks is reported in a spectra, it is modeled with a normal distribution with $\sigma/20$ and SN/10.

E Water Flow Simulation and Measurements

E.1 Flow Simulation Around Detector Components

Simulating water flow with Solidworks provides valuable information about the flow field around the cable and module. The simulation uses a CAD model of the STRAW-b module, cable, and mounting components, but some parts are excluded for simplicity, such as the data cable. Figure E.1 and Figure E.2 show the flow field while Figure E.3 and Figure E.4 show the gradient of the flow field.

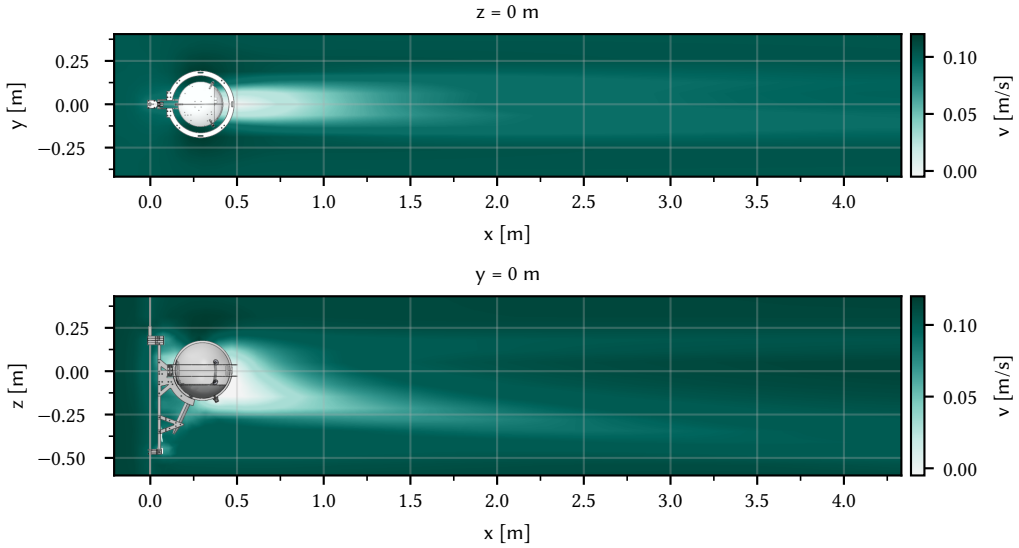

Figure E.1: Water flow around module and cable for a velocity of 10 $\frac{\text{cm}}{\text{s}}$. The underlying simulations was performed in Solidworks by Christian Spannfellner.

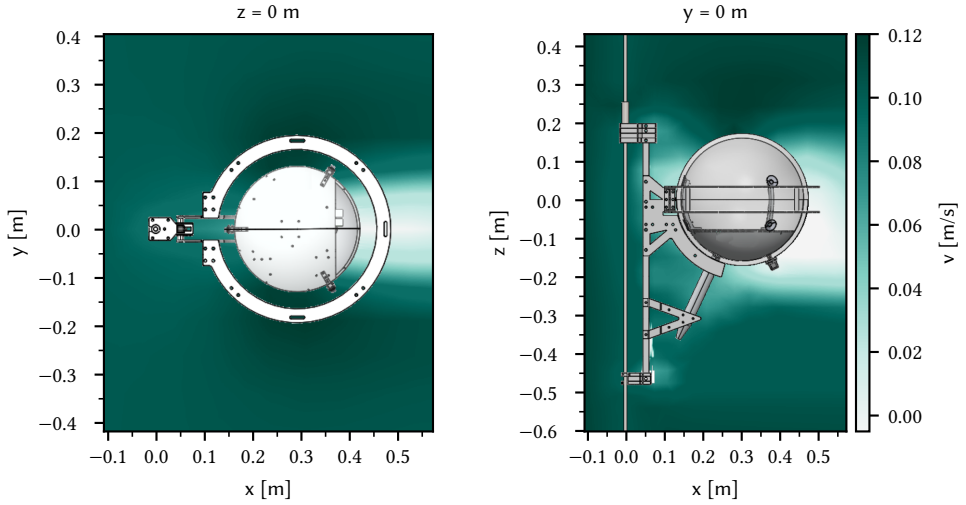

Figure E.2: Water flow around module and cable for a velocity of 10 $\frac{\text{cm}}{\text{s}}$. The underlying simulations was performed in Solidworks by Christian Spannfellner.

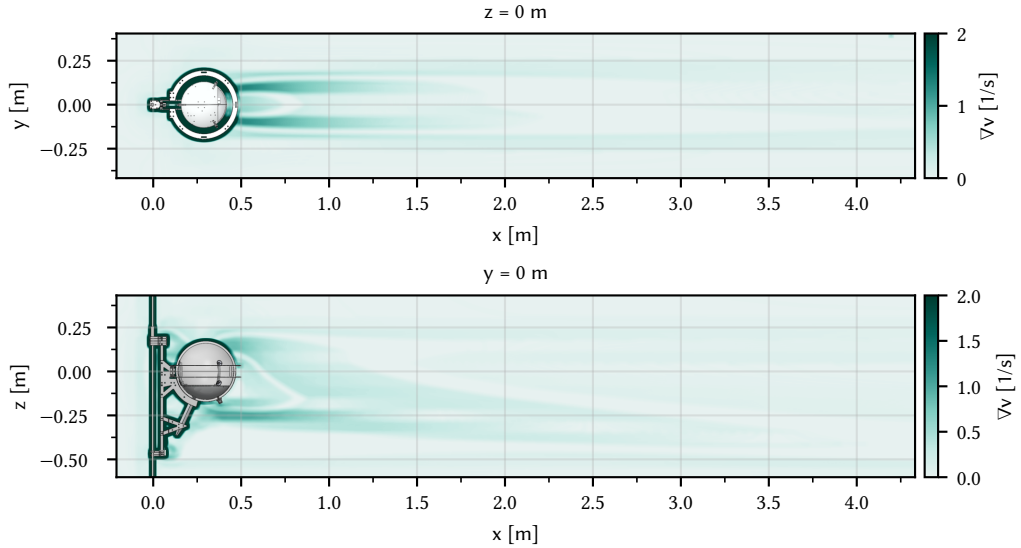

Figure E.3: Water flow gradient around module and cable for a velocity of 10 $\frac{\text{cm}}{\text{s}}$. The underlying simulations was performed in Solidworks by Christian Spannfellner.

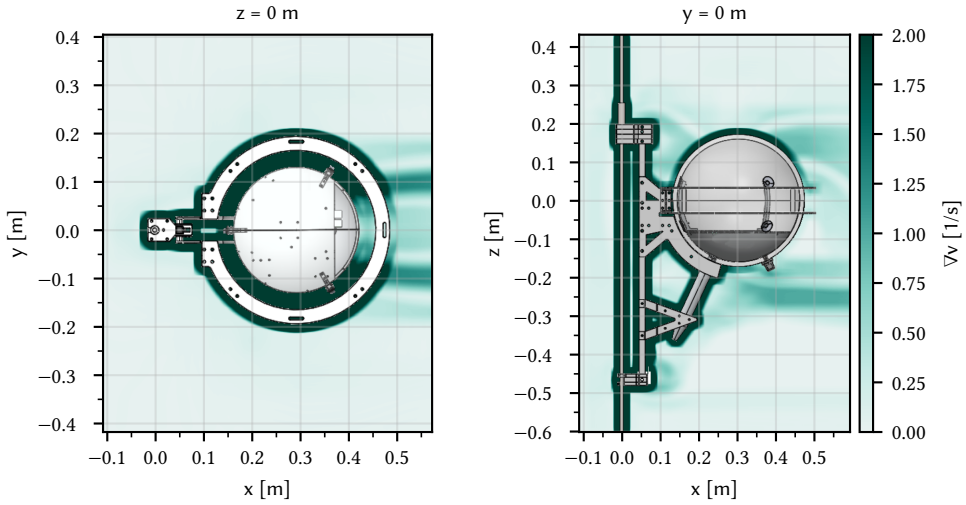

Figure E.4: Water flow gradient around module and cable for a velocity of 10 $\frac{\text{cm}}{\text{s}}$. The underlying simulations was performed in Solidworks by Christian Spannfellner.

E.2 Water Current and Module Heading

In Cascadia Basin, a currentmeter located around 2.5km away from STRAW and STRAW-b detectors measures the flow speed close to the ground. The magnetic sensors in STRAW-b modules provide a heading direction similar to the measured water flow. However, when flow speeds drop below 20 $\frac{\text{mm}}{\text{s}}$, the module heading starts to decouple from the flow direction, as shown in Figure E.5 where data from the currentmeter and PMT-Spectrometer 1 for a 10-day period are displayed.

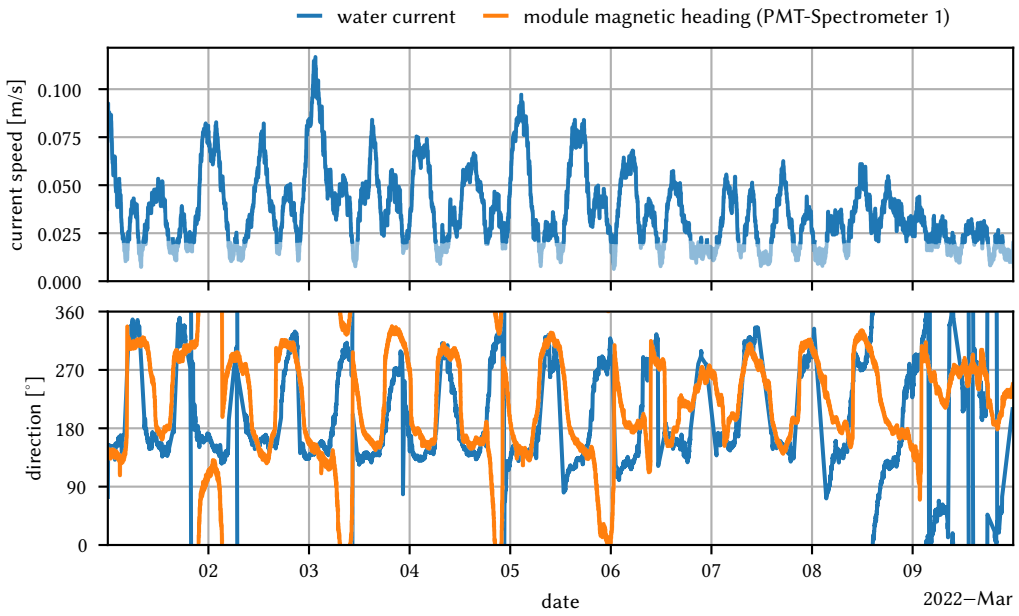

Figure E.5: Water flow measured by a currentmeter and magnetic heading of the module. The magnetic heading aligns with the flow direction for currents above 20 $\frac{\text{mm}}{\text{s}}$.

E.3 Module Oscillations from Vortex Streets

The modules experience vortex street oscillations downstream due to the influence of currents, as described in Section 8.4.2 and Section 8.4.3. Figure E.6 presents the linear detrended module heading during periods of significant oscillations, with amplitudes of up to 5°.

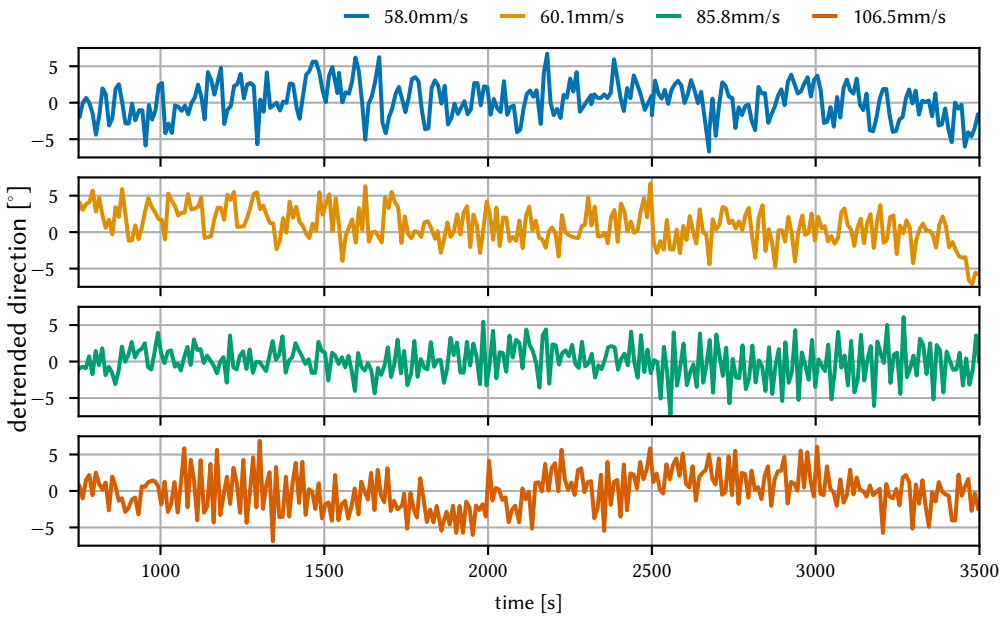

Figure E.6: Linear detrended module heading during periods of vortex street generation. The different periods are the same periods showed in Figure 8.14.

Abbreviations

ADC	analog to digital converter
ADCP	acoustic Doppler current profiler
AOV	angle of view
CMOS	complementary metal-oxide-semiconductor
CPLD	complex programmable logic device
DAQ	data acquisition
DB	data base
DCS	distributed control system
PDF	probability density function
DNA	deoxyribonucleic acid
FoV	field of view
FPGA	Field-Programmable Gate Array
FWHM	full width at half maximum
HDF5	Hierarchical Data Format
KM3NeT	Cubic Kilometer Neutrino Telescope
LED	light-emitting diode
MCTL	master control software
MI	mutual information
ML	machine learning
MSB	marine snow, sedimentation, and bio-fouling
NN	neural network
ONC	Ocean Networks Canada
PMT	photomultiplier tube
POCAM	Precision Optical Calibration Module
P-ONE	Pacific Ocean Neutrino Experiment
ROV	remotely operated underwater vehicle
sDOM	STRAW Digital Optical Module
SPI	serial peripheral interface
STRAW	Strings for Absorption Length in Water
STRAW-b	Strings for Absorption Length in Water b
TDC	time-to-digital converter
ToT	time-over-threshold
TRB	Trigger Readout Board
UV	ultraviolet
WOM	Wavelength-shifting Optical Module

List of Figures

List of Tables

Bibliography

[1] da Cruz Diogo, Holzapfel Kilian, and Thomson Jol. *UNDERCURRENTS*. SFB42, 2020. URL: https://sfb42.org/UNDERCURRENTS (visited on 03/26/2023).

[2] Diogo da Cruz and Thomson Jol. *Sonderforschungsbereich 42 (SFB42)*. SFB42, 2020. URL: https://sfb42.org/about (visited on 03/26/2023).

[3] *First Run 3 data-MC plots for the measurement of the top-quark pair production cross-section in pp collisions at centre-of-mass energy of 13.6 TeV with the ATLAS experiment at the LHC*. 2022. arXiv: 2211.09414 [hep-ex].

[4] M. G. Aartsen et al. "First Observation of PeV-Energy Neutrinos with IceCube". In: *Physical Review Letters* 111.2 (July 2013). DOI: 10.1103/physrevlett.111.021103. URL: https://doi.org/10.1103%2Fphysrevlett.111.021103.

[5] M. G. Aartsen et al. "Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector". In: *Science* 342.6161 (Nov. 2013). DOI: 10.1126/science.1242856. URL: https://doi.org/10.1126%2Fscience.1242856.

[6] Mark Aartsen et al. "Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert". In: *Science* 361.6398 (July 2018), pp. 147–151. DOI: 10.1126/science.aat2890. URL: https://doi.org/10.1126%2Fscience.aat2890.

[7] Mark Aartsen et al. "Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A". In: *Science* 361.6398 (July 2018). DOI: 10.1126/science.aat1378. URL: https://doi.org/10.1126%2Fscience.aat1378.

[8] P Padovani et al. "Dissecting the region around IceCube-170922A: the blazar TXS 0506+056 as the first cosmic neutrino source". In: *Monthly Notices of the Royal Astronomical Society* 480.1 (July 2018), pp. 192–203. DOI: 10.1093/mnras/sty1852. URL: https://doi.org/10.1093%2Fmnras%2Fsty1852.

[9] M. G. Aartsen et al. "Evidence for neutrino emission from the nearby active galaxy NGC 1068". In: *Science* 378.6619 (Nov. 2022), pp. 538–543. DOI: 10.1126/science.abg3395. URL: https://doi.org/10.1126%2Fscience.abg3395.

[10] Simone Biagi and the KM3NeT consortium. “The KM3NeT Project”. In: *Journal of Physics: Conference Series* 375.5 (July 2012). DOI: 10.1088/1742-6596/375/1/052036. URL: https://dx.doi.org/10.1088/1742-6596/375/1/052036.

[11] Elisa Resconi. *Towards a new neutrino telescope in the Pacific Ocean*. 2023. URL: https://www.pacific-neutrino.org/ (visited on 03/29/2023).

[12] M. Boehmer et al. “STRAW (STRings for Absorption length in Water): pathfinder for a neutrino telescope in the deep Pacific Ocean”. In: *Journal of Instrumentation* 14.02 (Feb. 2019), P02013–P02013. DOI: 10.1088/1748-0221/14/02/p02013. URL: https://doi.org/10.1088%2F1748-0221%2F14%2F02%2Fp02013.

[13] Nicolai Bailly et al. “Two-year optical site characterization for the Pacific Ocean Neutrino Experiment (P-ONE) in the Cascadia Basin”. In: *The European Physical Journal C* 81.12 (Dec. 2021). DOI: 10.1140/epjc/s10052-021-09872-5. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-021-09872-5.

[14] Matteo Agostini et al. “The Pacific Ocean Neutrino Experiment”. In: *Nature Astronomy* 4.10 (Sept. 2020), pp. 913–915. DOI: 10.1038/s41550-020-1182-4. URL: https://doi.org/10.1038%2Fs41550-020-1182-4.

[15] Christopher Barnes et al. “Challenges, benefits and opportunities in operating cabled ocean observatories: Perspectives from NEPTUNE Canada”. In: May 2013, pp. 1–7. DOI: 10.1109/UT.2011.5774134.

[16] Ocean Networks Canada. *Ocean Networks Canada’s world-leading ocean observatories bring data to the surface, accelerating scientific discovery and delivering solutions that support ocean-and-planet sustainability for future generations.* URL: https://www.oceannetworks.ca (visited on 03/28/2023).

[17] Elisa Resconi. *PLEνM - TOWARDS A PLANETARY NEUTRINO MONITORING SYSTEM*. Oct. 2019. DOI: 10.5281/zenodo.3520454. URL: https://doi.org/10.5281/zenodo.3520454.

[18] M. Bou-Cabo and J.A. Martínez-Mora. “The ANTARES neutrino telescope: Performance one year after its completion”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 617.1 (2010), pp. 505–506. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2009.10.017. URL: https://www.sciencedirect.com/science/article/pii/S0168900209019226.

[19] GVD Collaboration et al. *Neutrino Telescope in Lake Baikal: Present and Future*. 2019. arXiv: 1908.05427 [astro-ph.HE].

[20] M.G. Aartsen et al. “The IceCube Neutrino Observatory: instrumentation and online systems”. In: *Journal of Instrumentation* 12.03 (Mar. 2017), P03012–P03012. DOI: 10.1088/1748-0221/12/03/p03012. URL: https://doi.org/10.1088%2F1748-0221%2F12%2F03%2Fp03012.

[21] Felix Henningsen. “Optical Characterization of the Deep Pacific Ocean: Development of an Optical Sensor Array for a Future Neutrino Telescope”. MA thesis. Technical University of Munich, 2018.

[22] Andreas Gärtner. “Development of an optical sensor system for the characterisation of Cascadia Basin, Canada”. MA thesis. Technical University of Munich, 2018.

[23] G RICCOBENE et al. “Deep seawater inherent optical properties in the Southern Ionian Sea”. In: *Astroparticle Physics* 27.1 (Feb. 2007), pp. 1–9. DOI: 10.1016/j.astropartphys.2006.08.006. URL: https://doi.org/10.1016%2Fj.astropartphys.2006.08.006.

[24] E.G. Anassontzis et al. “Water transparency measurements in the deep Ionian Sea”. In: *Astroparticle Physics* 34.4 (2010), pp. 187–197. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2010.06.008. URL: https://www.sciencedirect.com/science/article/pii/S0927650510001209.

[25] Raymond C. Smith and Karen S. Baker. “Optical properties of the clearest natural waters (200–800 nm)”. In: *Appl. Opt.* 20.2 (Jan. 1981), pp. 177–184. DOI: 10.1364/AO.20.000177. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-20-2-177.

[26] Holzapfel Kilian and Thomson Jol. *Radioamnion*. TUM, SFB1258, ONC, 2023. URL: https://radioamnion.net (visited on 03/26/2023).

[27] Ocean Networks Canada. *Oceans 3.0, SeaTube Pro*. Ocean Networks Canada, 2020. URL: https://data.oceannetworks.ca/SeaTube?resourceTypeId=1000&resourceId=23543&diveId=4180&time=2020-09-11T17:45:18.000Z (visited on 03/26/2023).

[28] Christian Spannfellner. “Development of pathfinder missions and instruments for the Pacific Ocean Neutrino Experiment”. MA thesis. Technical University of Munich, 2020.

[29] Christopher Fink. “Development of a LIDAR System for Characterizing the Optical Properties of Seawater at Cascadian Basin”. Technical University of Munich, 2019.

[30] Eva Laura Winter. “Muon Tracker for the Second Pathfinder of the Pacific Ocean Neutrino Explorer”. Technical University of Munich, 2019.

[31] Ruohan Li. “Development of STRAW-b: Strings for Absorption length in Water for Future Neutrino Telescope in Deep Pacific Ocean”. MA thesis. Technical University of Munich, 2020.

[32] hamamatsu. *Mini-Spectrometer C12666MA*. 2023. URL: https://www.hamamatsu.com/eu/en/product/optical-sensors/spectrometers/mini-spectrometer/C12666MA.html (visited on 03/26/2023).

[33] Immacolata Carmen Rea. "The Pacific Ocean Neutrino Experiment: feasibility study for a new Neutrino Telescope at Cascadia Basin, NE Pacific Ocean". PhD thesis. Technical University of Munich, 2021.

[34] Virginia Gewin. "How to shape a productive scientist– artist collaboration". In: *Nature* (2021). URL: https://cds.cern.ch/record/2835093.

[35] A Neiser et al. "TRB3: a 264 channel high precision TDC platform and its applications". In: *Journal of Instrumentation* 8.12 (Dec. 2013), p. C12043. DOI: 10.1088/1748-0221/8/12/C12043. URL: https://dx.doi.org/10.1088/1748-0221/8/12/C12043.

[36] T.M. Stout and T.J. Williams. "Pioneering work in the field of computer process control". In: *IEEE Annals of the History of Computing* 17.1 (1995), pp. 6–18. DOI: 10.1109/85.366507.

[37] influxdata. *Time series data*. influxdata, 2023. URL: https://www.influxdata.com/what-is-time-series-data/ (visited on 03/26/2023).

[38] influxdata. *Grafana Guide. A Guide to the TIG Stack — Telegraf, InfluxDB, and Grafana*. influxdata, 2023. URL: https://www.influxdata.com/grafana/ (visited on 03/26/2023).

[39] The HDF group. *The HDF Group provides the feature-packed HDF5 technology suite backed by complementary supporting services.* The HDF group, 2023. URL: https://www.hdfgroup.org/ (visited on 03/26/2023).

[40] Ocean Networks Canada. *Oceans 3.0 Data Portal*. Ocean Networks Canada, 2020. URL: https://data.oceannetworks.ca (visited on 03/26/2023).

[41] Dennis Hevesi. "Bryce Bayer, Inventor of a Filter to Make Color Digital Pictures, Dies at 83". In: *The New York Times* (Nov. 29, 2012). URL: https://www.nytimes.com/2012/11/29/business/bryce-bayer-inventor-of-a-filter-to-make-color-digital-pictures-dies-at-83.html (visited on 07/29/2022).

[42] Woosik Kang, Christoph Tönnis, and Carsten Rott. *The camera system for the IceCube Upgrade*. Aug. 21, 2019. DOI: 10.48550/ARXIV.1908.07734. URL: https://arxiv.org/abs/1908.07734.

[43] Arducam. *Arducam*. ARDUCAM TECHNOLOGY CO., LIMITED, 2023. URL: https://www.arducam.com (visited on 03/26/2023).

[44] *Sony Commercializes the World's Highest Sensitivity CMOS Image Sensor for Automotive Cameras*. Sony, Oct. 2014. URL: https://www.sony.com/en/SonyInfo/News/Press/201410/14-106E/ (visited on 07/29/2022).

[45] *IMX225LQR*. English. Version Rev.0.2. Sony. 2014. URL: https://astroccd.org/wp-content/uploads/2016/01/IMX225LQR-C_E_TechnicalDatasheet_Rev0.2.pdf (visited on 07/29/2022).

[46] J.H. Taylor, M.R. Johnson, and C.G. Crawford. *DVD Demystified*. McGraw-Hill, 2006. ISBN: 9780071423984. URL: https://books.google.de/books?id=ikxuL2aX9cAC.

[47] Bryce E. Bayer. "Color Imaging Array". US3971065A. June 1976. URL: https://patents.google.com/patent/US3971065A/en.

[48] OpenCV team. *OpenCV*. OpenCV, 2023. URL: https://opencv.org (visited on 03/26/2023).

[49] LUMINUS. *SST-10-UV Product Datasheet*. LUMINUS, 2022. URL: https://download.luminus.com/datasheets/Luminus_SST-10-UV_Datasheet.pdf (visited on 10/10/2022).

[50] CREE LED. *XLamp XP-E2 LED Data sheet*. CREE LED, 2022. URL: https://assets.cree-led.com/a/ds/x/XLamp-XPE2.pdf (visited on 10/10/2022).

[51] CREE LED. *XLamp XP-G3 LED Data Sheet*. CREE LED, 2022. URL: https://assets.cree-led.com/a/ds/x/XLamp-XPG3.pdf (visited on 10/10/2022).

[52] Texas Instruments. *LM2759 1A Switched Capacitor Flash LED Driver with I2C Compatible Interface*. Texas Instruments, June 2008. URL: https://www.ti.com/lit/ds/symlink/lm2759.pdf?ts=1666863378902&ref_url=https%253A%252F%252Fwww.google.com%252F (visited on 10/10/2022).

[53] W. Shockley. "The theory of p-n junctions in semiconductors and p-n junction transistors". In: *The Bell System Technical Journal* 28.3 (1949), pp. 435–489. DOI: 10.1002/j.1538-7305.1949.tb03645.x.

[54] Thorlabs. *N-BK7 High Precision Windows*. URL: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=1117 (visited on 10/10/2022).

[55] Kenro Miyamoto. "Fish Eye Lens". In: *J. Opt. Soc. Am.* 54.8 (Aug. 1964), pp. 1060–1061. DOI: 10.1364/JOSA.54.001060. URL: http://opg.optica.org/abstract.cfm?URI=josa-54-8-1060.

[56] Ciarán Hughes et al. "Accuracy of fish-eye lens models". In: *Appl. Opt.* 49.17 (June 2010), pp. 3338–3347. DOI: 10.1364/AO.49.003338. URL: http://opg.optica.org/ao/abstract.cfm?URI=ao-49-17-3338.

[57] Zhengyou Zhang. "A Flexible New Technique for Camera Calibration". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (Dec. 2000), pp. 1330–1334. URL: https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/.

[58] R. Tsai. “A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses”. In: *IEEE Journal on Robotics and Automation* 3.4 (1987), pp. 323–344. DOI: 10.1109/JRA.1987.1087109.

[59] Boris Peter Selby et al. “Patient positioning with X-ray detector self-calibration for image guided therapy”. In: *Australasian Physical & Engineering Sciences in Medicine* 34.3 (Aug. 2, 2011), p. 391. DOI: 10.1007/s13246-011-0090-4. URL: https://doi.org/10.1007/s13246-011-0090-4.

[60] Arducam. *Arducam M12 lens kit*. URL: https://www.arducam.com/doc/m12_lens_kit.pdf (visited on 07/30/2022).

[61] Schott. *Schott N-BK 7 517642.251*. URL: https://www.schott.com/shop/medias/schott-datasheet-n-bk7-eng.pdf (visited on 09/24/2022).

[62] “Bernoulli Distribution”. In: *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, 2008, pp. 36–37. ISBN: 978-0-387-32833-1. DOI: 10.1007/978-0-387-32833-1_24. URL: https://doi.org/10.1007/978-0-387-32833-1_24.

[63] Brian C. Ross. “Mutual Information between Discrete and Continuous Data Sets”. In: *PLOS ONE* 9.2 (Feb. 2014), pp. 1–5. DOI: 10.1371/journal.pone.0087357. URL: https://doi.org/10.1371/journal.pone.0087357.

[64] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. “Estimating mutual information”. In: *Phys. Rev. E* 69 (6 June 2004), p. 066138. DOI: 10.1103/PhysRevE.69.066138. URL: https://link.aps.org/doi/10.1103/PhysRevE.69.066138.

[65] Justin Johnson and Taghi Khoshgoftaar. “Survey on deep learning with class imbalance”. In: *Journal of Big Data* 6 (Mar. 2019), p. 27. DOI: 10.1186/s40537-019-0192-5.

[66] Yoshua Bengio. *Practical recommendations for gradient-based training of deep architectures*. 2012. DOI: 10.48550/ARXIV.1206.5533. URL: https://arxiv.org/abs/1206.5533.

[67] Enlu Lin, Qiong Chen, and Xiaoming Qi. *Deep Reinforcement Learning for Imbalanced Classification*. 2019. DOI: 10.48550/ARXIV.1901.01379. URL: https://arxiv.org/abs/1901.01379.

[68] J. Brownlee. *Probability for Machine Learning: Discover How To Harness Uncertainty With Python*. Machine Learning Mastery, 2019. URL: https://books.google.de/books?id=uU2xDwAAQBAJ.

[69] P. Vietti. “New Hull Coatings Cut Fuel Use, Protect Environment”. In: (2009), pp. 36–38. URL: https://web.archive.org/web/20111005011946/http://www.enviro-navair.navy.mil/

currents/fall2009/Fall09_New_Hull_Coatings.pdf (visited on 10/02/2022).

[70] Kevin Long Neal Blossom Dr. Colin Anderson. "The Shift in Antifouling Coating Regulations: from Risk to Efficacy". In: (Nov. 2018). URL: https://www.chemet.com/assets/1/6/The_Shift_in_Antifouling_Coating_Regulations_from_Risk_to_Efficacy__Oct_29_2018.pdf (visited on 10/02/2022).

[71] "Design and field application of a UV-LED based optical fiber biofilm sensor". In: *Biosensors and Bioelectronics* 33.1 (2012), pp. 172–178. ISSN: 0956-5663. DOI: https://doi.org/10.1016/j.bios.2011.12.048. URL: https://www.sciencedirect.com/science/article/pii/S0956566311008700.

[72] Hari Venugopalan. "Photonic Frontiers: LEDs - UVC LEDs reduce marine biofouling". In: (July 2016). URL: https://www.laserfocusworld.com/lasers-sources/article/16547014/photonic-frontiers-leds-uvc-leds-reduce-marine-biofouling (visited on 10/03/2022).

[73] G. Bernhard et al. "Measurements of spectral solar UV irradiance in tropical-Australia". In: *Journal of Geophysical Research: Atmospheres* 102.D7 (1997), pp. 8719–8730. DOI: https://doi.org/10.1029/97JD00072. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/97JD00072. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97JD00072.

[74] R. W. Hart and R. B. Setlow. "Correlation Between Deoxyribonucleic Acid Excision-Repair and Life-Span in a Number of Mammalian Species". In: *Proceedings of the National Academy of Sciences* 71.6 (1974), pp. 2169–2173. URL: https://www.pnas.org/doi/abs/10.1073/pnas.71.6.2169.

[75] Bankanidhi Sahoo et al. "Protein aggregation probed by two-photon fluorescence correlation spectroscopy of native tryptophan". In: *The Journal of chemical physics* 129 (Sept. 2008), p. 075103. DOI: 10.1063/1.2969110.

[76] Great Britain Foreign, Commonwealth Office, and the Stationery Office. *The Stationery Office Annual Catalogue*. On the Control of Harmful Anti-Fouling Systems on Ships. Stationery Office, 2012. ISBN: 978-0-115-02074-2. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/236102/8284.pdf (visited on 10/13/2022).

[77] Shaoyi Jiang and Zhiqiang Cao. "Ultralow-Fouling, Functionalizable, and Hydrolyzable Zwitterionic Materials and Their Derivatives for Biological Applications". In: *Advanced Materials* 22.9 (2010), pp. 920–932. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.200901407.

[78] D. Sommerville. “Development of a Site Specific Biofouling Control Program for the Diablo Canyon Power Plant”. In: *OCEANS ’86*. 1986, pp. 227–231. DOI: 10.1109/OCEANS.1986.1160543.

[79] E. Newton (Edmund Newton) Harvey. *A history of luminescence from the earliest times until 1900*. Philadelphia, American Philosophical Society, 1957, 1957, p. 770. URL: https://www.biodiversitylibrary.org/item/49742.

[80] M. I. Latz, T. M. Frank, and J. F. Case. “Spectral composition of bioluminescence of epipelagic organisms from the Sargasso Sea”. In: *Marine Biology* 3 (June 1988), pp. 441–446. DOI: 10.1007/BF00391120. URL: https://doi.org/10.1007/BF00391120.

[81] Anaid Gouveneaux. “Bioluminescence of Tomopteridae species (Annelida): multidisciplinary approach”. PhD thesis. Aug. 2016.

[82] Peter J. Herring. “The spectral characteristics of luminous marine organisms”. In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 220 (1983), pp. 183–217.

[83] George M. Hale and Marvin R. Querry. “Optical Constants of Water in the 200-nm to 200-μm Wavelength Region”. In: *Appl. Opt.* 12.3 (Mar. 1973), pp. 555–563. DOI: 10.1364/AO.12.000555. URL: https://opg.optica.org/ao/abstract.cfm?URI=ao-12-3-555.

[84] Warren Francis, Meghan Powers, and Steven Haddock. “Characterization of an anthraquinone fluor from the bioluminescent, pelagic polychaete Tomopteris”. In: *Luminescence* 29 (Dec. 2014). DOI: 10.1002/bio.2671.

[85] Ryohei Ito. *Leachia Lesueur*. Ryohei Ito, 2021. URL: http://ryoheiito-photography.com/work/%E3%83%88%E3%82%A6%E3%82%AC%E3%82%BF%E3%82%A4%E3%82%AB-leachia-pacifica/ (visited on 02/27/2023).

[86] Steve Jurvetson. *Two frolicking comb jellies (Beroe spp.) at the Monterey Bay Aquarium.* wikipedia, 2020. URL: https://en.wikipedia.org/wiki/Coelenterata#/media/File:Comb_jellies-mba.jpg (visited on 02/27/2023).

[87] Alexander Semenov. *Coldwater Science*. coldwater.science, 2020. URL: http://coldwater.science/project/tunicata (visited on 02/27/2023).

[88] Séverine Martini and Steven H. D. Haddock. “Quantification of bioluminescence from the surface to the deep sea demonstrates its predominance as an ecological trait”. In: *Scientific Reports* 7.1 (2017), p. 45750. DOI: 10.1038/srep45750. URL: https://doi.org/10.1038/srep45750.

[89] C.F.E. Roper et al. “FAO Species Catalogue, Vol. 3. Cephalopods of the World. An Annotated and Illustrated Catalogue of Species of Interest to Fisheries”. In: *XF2006244397* 3 (Jan. 1984).

[90] Richard E. andKatharina M. Mangold Young and Michael Vecchione. *Leachia Lesueur*. The Tree of Life Web Project, Mar. 2018. URL: http://tolweb.org/Leachia/19544/2018.03.29 (visited on 02/27/2023).

[91] J. Rohr, J. Losee, and G. Anderson. "The Response of Bioluminescent Organisms to Fully Developed Pipe Flow." In: (Aug. 1994). URL: https://apps.dtic.mil/sti/pdfs/ADA291363.pdf.

[92] Elisa Marie Maldonado and Michael I. Latz. "Shear-Stress Dependence of Dinoflagellate Bioluminescence". In: *The Biological Bulletin* 212 (2007), pp. 242–249.

[93] Jim Rohr et al. "Experimental Approaches Towards Interpreting Dolphin-Stimulated Bioluminescence". In: *Journal of Experimental Biology* 201.9 (May 1998), pp. 1447–1460. ISSN: 0022-0949. DOI: 10.1242/jeb.201.9.1447. eprint: https://journals.biologists.com/jeb/article-pdf/201/9/1447/2580097/jexbio_201_9_1447.pdf. URL: https://doi.org/10.1242/jeb.201.9.1447.

[94] Stephan Meighen-Berger, Li Ruohan, and Golo Wimmer. *Bioluminescence modeling for deep sea experiments*. 2021. arXiv: 2103.03816 [q-bio.PE].

[95] Ruslan Abdulkadirov and Pavel Lyakhov. "Estimates of Mild Solutions of Navier and Stokes Equations in Weak Herz-Type Besov and Morrey Spaces". In: *Mathematics* 10 (2022). DOI: 10.3390/math10050680. URL: https://www.mdpi.com/2227-7390/10/5/680.

[96] Nancy Hall. *Drag of a Sphere*. NASA, 2020. URL: https://www1.grc.nasa.gov/beginners-guide-to-aeronautics/drag-of-a-sphere/ (visited on 02/27/2023).

[97] K. Alambra. *Water Viscosity Calculator*. omnicalculator, 2023. URL: https://www.omnicalculator.com/physics/water-viscosity (visited on 03/26/2023).

[98] J. Southard. *Introduction to Fluid Motions and Sediment Transport*. MIT OpenCourseware, 2022. URL: https://geo.libretexts.org/Bookshelves/Sedimentology/Book%3A_Introduction_to_Fluid_Motions_and_Sediment_Transport_(Southard).

[99] Th. v. Karmán. "Über den Mechanismus des Wiederstandes, den ein bewegter Korper in einer Flüssigkeit erfahrt". In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1912 (1912), pp. 547–556. URL: http://eudml.org/doc/58837.

[100] Bob Cahalan. *Kármán vortex street caused by wind flowing around the Juan Fernández Islands off the Chilean coast*. NASA, 1999. URL: http://earthobservatory.nasa.gov/Newsroom/NewImages/images.php3?img_id=3328 (visited on 03/24/2023).

[101] Sal Rodriguez, Pro Domino, and Mohamed El-Genk. “FLUID FLOW AND HEAT TRANSFER ANALYSIS OF THE VHTR LOWER PLENUM USING THE FUEGO CFD CODE”. In: (May 2012).

[102] Kun Il Park. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-68075-0. DOI: 10.1007/978-3-319-68075-0_1. URL: https://doi.org/10.1007/978-3-319-68075-0_1.

[103] Imants G. Priede et al. “The potential influence of bioluminescence from marine animals on a deep-sea underwater neutrino telescope array in the Mediterranean Sea”. In: *Deep Sea Research Part I: Oceanographic Research Papers* 55.11 (2008), pp. 1474–1483. ISSN: 0967-0637. DOI: https://doi.org/10.1016/j.dsr.2008.07.001. URL: https://www.sciencedirect.com/science/article/pii/S0967063708001465.

[104] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2001. ISBN: 9780387952840. URL: https://books.google.de/books?id=VRzITwgNV2UC.