Technische Universität München

Lehrstuhl für Theoretische Informationstechnik

Quantum communication under channel uncertainty

Janis Christian Gregor Nötzel

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. sc. techn. Gerhard Kramer
Prüfer der Dissertation:	1. Univ.-Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
	2. Univ.-Prof. Dr. rer. nat. Michael M. Wolf

Die Dissertation wurde am 12. 04. 2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 06. 09. 2012 angenommen.

Danksagung

An dieser Stelle möchte ich mich bei all denen bedanken, die die Entstehung dieser Arbeit ermöglicht und gefördert haben:

An erster Stelle bei Igor Bjelaković für die große Geduld und Ruhe, die er über Jahre hinweg in all unseren Diskussionen auf- und eingebracht hat und für seine weitsichtigen und realistischen Zielsetzungen, die immer wieder zu spannenden und gleichzeitig lohnenden Fragestellungen führten.
Es war mir eine grosse Freude.

Mein Dank gilt auch Holger Boche für seinen nie versiegenden, vorbildhaften wissenschaftlichen Enthusiasmus, sein Engagement, seine aufmunternden Worte und, natürlich, für die grosse Chance, die er mir durch die Einstellung in seine Arbeitsgruppe gegeben hat.

Auch wenn ich ihn nur kurz kennenlernen konnte, hatte Prof. Ahlswede einen ganz besonderen Einfluss. Die Zusammenarbeit mit ihm hat mir in kürzester Zeit einen elementaren Einblick in etwas ermöglicht, dass ich hier als "wissenschaftliches Denken" bezeichnen möchte. An ihn geht daher ein respektvoller Dank und Gruß, auch wenn er ihn nicht mehr lesen kann.

Weiter möchte ich mich bei Prof. Michael Wolf bedanken dafür, dass er als Zweitgutachter meiner Arbeit fungiert und so meine Promotion an der TU München überhaupt erst ermöglicht. In letzter Minute konnte ich durch das Studium seiner Arbeiten sowie die Zusammenarbeit mit Igor Bjelaković, Holger Boche und Gisbert Janßen noch einen langgehegten Traum in die Tat umsetzen: Eine Anwendung der Extremalitätsbedingungen für vollständig positive Abbildungen von Choi.

Abschliessend möchte ich dem Vorsitzenden des Prüfungsausschusses, Prof. Kramer, für seine Arbeit danken.

Zusammenfassung

Diese Arbeit enthält Resultate sowohl bezüglich Übertragung von Verschränkung sowie Unterräumen als auch Erzeugung von Verschränkung im Grenzwert unendlich vieler Nutzungen zweier Klassen von Quantenkanälen. Die Gemeinsamkeit beider Klassen ist am besten beschrieben als "Kanalunsicherheit". Des weiteren wird die Verbindung zwischen einer der beiden Kanalklassen und der Null-Fehler Kapazität stationärer gedächtnisloser Kanäle untersucht.

Kapitel 3 behandelt das Modell des Compound-Quantenkanals. Dieser besteht aus einer Menge stationärer, gedächtnisloser Quantenkanäle. Jede der oben genannten Aufgabenstellungen wird in einem Zweinutzerszenario (ein Sender, ein Empfänger) bearbeitet. Kommunikation zwischen Sender und Empfänger ist nur in einer Richtung erlaubt.
Ein Code wird nur dann als "gut" betrachtet, wenn er gut ist für jeden einzelnen stationären gedächtnislosen Kanal aus der vorgegebenen Menge.
Es werden drei Grundszenarien unterschieden. Im ersten weiss der Empfänger exakt, welcher Kanal verwendet wird. Im zweiten kennt nur der Sender den Kanal, im dritten ist keiner der Beiden informiert. Für alle neun verschiedenen Kombinationen von Kommunikationsaufgabe und Zustand von Sender und Empfänger wird jeweils die Umkehrung und der direkte Teil eines Kodierungssatzes bewiesen.
Diese Resultate können interpretiert werden als Stetigkeitsaussage für die Kapazitätsfunktion der stationären gedächtnislosen Kanäle auf dem Kodierungslevel.
Die Gleichheit der Kapazitäten für Verschränkungs- und für (starke) Unterraumübertragung wird gezeigt.
Abschliessend wird der Begriff der Symmetrisierbarkeit eines Kanals, der ursprünglich für beliebig variierende klassische Kanäle (AVCs) entwickelt wurde, erweitert auf den Fall des Compound Quantenkanals. Überraschenderweise findet er hier eine nichttriviale Anwendung und beantwortet die Frage, wann genau ein solcher Kanal positive Kapazität für klassische Nachrichtenübertragung hat.

In Kapitel 4 wird der beliebig variierende Quantenkanal (AVQC) untersucht. Eine (unter anderem permutationsinvariante) Menge gedächtnisloser Quantenkanäle wird vorgegeben, und nur diese Menge ist sowohl Sender als auch Empfänger bekannt. Wieder ist nur Vorwärtskommunikation erlaubt, aber diesmal dürfen Sender und Empfänger eine beliebig grosse Menge gemeinsamen Zufalls verwenden. So wie schon im vorhergehenden Modell wird auch hier ein Code nur dann als "gut" betrachtet, wenn er "gut" für jeden Kanal aus der Menge ist.
Aufgrund der erhöhten Komplexität der Situation liegt der Fokus ausschliesslich auf dem Fall, in dem weder Sender noch Empfänger über erweiterte Kanalkenntnis verfügen. Für die drei Kommunikationsszenarien, die sich aus dieser Situation ergeben, wird sowohl die Umkehrung als auch der direkte Teil eines Kodierungssatzes bewiesen.
Durch die Reduktion der verfügbaren Menge an gemeinsamem Zufall auf Null erfolgt schliesslich der Übergang zu den wohlbekannten deterministischen Kodierungsverfahren. Es wird gezeigt das, für den Fall, dass asymptotisch fehlerfreie Übertragung von nur polynomiell vielen klassische Nachrichten (in der Anzahl der Kanalnutzungen) möglich ist, der gemeinsam genutzte Zufall keinen Kapazitätsgewinn bringt, da er durch die Übertragung klassischer Nachrichten simuliert werden kann.
Dies führt zu einer Quantenversion der klassischen Ahlswede-Dichotomie für AVCs.
Ein erstaunliches Resultat in diesem Abschnitt der Arbeit ist die Äquivalenz von (starker) Unterraum- und Verschränkungsübertragungskapazität.
Zwei Fragen werden lediglich angeschnitten: Erstens, wann die deterministische Kapazität für Verschränkungsübertragung über einen AVQC gleich ihrer randomisierten Version ist. Dies beinhaltet die Definition unterschiedlicher Varianten der "Symmetrisierbarkeit", die notwendige und hinreichende bzw. hinreichende Kriterien für die Positivität verschiedener Kapazitäten eines AVQCs liefern. Zweitens wird die Frage aufgeworfen, wann die Kapazitätsformel eine nicht regularisierte Form annimmt. Zwei Bedingungen hierfür werden angegeben.

Ein Nicht-triviales Beispiel eines AVQCs wird anhand des Erasure-AVQCs diskutiert.
Die Verbindung (in Analogie zu der von Ahlswede für AVCs gefundenen) von AVQCs zur Null-Fehler-Verschränkungsübertragung über stationäre gedächtnislose Kanäle wird untersucht. Es wird gezeigt, dass eine einfache Verallgemeinerung nicht möglich ist. Die Arbeit schliesst mit einer kurzen Beschreibung des qualitativen Verhaltens von Null-Fehler-Kapazitäten.

Abstract

These pages summarize results concerning transmission of entanglement and subspaces as well as generation of entanglement in the limit of asymptotically many uses of two classes of quantum channels. The unifying feature of both of these classes is, in plain words, best described as *channel uncertainty*. Additionally, the connection of one of the two classes to the theory of zero-error communication over stationary memoryless quantumchannels is examined.

Chapter 3 starts with the compound quantum channel. This is a collection of stationary, memoryless quantum channels. The task is either one of the above and communication is strictly restricted to forward communication over the compound quantum channel. A code is considered to be good only, if it is good for every single one of the stationary memoryless channels taken from the collection.
Three different situations are investigated. In the first one, the receiver knows exactly which channel is in use. In the second situation, only the sender knows the exact channel. Thirdly, the case where both users are uninformed is considered.
For all the nine possible different combinations of communication task and situation that sender and receiver are in, both the converse and the direct part of a coding theorem are proven.
These results can also be interpreted as a continuity result for the capacity function of stationary memoryless quantum channels on the operational or coding level.
Equality of strong subspace transmission- and entanglement transmission capacity is proven.
The notion of symmetrizability that was developed for classical arbitrarily varying channels in order to tell exactly when a given communication task is possible at a positive rate is extended and, rather surprisingly, finds an application for compound quantum channels.

Chapter 4 considers an even more complex model: The arbitrarily varying quantum channel (AVQC). A specific set of non-stationary, but still memoryless quantum channels is given, and only this set is known to both sender and receiver. Again, only forward communication is allowed, but this time sender and receiver may use an arbitrary amount of shared classical randomness during their communication. Like before, a code is considered to be good only if it performs well for every channel out of the whole set. Due to the increased complexity of the situation, attention is restricted to the case of uninformed users. For the three communication scenarios arising for uninformed users with common randomness, a coding theorem is proven.
By restricting the amount of shared randomness to zero, the connection to the usual deterministic coding schemes is made. Moreover, in case that asymptotically error-free deterministic transmission of only polynomially (in channel uses) many classical messages is possible, it is shown that shared classical randomness is superfluous since it can be simulated by sending classical messages.
This leads to a quantum version of the classical Ahlswede-dichotomy for arbitrarily varying channels.
As a little surprise, it is shown that the strong subspace transmission capacity of an AVQC equals its entanglement transmission capacity.
Two questions are addressed but not completely solved: First, when exactly the deterministic capacity for transmission of entanglement over an AVQC equals the random version thereof. This includes several different notions of symmetrizability which are defined in order to clarify exactly when certain communication tasks are possible at a positive rate. Second, under which conditions a single-letter capacity formula holds. Two conditions for the existence of such a formula are given.
As a nontrivial example, the erasure - arbitrarily varying quantum channel is examined.
The connection to zero-error communication over quantum channels is made and the zero-error capacity of a certain channel (depending on the AVQC one is analyzing) established as a lower bound on the capacity for transmission of entanglement over that particular AVQC. Some light is shed on the qualitative behaviour of zero-error capacities.

Contents

1 Notation and conventions **1**

2 Introduction **4**
2.1 The quantum compound channel . . . 4
2.1.1 Why should we talk about that model at all? . . . 4
2.1.2 Previous work . . . 5
2.1.3 Outline (basic ideas and results for quantum compound channels) . . . 6
2.2 The arbitrarily varying quantum channel . . . 8
2.2.1 A system-theoretic motivation . . . 8
2.2.2 Outline (basic ideas and results for arbitrarily varying quantum channels) . . . 9
2.2.3 Previous work . . . 11

3 The compound quantum channel **12**
3.1 Definitions and main result . . . 12
3.1.1 The informed decoder . . . 12
3.1.2 The informed encoder . . . 13
3.1.3 The case of uninformed users . . . 14
3.1.4 Main result . . . 15
3.2 One-shot results . . . 16
3.2.1 One-shot coding result for a single channel . . . 16
3.2.2 One-shot coding result for uninformed users . . . 17
3.2.3 One-shot coding result for informed encoder . . . 21
3.2.4 Entanglement fidelity . . . 25
3.3 Direct part of the coding theorem for finitely many channels . . . 28
3.3.1 Typical projections . . . 28
3.3.2 Typical kraus operators . . . 29
3.3.3 The case of uninformed users . . . 30
3.3.4 The informed encoder . . . 33
3.4 Finite approximations in the set of quantum channels . . . 34
3.4.1 The compound BSST-lemma . . . 37
3.5 Direct parts of the coding theorems for general compound quantum channels . . . 39
3.5.1 The case of informed decoder and uninformed users . . . 39
3.5.2 The informed encoder . . . 41
3.6 Converse parts of the coding theorems for general quantum compound channels . . . 46
3.6.1 Converse for informed decoder and uninformed users . . . 46
3.6.2 The informed encoder . . . 49
3.7 Continuity of compound capacity . . . 49
3.8 Entanglement-generating capacity of compound channels . . . 50
3.9 Equivalence of strong subspace and entanglement transmission . . . 51
3.10 A symmetrizability condition for compound quantum channels . . . 55

4 The arbitrarily varying quantum channel **57**
4.1 Basic definitions and main results . . . 57
4.1.1 Entanglement transmission . . . 57
4.1.2 Strong subspace transmission . . . 59
4.1.3 Zero-error capacities . . . 60
4.1.4 Entanglement generation . . . 61
4.2 Equivalence of strong subspace and entanglement transmission . . . 62

4.3 Proof of the converse part . . . 63
4.3.1 Converse for the finite AVQC . . . 64
4.3.2 Case $|\mathfrak{I}| = \infty$ . . . 65
4.4 Achievability of entanglement transmission rate I: Random codes . . . 65
4.5 Achievability of entanglement transmission rate II: Derandomization . . . 72
4.6 Zero-capacity-conditions: Symmetrizability . . . 75
4.6.1 Classical capacity with deterministic codes and average error . . . 76
4.6.2 Classical capacity with deterministic codes and maximal error . . . 78
4.6.3 Entanglement transmission capacity with random codes . . . 82
4.7 Conditions for single-letter-capacities . . . 85
4.8 An example and an application to zero-error capacities . . . 87
4.8.1 Erasure-AVQC . . . 87
4.8.2 Qualitative behavior of zero-error capacities . . . 89
4.8.3 Discontinuity of quantum Lovász $\tilde{\theta}$ function & zero-error distillable entanglement . 93
4.9 Entanglement generation . . . 95

5 Conclusions and open problems **97**
5.1 Conclusion for the compound quantum channel . . . 97
5.2 Conclusion for the arbitrarily varying quantum channel . . . 97
5.3 Open problems . . . 97

6 Appendix **100**

7 References **102**

1 Notation and conventions

Hilbert space. All Hilbert spaces are assumed to have finite dimension and are over the field $\mathbb{C}$.

Linear operator. The set of linear operators from $\mathcal{H}$ to $\mathcal{H}$ is denoted $\mathcal{B}(\mathcal{H})$. The **adjoint** to $b \in \mathcal{B}(\mathcal{H})$ is marked by a star and written b^*.

States. $\mathcal{S}(\mathcal{H})$ is the set of states, i.e. positive semi-definite operators with trace 1 acting on the Hilbert space $\mathcal{H}$.
Pure states are given by projections onto one-dimensional subspaces. A vector $x \in \mathcal{H}$ of unit length spanning such a subspace will therefore be referred to as a state vector, the corresponding state will be written $|x\rangle\langle x|$.
To each subspace $\mathcal{F}$ of $\mathcal{H}$ we associate the unique projection $q_{\mathcal{F}}$ whose range is the subspace $\mathcal{F}$ and we write $\pi_{\mathcal{F}}$ for the maximally mixed state on $\mathcal{F}$, i.e. $\pi_{\mathcal{F}} := \frac{q_{\mathcal{F}}}{\mathrm{tr}(q_{\mathcal{F}})}$.

Unit sphere. For an arbitrary Hilbert space $\mathcal{H}$ with inner produckt $\langle \cdot, \cdot \rangle$, we write $S(\mathcal{H})$ for its unit sphere, e.g. $S(\mathcal{H}) := \{x \in \mathcal{H} : \langle x, x \rangle = 1\}$.

Completely positive trace preserving maps. The set of completely positive trace preserving maps (CPTP maps) between $\mathcal{B}(\mathcal{H})$ and $\mathcal{B}(\mathcal{K})$ is denoted $\mathcal{C}(\mathcal{H}, \mathcal{K})$. They are also called channels. The hilbert space $\mathcal{H}$ plays the role of the input system to the channel (traditionally owned by Alice) and $\mathcal{K}$ is the channel's output Hilbert space (usually in Bob's possession).
Completely positive trace decreasing maps. $\mathcal{C}^{\downarrow}(\mathcal{H}, \mathcal{K})$ stands for the set of completely positive trace decreasing maps between $\mathcal{B}(\mathcal{H})$ and $\mathcal{B}(\mathcal{K})$.

Unitary operators. $\mathfrak{U}(\mathcal{H})$ will denote in what follows the group of unitary operators acting on $\mathcal{H}$. For a Hilbert space $\mathcal{G} \subset \mathcal{H}$ we will always identify $\mathfrak{U}(\mathcal{G})$ with a subgroup of $\mathfrak{U}(\mathcal{H})$ in the canonical way.
Projections. A projection $q \in \mathcal{B}(\mathcal{H})$ is an operator satisfying $q = q^2$ *and* $p = p^*$ (non-orhtogonal projections do simply not occur).
For any such projection q we set $q^{\perp} := \mathbf{1}_{\mathcal{H}} - q$.
Each projection $q \in \mathcal{B}(\mathcal{H})$ defines a completely positive trace decreasing map $\mathcal{Q} \in \mathcal{C}(\mathcal{H}, \mathcal{H})$ given by $\mathcal{Q}(a) := qaq$ for all $a \in \mathcal{B}(\mathcal{H})$.
In a similar fashion any $u \in \mathfrak{U}(\mathcal{H})$ defines a $\mathcal{U} \in \mathcal{C}(\mathcal{H}, \mathcal{H})$ by $\mathcal{U}(a) := uau^*$ for $a \in \mathcal{B}(\mathcal{H})$.

Entropy. We use the base two logarithm which is denoted by log. The von Neumann entropy of a state $\rho \in \mathcal{S}(\mathcal{H})$ is given by

$$S(\rho) := -\mathrm{tr}(\rho \log \rho). \tag{1}$$

Coherent information. The coherent information for $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ and $\rho \in \mathcal{S}(\mathcal{H})$ is defined by

$$I_c(\rho, \mathcal{N}) := S(\mathcal{N}(\rho)) - S((id_{\mathcal{H}} \otimes \mathcal{N})(|\psi\rangle\langle\psi|)), \tag{2}$$

where $\psi \in \mathcal{H} \otimes \mathcal{H}$ is an arbitrary purification of the state ρ. Following the usual conventions we let $S_e(\rho, \mathcal{N}) := S((id_{\mathcal{H}} \otimes \mathcal{N})(|\psi\rangle\langle\psi|))$ denote the entropy exchange. A useful equivalent definition of $I_c(\rho, \mathcal{N})$ is given in terms of $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ and a *complementary channel* $\widehat{\mathcal{N}} \in \mathcal{C}(\mathcal{H}, \mathcal{H}_e)$ where $\mathcal{H}_e$ denotes the Hilbert space of an environment: Due to Stinespring's dilation theorem $\mathcal{N}$ can be represented as $\mathcal{N}(\rho) = \mathrm{tr}_{\mathcal{H}_e}(v\rho v^*)$ for $\rho \in \mathcal{S}(\mathcal{H})$ where $v : \mathcal{H} \to \mathcal{K} \otimes \mathcal{H}_e$ is a linear isometry. The complementary channel $\widehat{\mathcal{N}} \in \mathcal{C}(\mathcal{H}, \mathcal{H}_e)$ corresponding to the specific choice of isometry and the given $\mathcal{N}$ is then defined by

$$\widehat{\mathcal{N}}(\rho) := \mathrm{tr}_{\mathcal{H}}(v\rho v^*) \qquad (\forall \rho \in \mathcal{S}(\mathcal{H})). \tag{3}$$

Using a complementary channel, the coherent information can be written as

$$I_c(\rho,\mathcal{N}) = S(\mathcal{N}(\rho)) - S(\widehat{\mathcal{N}}(\rho)) \tag{4}$$

or, equivalently,

$$I_c(\rho,\mathcal{N}) = -I_c(\rho,\widehat{\mathcal{N}}). \tag{5}$$

Fidelity. As a measure of closeness between two states $\rho,\sigma \in \mathcal{S}(\mathcal{H})$ we use the fidelity $F(\rho,\sigma) := ||\sqrt{\rho}\sqrt{\sigma}||_1^2$. The fidelity is symmetric in the input and for a pure state $\rho = |\phi\rangle\langle\phi|$ we have $F(|\phi\rangle\langle\phi|,\sigma) = \langle\phi,\sigma\phi\rangle$.
It is related (see [32]) to the one-norm $\|\cdot\|_1$ by the inequalities

$$1 - \sqrt{F(\rho,\sigma)} \leq \frac{1}{2}\|\rho-\sigma\|_1 \leq \sqrt{1-F(\rho,\sigma)}, \tag{6}$$

which hold independent of $\dim\mathcal{H}$ for arbitrary Hilbert spaces $\mathcal{H}$ and every pair $\rho,\sigma\in\mathcal{S}(\mathcal{H})$.

Entanglement fidelity. A closely related quantity is the entanglement fidelity. For $\rho \in \mathcal{S}(\mathcal{H})$ and $\mathcal{N}\in\mathcal{C}^{\downarrow}(\mathcal{H},\mathcal{H})$ it is given by

$$F_e(\rho,\mathcal{N}) := \langle\psi,(id_{\mathcal{H}}\otimes\mathcal{N})(|\psi\rangle\langle\psi|)\psi\rangle, \tag{7}$$

with $\psi\in\mathcal{H}\otimes\mathcal{H}$ being an arbitrary purification of the state ρ.

Diamond norm. For the approximation of arbitrary sets of channels by finite sets we use the diamond norm $||\cdot||_\diamond$, which is given by

$$||\mathcal{N}||_\diamond := \sup_{n\in\mathbb{N}} \max_{a\in\mathcal{B}(\mathbb{C}^n\otimes\mathcal{H}),||a||_1=1} ||(\mathrm{id}_n\otimes\mathcal{N})(a)||_1, \tag{8}$$

where $\mathrm{id}_n : \mathcal{B}(\mathbb{C}^n)\to\mathcal{B}(\mathbb{C}^n)$ is the identity channel, and $\mathcal{N}:\mathcal{B}(\mathcal{H})\to\mathcal{B}(\mathcal{K})$ is any linear map, not necessarily completely positive.
The merits of $||\cdot||_\diamond$ are due to the following facts (cf. [46]). First, $||\mathcal{N}||_\diamond = 1$ for all $\mathcal{N}\in\mathcal{C}(\mathcal{H},\mathcal{K})$. Thus, $\mathcal{C}(\mathcal{H},\mathcal{K})\subset S_\diamond$, where $S_\diamond$ denotes the unit sphere of the normed space $(\mathcal{B}(\mathcal{B}(\mathcal{H}),\mathcal{B}(\mathcal{K})),||\cdot||_\diamond)$. Moreover, $||\mathcal{N}_1\otimes\mathcal{N}_2||_\diamond = ||\mathcal{N}_1||_\diamond||\mathcal{N}_2||_\diamond$ for arbitrary linear maps $\mathcal{N}_1,\mathcal{N}_2:\mathcal{B}(\mathcal{H})\to\mathcal{B}(\mathcal{K})$. Finally, the supremum in (8) needs only be taken over n that range over $\{1,2,\ldots,\dim\mathcal{H}\}$.

Norm closure. For a set $\mathfrak{I}\subset\mathcal{C}(\mathcal{H},\mathcal{K})$ we write $\overline{\mathfrak{I}}$ for its closure in $\|\cdot\|_\diamond$.

Distance between sets of channels. We use the diamond norm to define the function $D_\diamond(\cdot,\cdot)$ on $\{(\mathfrak{I},\mathfrak{I}') : \mathfrak{I},\mathfrak{I}'\subset\mathcal{C}(\mathcal{H},\mathcal{K})\}$, which is for $\mathfrak{I},\mathfrak{I}'\subset\mathcal{C}(\mathcal{H},\mathcal{K})$ given by

$$D_\diamond(\mathfrak{I},\mathfrak{I}') := \max\{\sup_{\mathcal{N}\in\mathfrak{I}}\inf_{\mathcal{N}'\in\mathfrak{I}'}||\mathcal{N}-\mathcal{N}'||_\diamond,\ \sup_{\mathcal{N}'\in\mathfrak{I}'}\inf_{\mathcal{N}\in\mathfrak{I}}||\mathcal{N}-\mathcal{N}'||_\diamond\}. \tag{9}$$

The function $D_\diamond$, when restricted to $\{\mathfrak{I}:\mathfrak{I}\subset\mathcal{C}(\mathcal{H},\mathcal{K}),\ \mathfrak{I}=\overline{\mathfrak{I}}\}$, defines a metric which is basically the Hausdorff distance induced by the diamond norm.
Obviously, for arbitrary $\mathfrak{I},\mathfrak{I}'\subset\mathcal{C}(\mathcal{H},\mathcal{K})$, $D_\diamond(\mathfrak{I},\mathfrak{I}')\leq\epsilon$ implies that for every $\mathcal{N}\in\mathfrak{I}$ ($\mathcal{N}'\in\mathfrak{I}'$) there exists $\mathcal{N}'\in\mathfrak{I}'$ ($\mathcal{N}\in\mathfrak{I}$) such that $||\mathcal{N}-\mathcal{N}'||_\diamond\leq 2\epsilon$. If $\mathfrak{I}=\overline{\mathfrak{I}}$, $\mathfrak{I}'=\overline{\mathfrak{I}'}$ holds we even have $||\mathcal{N}-\mathcal{N}'||_\diamond\leq\epsilon$. In this way $D_\diamond$ gives a measure of distance between two compound channels.

For any set $\mathfrak{I}\subset\mathcal{C}(\mathcal{H},\mathcal{K})$ and $l\in\mathbb{N}$ we set

$$\mathfrak{I}^{\otimes l} := \{\mathcal{N}^{\otimes l} : \mathcal{N}\in\mathfrak{I}\}. \tag{10}$$

Probability distributions. For a finite set A the notation $\mathfrak{P}(A)$ is reserved for the set of probability distributions on A.

Cardinality of a set. For a set A, $|A|$ denotes its cardinality.

Cartesian product of sets. For an arbitrary set $\mathbf{S}$ and $l \in \mathbb{N}$, $\mathbf{S}^l := \{(s_1, \ldots, s_l) : s_i \in \mathbf{S}\ \forall i \in \{1, \ldots, l\}\}$. We also write s^l for the elements of $\mathbf{S}^l$.

Convex hull. For an arbitrary set $\mathfrak{I}$ of CPTP maps we denote by $\mathrm{conv}(\mathfrak{I})$ its convex hull (see [68] for the definition) and note that in case that $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ is a finite set we have

$$\mathrm{conv}(\mathfrak{I}) = \left\{ \mathcal{N}_q \in \mathcal{C}(\mathcal{H}, \mathcal{K}) : \mathcal{N}_q = \sum_{s\in\mathbf{S}} q(s)\mathcal{N}_s,\ q \in \mathfrak{P}(\mathbf{S}) \right\}, \tag{11}$$

an equality that we will make use of in the approximation of infinite AVQC's by finite ones.

Relative interior. Finally, we need some simple topological notions for convex sets in finite dimensional normed space $(V, ||\cdot||)$ over the field of real or complex numbers which we borrow from [68]. Let $F \subset V$ be convex. $x \in F$ is said to be a relative interior point of F if there is $r > 0$ such that $B(x, r) \cap \mathrm{aff}\, F \subset F$. Here $B(x, r)$ denotes the open ball of radius r with the center x and $\mathrm{aff}\, F$ stands for the affine hull of F. The set of relative interior points of F is called the relative interior of F and is denoted by $\mathrm{ri}\, F$.

Relative boundary. The relative boundary of F, $\mathrm{rebd}\, F$, is the set difference between the closure of F and $\mathrm{ri}\, F$.

Blow-up. For a set $A \subset V$ and $\delta \geq 0$ we define the parallel set or the blow-up $(A)_\delta$ of A by

$$(A)_\delta := \{x \in V : ||x - y|| \leq \delta \text{ for some } y \in A\}. \tag{12}$$

2 Introduction

The following two sections motivate, describe and put into a historical context the two channel models that will be dealt with in the rest of this thesis. Although the compound quantum channel later on serves as a building block in the derivation of our results for arbitrarily varying quantum channels, the situation described by that model is interesting in its own right, as will be made clear from our argumentation.

2.1 The quantum compound channel

2.1.1 Why should we talk about that model at all?

Channel uncertainty is a big issue in the design of reliable communication systems. As an example, let us take a look at the rather simple scenario that arises from considering an optical fibre which is, for communication purposes, described by a handful of mathematical parameters. Each single one of these parameters will be known only with a certain precision due to measurement errors. Any coding scheme used with this fibre will thus have to work for every possible set of parameters within some specified range, and the true value of the parameters remains unknown. This is a possible example for a compound quantum channel.
One could expect that, at least in principle, it should be possible to gain perfect channel knowledge. We will argue below that this is not the case. This shows that the compound quantum channel not just merely serves as a building block in the proof of achievability of the random entanglement transmission capacity of the arbitrarily varying quantum channel, but rather is an important model in its own right.
Consider the following scenario. Sender and receiver are given some channel for forward communication (This might for example be a USB cable and the sender wants to save some data on a portable harddrive belonging to the receiver). They are guaranteed that, for the task of sending classical information over it, it can be described by either one of two stochastic matrices $(W_1(b|a))_{a\in\mathbf{A},b\in\mathbf{B}}$ or $(W_2(b|a))_{a\in\mathbf{A},b\in\mathbf{B}}$ (where both $\mathbf{A}$ and $\mathbf{B}$ are finite sets) and, clearly, $W_2 \neq W_1$.
Now sender and receiver talk to each other and they both agree that this is unbearable. In fact, they would like it much better to *perfectly* know which channel it is *before* the start of their communication task. We assume that this kind of direct communication between them is costly. Therefore, they want to set up some procedure that, using forward communication over the channel, figures out which of the two possible descriptions is the proper one (at receivers side). After that, the receiver tells the sender the index (1 or 2) of the proper description, and they start their second task (the one they originally intended to perform: saving data to the portable harddrive).
Although in this work we shall be concerned with variants of the second task only, let us briefly formalize what they are up to now in order to fulfill their first task:

1. Find the smallest $l \in \mathbb{N}$ such that there exist $a^l \in \mathbf{A}^l$, $D_1, D_2 \subset \mathbf{B}^l$, $D_1 \bigcap D_2 = \emptyset$ such that for $i = 1, 2$ we get
2. $\sum_{b^l\in D_i} \prod_{k=1}^{l} W_i(b_k|a_k) = 1$.

For sake of simplicity, let $\mathbf{A} = \{a\}$ (yes, this *is* trivial). Then, setting $p_1(b) := W_1(b|a)$, $p_2(b) := W_2(b|a)$ (for all $b \in \mathbf{B}$) and $p_i^{\otimes l}(X) := \sum_{b\in\mathbf{X}} \prod_{i=1}^{l} p_i(b_i)$ (for all $\mathbf{X} \subset \mathbf{B}^l$) we see that our task turns into

1'. Find the smallest $l \in \mathbb{N}$ such that there exist $D_1, D_2 \subset \mathbf{B}^l$, $D_1 \bigcap D_2 = \emptyset$ such that
2'. $p_i^{\otimes l}(D_i) = 1$ for $i = 1, 2$.

Let us have a look at a reformulation of Corollary 1.2 in [20]:

Theorem 1 (Stein's Lemma)**.** *For every* $0 < \varepsilon < 1$,

$$\lim_{l\to\infty} \frac{1}{l} \log(\min_{D_1 \subset \mathbf{B}^l : p_1^{\otimes l}(D_1) \geq 1-\varepsilon} p_2^{\otimes l}(D_1)) = -\sum_{b\in\mathbf{B}} p_1(b) \log(\frac{p_1(b)}{p_2(b)}). \tag{13}$$

We make the technical assumption that $p_1(b) > 0 \Rightarrow p_2(b) > 0$ $(\forall b \in \mathbf{B})$. Take some fixed $\varepsilon \in (0,1)$. Then by the above Theorem and our assumption there is $c > 0$ such that for all large enough l, for every set D_1 such that $p_1^{\otimes l}(D_1) \geq 1 - \varepsilon$ holds, and for every D_2 with $D_2 \bigcap D_1 = \emptyset$ we get

$$p_2^{\otimes l}(D_2) = 1 - p_2^{\otimes l}(D_2^{\complement}) \leq 1 - p_2^{\otimes l}(D_1) \leq 1 - 2^{-lc}. \tag{14}$$

This should convince the reader that our task is not possible at all in general. The best result we can hope for in general is that with *probability* exponentially close (in l) to one we can be sure that the index i is correctly identified.
Assuming that each use of the channel takes some time we see, that we can in fact *never* be totally sure in our whole life which channel we are transmitting over.
Let us switch to a more general situation, where an arbitrary possibly infinite set of stochastic matrices is the set of possible proper descriptions of the channel. Then by the above discussion, the best we should look out for is that (without giving or intending to give any proof here, since this would lead us far away from our original goal), again, with *probability* exponentially close (in l) to one, the *compound* channel we will be transmitting over after a test taking the time of l channel uses is given (for some stochastic matrix $\overline{W}$) by a set $\mathcal{W} = \{W(\cdot|\cdot) : \sum_{b\in\mathbf{B}} |W(b|a) - \overline{W}(b|a)| \leq \delta_l \ \forall a \in \mathbf{A}\}$, where $(\delta_l)_{l\in\mathbb{N}}$ is a sequence satisfying $\delta_l \searrow 0$.
Now that the compound channel is fully established as a fundamental model for communication, let us get into some more detail on this object. Especially, from now on our basic objects will be *quantum* channels.

2.1.2 Previous work

The determination of capacities of quantum channels in various settings has been a field of intense work over the last decade. To any quantum channel we can associate a complete zoo of different notions of capacity depending on what is to be achieved by transmitting something over the channel and which figure of merit is chosen as the criterion for the success of the particular quantum communication task. For example we may try to determine the maximum number of classical messages that can be reliably distinguished at the output of the channel. This leads to the notion of classical capacity of a quantum channel.
Instead, we might wish to establish secure classical communication over a quantum channel, giving rise to the definition of a channel's private capacity.
In both cases, the two commonly used measures of success for transmission of classical messages are average- and maximal error probability.
On the other hand, in the realm of quantum communication, one may also ask what the maximal amount of entanglement is that can be generated or transmitted over a given quantum channel, leading to the notions of entanglement generation and entanglement transmission capacity. Other examples of quantum capacities are strong subspace transmission and average subspace transmission capacities.
Such quantum communication tasks are needed, for example, to support computation in quantum circuits or to provide the best possible supply of pure entanglement in a noisy environment. Fortunately, these genuinely quantum mechanical capacities are shown to be equal for perfectly known single user channels [10], [50].
Most of the work done so far on quantum channel capacities relies on the assumption that the channel is *perfectly known* to sender and receiver, and as we just found out, this is hardly ever the case in any

application.
First results indicating that coherent information was to play a role in the determination of the quantum capacity of memoryless channels were established by Schumacher and Nielsen [60] and, independently, by Lloyd [53] who was the first to conjecture that indeed the regularized coherent information would give the correct formula for the quantum capacity and gave strong heuristic evidence to his claim. In 1998 and 2000 Barnum, Knill, and Nielsen and Barnum, Nielsen, and Schumacher [10], [11] gave the first upper bound on the capacity of a memoryless channel in terms of the regularized coherent information. Later on, Shor [65] and Devetak [24] offered two independent approaches to the achievability part of the coding theorem. Despite the fact that the regularized coherent information was identified as the capacity of memoryless quantum channels many other approaches to the coding theorem have been offered subsequently, for example Devetak and Winter [26] and Hayden, Shor, and Winter [36]. Of particular interest for the present work were the developments by Klesse [47] and Hayden, Horodecki, Winter, and Yard [35] based on the decoupling idea which can be traced back to Schumacher and Westmoreland [62]. In fact, the main purpose of this work was to show that the decoupling idea can be utilized to prove the existence of reliable universal quantum codes for entanglement transmission and generation.
The capacity for transmission of classical messages over memoryless quantum channels has been determined in the pioneering work by Holevo [37], [38] and Schumacher and Westmoreland [61]. Their results have been substantially sharpened by Winter [70] and Ogawa and Nagaoka [57] who gave independent proofs of the strong converse to the coding theorem.
The capacity of compound channels in the classical setting was determined by Wolfowitz [72, 73] and Blackwell, Breiman, and Thomasian [17]. The full coding theorem for transmission of classical information via compound quantum channels was proven in [14]. Subsequently, Hayashi [34] obtained a closely related result with a completely different proof technique based on the Schur-Weyl duality from representation theory and the packing lemma from [20].

2.1.3 Outline (basic ideas and results for quantum compound channels)

While the classical capacity of compound quantum channels has been determined only recently in [14], in the first part of this work the focus will be on entanglement-generation, entanglement transmission and strong subspace transmission capacities.
All three of them will be shown to be equal and a regularized formula not unlike that for a single memoryless channel will be given.
Let us now dive a little deeper into the technical issues.
The underlying idea of the coding theorems for the compound quantum channel is, as in the case of a single memoryless channel: decoupling.
This fundamental idea is explained in great detail in [28]. We give a brief scetch as follows. For a channel $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ and a state $\rho \in \mathcal{S}(\mathcal{H})$, take an environment E, a stinespring isometry W for that particular environment, the corresponding complimentary channel $\widehat{\mathcal{N}} \in \mathcal{C}(\mathcal{H}, E)$ and a purification $\psi \in \mathcal{H}' \otimes \mathcal{H}$ of ρ, where $\mathcal{H}'$ is just a copy of $\mathcal{H}$. If $id_{\mathcal{H}'} \otimes \widehat{\mathcal{N}}(|\psi\rangle\langle\psi|)$ is almost a product state, then Uhlmanns Theorem [66] guarantees the existence of a unitary U acting on $\mathcal{K} \otimes E'$ for some cleverly chosen second environment E' such that a recovery $\mathcal{R} \in \mathcal{C}(\mathcal{K}, \mathcal{H})$ defined using an isometry V in the stinespring representation of $\mathcal{R}$ which is itself induced by U allows approximate transmission of entanglement over $\mathcal{N}$.
More informally, write $\mathcal{W}(\cdot) := W(\cdot)W^*$. If we start from a highly correlated pure state $|\psi\rangle\langle\psi|$ and transform it to $id \otimes \mathcal{W}(|\psi\rangle\langle\psi|)$, then this state must still hold strong correlations between $E \otimes \mathcal{K}$ and $\mathcal{H}$. So, if its reduction to $E \otimes \mathcal{H}$ is almost decorrelated, then all the correlations must be in $\mathcal{K} \otimes \mathcal{H}$, hence there must be a way to recover them. A thorough application of this idea can be found in [47].
We derive two modification of Klesse's one-shot coding result [47] that are adapted to arithmetic averages of channels in section 3.2. One guarantees existence of good codes for uninformed users, the other for an informed encoder. By noting that entanglement fidelity is affine in the channel, this delivers a tool to (asymptotically) cope with finite compound channels in these two cases.

There is an issue with the dependence of the so derived bounds on the block length in section 3.3: in order to get the correct asymptotics, we have to project onto typical subspaces of suitable output states of the individual channels. Therefore, it turns out that we effectively end up in the scenario with *informed decoder*.
Luckily, we found that these projections can simply be removed without decreasing the entanglement fidelity too much (see the end of section 3.2) and we really get universal (i.e. uninformed) decoders for our coding problems. The direct part for the informed decoder follows directly from that for uninformed users, since by the converse part their capacity is upper bounded by the same number.
In order to get the correct capacity formula, one has to pass from a maximization over maximally mixed states to one over arbitrary states. This is in the case of a single memoryless channel done by an application of the BSST Lemma [12]. In our case, an additional minimization over the set of channels prevents a direct application of the BSST Lemma. Therefore, an appropriate generalization (called compound BSST Lemma) is provided in section 3.4.
This generalization is possible thanks to discretization and approximation techniques based on τ-nets in the set of quantum channels that get applied also when passing from finite to infinite compound channels in section 3.5, which also contains an application of the continuity result [51] for the entanglement transmission capacity of the stationary memoryless channel.
This continuity result is used as well in order to establish a corresponding result for the entanglement transmission capacity of compound quantum channels in section 3.7. Once we know this capacity, it is only a short distance of two pages to get the corresponding result for entanglement generation.
From there, we pass to the equivalence of strong subspace and entanglement transmission. The corresponding fundamental Lemma 66 can in its original form be found in the work [5] done together by R. Ahlswede, I. Bjelakovic, H. Boche and the author. Its proof exploits concentration phenomena on the unit sphere of high dimensional Hilbert spaces and the fact that the success criterion entanglement fidelity can approximately be seen as an averaged form of the success criterion strong subspace transmission [42].
Finally, the notion of symmetrizability is defined for compound quantum channels in section 3.10. This definition gives us a criterion saying exactly when the capacity for transmission of classical messages using the average error criterion of a compound quantum channel equals zero. Symmetrizability has originally been developed to cope with the zero-capacity question in the case of arbitrarily varying channels and this is the reason why the proof of the statement of this section is delayed until section 4.6.
At this point, it has to be said that symmetrizability is, in its present form, comparable to someone trying to eat without knowing what food is. We are trying to give criteria for the entanglement transmission capacity of highly complex channel models to be equal to zero (or some other number), without even knowing when this is the case for such simple stationary memoryless channels as the depolarizing channel. Or, to put it in the words of Ahlswede, when he started questioning his collaborators about known and unknown facts, finding out that there was no simple criterion telling exactly when the entanglement transmission capacity of a single, stationary and memoryless quantum channel equals zero: "Wenn wir noch nicht einmal *das* wissen...".
Maybe, saying that he had in mind the classical symmetrizability criterion (Separation Lemma) from [2] underlying that given for AVQCs in subsection 4.6.2 for classical messages and the maximal error probability criterion. The proof connecting the definition to its statement about the corresponding capacity explicitly uses the fact that the class of binary symmetric channels is very well understood - something we can only dream of for its counterpart, the depolarizing channels.
This clearly is an area for future research. A more detailed description of the problem will be given in section 5.1.

2.2 The arbitrarily varying quantum channel

2.2.1 A system-theoretic motivation

The second part of the thesis contains work of R. Ahlswede, I. Bjelakovic, H. Boche and the author. Let us start with a rather simple situation. We are given three parties - a legitimate sender and receiver ($\mathfrak{S}_1$ and $\mathfrak{R}_1$) and an evil guy ($\mathfrak{S}_2$), who is trying to interrupt the communication between $\mathfrak{S}_1$ and $\mathfrak{R}_1$. The situation can be described in mathematical terms by Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ and $\mathcal{K}_1$ which are accessed by $\mathfrak{S}_1$, $\mathfrak{S}_2$ and $\mathfrak{R}_1$, while the connection between them is given by a channel $\mathcal{N} \in \mathcal{C}(\mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{K}_1)$.
For arbitrary $l \in \mathbb{N}$ and $\rho \in \mathcal{S}(\mathcal{H}_2^{\otimes l})$ we further define the channels

$$\mathcal{N}_\rho^l(\cdot) := \mathcal{N}^{\otimes l}(\cdot \otimes \rho). \tag{15}$$

Using these channels, $\mathfrak{S}_1$ will try to send one half of a maximally entangled state to $\mathfrak{R}_1$, no matter what the bad guy $\mathfrak{S}_2$ puts in on his side of the channel. There are no restrictions to the powers of $\mathfrak{S}_2$, except that he can only access his Hilbert space $\mathcal{H}_2$.

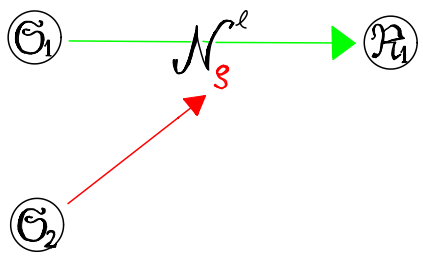

Figure 1: The full model of an arbitrarily varying quantum channel (The good, the bad and the receiver).

A coarse description of the usual (quantum) Shannon-information theoretic setup is given by the following task:
Given a sequence $(\pi_{\mathcal{F}_l})_{l\in\mathbb{N}}$ of maximally mixed states on Hilbert spaces $\mathcal{F}_l$ satisfying $\liminf_{l\to\infty} \frac{1}{l} \log \dim \mathcal{F}_l = R \in \mathbb{R}_+$, show the existence of sequences of encoding and recovery maps $(\mathcal{P}^l)_{l\in\mathbb{N}}$, $(\mathcal{R}^l)_{l\in\mathbb{N}}$ such that

$$\lim_{l\to\infty} \inf_{\rho\in\mathcal{S}(\mathcal{H}_2^{\otimes l})} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_\rho^l \circ \mathcal{P}^l) = 1. \tag{16}$$

The structure of the sets

$$\{\mathcal{N}_\rho^l\}_{\rho\in\mathcal{S}(\mathcal{H}_2^{\otimes l})} \qquad (l \in \mathbb{N}) \tag{17}$$

being rather complex, we reduce the abilities of $\mathfrak{S}_2$ - he shall be given a subset $\mathbf{S}_2 \subset \mathcal{S}(\mathcal{H}_2)$ and his inputs get restricted to states taken from the sets

$$\mathbf{S}_{2,l} := \{\rho_1 \otimes \ldots \otimes \rho_l : \rho_i \in \mathbf{S}_2 \ \forall i \in \{1, \ldots, l\}\} \qquad (l \in \mathbb{N}). \tag{18}$$

It is evident that (16) then reduces to

$$\lim_{l\to\infty} \inf_{s^l\in\mathbf{S}^l} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ (\otimes_{i=1}^l \mathcal{N}_{s_i}) \circ \mathcal{P}^l) = 1, \tag{19}$$

where $\mathbf{S}$ is an index set for $\mathbf{S}_2$ (implying $\mathbf{S}_2 = \{\rho_s\}_{s\in\mathbf{S}}$) and $\mathcal{N}_s := \mathcal{N}^1_{\rho_s}$ for all $s \in \mathbf{S}$. In the classical case, due to non-existence of entanglement, the restriction (18) is not necessary at all and the approach directly leads to the model of an AVC as introduced in [18].
In our case, the full model will be stated as an open problem in section 5.2.

A slightly different look at the situation stemming from todays cellular networks is given by the following. We assume that there are two non-cooperating senders $\mathfrak{S}_1$ and $\mathfrak{S}_2$, each of which is transmitting quantum states to a corresponding receiver ($\mathfrak{R}_1$ and $\mathfrak{R}_2$). Usually, this leads to 'interference': The actions of $\mathfrak{S}_2$ have an influence on the communication between $\mathfrak{S}_1$ and $\mathfrak{R}_1$ and vice versa.
Let us concentrate on the first communication link and simply regard the second one as some complicated form of noise. Our starting point is a stationary memoryless quantum channel specified by

$$\mathcal{E} \in \mathcal{C}(\mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{K}_1 \otimes \mathcal{K}_2), \tag{20}$$

where the senders $\mathfrak{S}_i$ have access to $\mathcal{H}_i^{\otimes l}$ and the receivers act on the output Hilbert spaces $\mathcal{K}_i^{\otimes l}$, $i = 1, 2$, $l \in \mathbb{N}$. Since we concentrate on the first link, our basic model is the (stationary and memoryless) channel $\mathcal{N} \in \mathcal{C}(\mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{K}_1)$ defined by

$$\mathcal{N} := \mathrm{tr}_{\mathcal{K}_2} \circ \mathcal{E}. \tag{21}$$

Again, for arbitrary $l \in \mathbb{N}$ and $\rho \in \mathcal{S}(\mathcal{H}_2^{\otimes l})$ we define the channels

$$\mathcal{N}^l_\rho(\cdot) := \mathcal{N}^{\otimes l}(\cdot \otimes \rho), \tag{22}$$

and once more, $\mathfrak{S}_1$ will try to send one half of a maximally entangled state to $\mathfrak{R}_1$, no matter what $\mathfrak{S}_2$ puts in on his side of the channel.

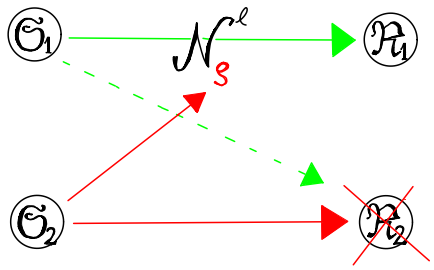

Figure 2: How to derive the model of an arbitrarily varying quantum channel from a stationary memoryless one with four users.

In this model, $\mathfrak{S}_1$ and $\mathfrak{R}_1$ may come to the conclusion that $\mathfrak{S}_2$, although not being helpful at all, is at least following basic rules or standards that have been predefined. Thus, they may conclude, there will be certain states that $\mathfrak{S}_2$ will never send. This could lead to the model of an arbitrarily varying quantum channel with state constraints.

For now, using our reduced quantum model, we are led to the following setup:

2.2.2 Outline (basic ideas and results for arbitrarily varying quantum channels)

There is a set of quantum channels $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ which is known to both the (legitimate) sender and receiver ($\mathfrak{S}_1$ and $\mathfrak{R}_1$). The goal of the sender is to transmit one half of a maximally entangled pure

state ψ, suitably encoded, by l-fold usage of the (unknown) channel. An entity ($\mathfrak{S}_2$), which we now call the adversary for simplicity, can choose a sequence $s^l = (s_1, \ldots, s_l) \in \mathbf{S}^l$ at her/his will which results in the selection of the channel $\mathcal{N}_{s^l} = \otimes_{i=1}^{l} \mathcal{N}_{s_i}$. The encoded version of ψ is then fed into $\mathcal{N}_{s^l}$ and the receiver's goal is to recover the input state, of course without knowing the sequence s^l being selected by the adversary. Implicit in this informal description of the communication scenario is that we suppose that the adversary knows the code which is used for entanglement transmission. Therefore, the communicators are forced to use entanglement transmission protocols that are reliable for the whole family $\mathfrak{I}^{(l)} = \{\mathcal{N}_{s^l}\}_{s^l \in \mathbf{S}^l}$ of memoryless and partly non-stationary channels. In other words, the desired entanglement transmission protocol should be resistant to the effect of arbitrarily varying noise represented by the family $\mathfrak{I}^{(l)} = \{\mathcal{N}_{s^l}\}_{s^l \in \mathbf{S}^l}$. Even in the simplest non-trivial case of a finite set $\mathfrak{I}$ with $|\mathfrak{I}| > 1$ we have to deal for each block length l with exponentially many quantum channels simultaneously.
The main contribution of this second part of the thesis is a generalization of Ahlswede's dichotomy [2] which can be stated as follows:
First, the common-randomness-assisted entanglement transmission capacity of the AVQC $(\mathfrak{I}^{(l)})_{l \in \mathbb{N}}$ is equal to the entanglement transmission capacity of the compound channel built up from $\mathrm{conv}(\mathfrak{I})$, i.e. the uncountable family of stationary, memoryless channels that lie in the convex hull of $\mathfrak{I}$.
Second, if the deterministic capacity for transmission of messages with asymptotically vanishing average error over an AVQC is greater than zero, its capacity for transmission of entanglement with deterministic codes is equal to its common-randomness-assisted capacity for transmission of entanglement.

The proof of the direct part as well as the proof of the converse rely substantially on the corresponding results for compound quantum channels developed in the first part of the thesis. The link between the compound and arbitrarily varying channel models needed in the achievability proofs is given by the powerful robustification technique of [3] and [4], which is stated as Theorem 94 in section 4.4.
The idea behind the second part of the theorem is the following. If the deterministic capacity for message transmission, with average error probability as the success criterion, of $\mathfrak{I}$ is greater than zero, then sender and receiver can use a few (sub-exponentially many) bits to derandomize a given common-randomness-assisted code for transmission of entanglement. A mathematically rigorous treatment of this idea is found in section 4.5.
As a supplement to the coding theorem, we derive a multi-letter necessary and sufficient condition for the deterministic capacity, with average error, for message transmission of a (finite) AVQC to be zero in section 4.6. For sake of completeness, we also include a necessary and sufficient condition for the deterministic capacity for message transmission with *maximal* error probability to be equal to zero. Moreover, we present a first attempt to derive a non-trivial sufficient condition for the common-randomness-assisted capacity for transmission of entanglement to be zero, which we call qc-symmetrizability. Our feeling in this matter is that the definition of that kind of symmetrizability is too narrow to have any chance to be necessary and sufficient. This is basically because according to that definition the adversary does not use all the freedom he is given by the channel model to prevent the common-randomness-assisted entanglement transmission.
The most surprising result included here is a striking difference to the classical theory: entanglement transmission with entanglement fidelity as the criterion of success is widely acknowledged as a fully quantum counterpart to message transmission with average error as a criterion for failure of transmission, while the counterpart of strong subspace transmission should be maximal error probability.
The two classical criteria have been proven to be asymptotically equivalent e.g. for single memoryless channels. For transmission over an AVC they lead to different capacities, as can be seen from Example 2 in [2]. The AVC given there has zero capacity for message transmission with asymptotically vanishing maximal error probability, but from Theorem 3, part a) it can be seen that it has positive capacity for message transmission with asymptotically vanishing average error.
In the quantum case, asymptotic equivalence of entanglement and strong subspace transmission for single quantum channels has already been proven in [10]. Our results from section 4.2 show, that they are - in

contrast to the classical theory - also (asymptotically) equivalent criteria w.r.t. AVQCs.
It is no surprise then, that the connection between arbitrarily varying channels and zero-error capacities that is valid in the classical case [1] only partly survives in the quantum regime. This connection is explored in the last part of the paper. It is a personal interest of the author to point out that it contains one of only few applications of the extremality condition for completely positive maps that was developed by Choi in [19]. Additionally, we show in section 4.8 that quantum, classical, and entanglement-assisted zero-error capacities of quantum channels are generically zero and are discontinuous at every positivity point. This is obvious for the classical zero-error capacity in Shannon's original setting [63]. In the quantum case we employ some simple facts from convex geometry combined with methods motivated by the theory of arbitrarily varying channels to obtain this conclusion in an extremely simple way directly from the corresponding definitions of zero-error quantum capacities. It should be mentioned at this point that these results can as well be obtained rather easily using the concept of non-commutative graphs (again accompanied by some convex geometry) that has been systematically explored in the recent work [27]. The fact that the quantum zero-error capacity is generically zero shows that the channels for which it is possible to satisfy the Knill-Laflamme condition [48] on a subspace of dimension greater or equal than 2 are exceptional.
We also list two properties that lead to a single-letter capacity formula of an AVQC in section 4.7 and compute the (deterministic) entanglement transmission capacity of an erasure AVQC (section 4.8).
A small add-on is given in section 4.9, where we show that the common-randomness-assisted entanglement generation capacity of an AVQC equals its random entanglement transmission capacity.

2.2.3 Previous work

The model of an arbitrarily varying channel has been introduced by Blackwell, Breiman and Thomasian [18] in 1960. They derived a formula for the capacity of an AVC with random codes and asymptotically vanishing average error probability. They also wrote down an explicit example of an AVC whose deterministic capacity is zero, while having nonzero capacity when using random codes.
Later landmarks in the development of coding theorems for AVCs have been the papers by Kiefer and Wolfowitz [45], who found a necessary and sufficient condition for an AVC to have nonzero capacity with deterministic codes and asymptotically vanishing maximal error probability.
The maximal error probability criterion was further investigated in [7] by Ahlswede and Wolfowitz, who completely determined the capacity of AVCs with binary output alphabet under that criterion. A solution for arbitrarily large alphabets does not seem to exist until now. It should be mentioned that such a solution would include the solution to Shannon's zero error capacity problem [63], as pointed out in [1].
In our approach we use the powerful elimination technique developed by Ahlswede in 1978 [2] that, together with the random coding results of [18] enabled him to prove the following dichotomy result for AVCs: It stated that the capacity of an AVC (under the average error probability criterion) is either zero or equals its random coding capacity. Together with Ahlswede's robustificaion technique [3, 4], the elimination technique led to a rather streamlined approach that, in this combination, has first been successfully used in [4].
After the discoveries of [2], an important open question was, when exactly the deterministic capacity with vanishing average error is equal to zero. In 1985, a first step towards a solution was made by Ericson [29], who came up with a sufficient condition that was proven to be necessary by Csiszar and Narayan [21] in 1989.
The model of an arbitrarily varying channel with classical input and quantum output has first been considered in 2007 by Ahlswede and Blinovsky [6]. They considered the transmission of messages under the average error criterion and gave a complete solution of the problem, i.e. a single-letter capacity formula, including a necessary and sufficient condition for the case of zero capacity.

3 The compound quantum channel

3.1 Definitions and main result

Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$. The memoryless compound channel associated with $\mathfrak{I}$ is given by the family $\{\mathcal{N}^{\otimes l} : \mathcal{S}(\mathcal{H}^{\otimes l}) \to \mathcal{S}(\mathcal{K}^{\otimes l})\}_{l\in\mathbb{N},\mathcal{N}\in\mathfrak{I}}$. In the rest of this work we will simply write $\mathfrak{I}$ or, if ambiguities have to be avoided, *the compound channel* $\mathfrak{I}$ for that family.
Each compound channel can be used in three different scenarios:

1. the *informed decoder*
2. the *informed encoder*
3. the case of *uninformed users.*

In the following three sections we will give definitions of codes and capacity for these cases.

3.1.1 The informed decoder

Definition 2. *An* (l, k_l)-code *for* $\mathfrak{I}$ *with* informed decoder *is a pair* $(\mathcal{P}^l, \{\mathcal{R}^l_{\mathcal{N}} : \mathcal{N} \in \mathfrak{I}\})$ *where:*

1. $\mathcal{P}^l : \mathcal{B}(\mathcal{F}_l) \to \mathcal{B}(\mathcal{H})^{\otimes l}$ *is a CPTP map for some Hilbert space* $\mathcal{F}_l$ *with* $k_l = \dim \mathcal{F}_l$.
2. $\mathcal{R}^l_{\mathcal{N}} : \mathcal{B}(\mathcal{K})^{\otimes l} \to \mathcal{B}(\mathcal{F}'_l)$ *is a CPTP map for each* $\mathcal{N} \in \mathfrak{I}$ *where the Hilbert space* $\mathcal{F}'_l$ *satisfies* $\mathcal{F}_l \subset \mathcal{F}'_l$.

Remark 3. *In what follows the operations* $\mathcal{R}^l_{\mathcal{N}}$ *are referred to as recovery (or decoding) operations. Since the decoder knows which channel is actually used during transmission, they are allowed to depend on the channel.*
Note at this point that we deviate from the standard assumption that $\mathcal{F}_l = \mathcal{F}'_l$. *We allow* $\mathcal{F}_l \subsetneq \mathcal{F}'_l$ *for convenience only since it allows more flexibility in code construction. It is readily seen from the definition of achievable rates and capacity below that the assumption* $\mathcal{F}_l \subsetneq \mathcal{F}'_l$ *cannot lead to a higher capacity of* $\mathfrak{I}$ *in any of the three cases that we are dealing with.*

Definition 4. *A non-negative number* R *is called an achievable rate for transmission of entanglement over* $\mathfrak{I}$ *with informed decoder if there is a sequence of* (l, k_l)*-codes such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l_{\mathcal{N}} \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l) = 1$

holds.

Definition 5. *The* entanglement transmission capacity $Q_{ID}(\mathfrak{I})$ *of the compound channel* $\mathfrak{I}$ *with informed decoder is given by*

$$Q_{ID}(\mathfrak{I}) := \sup\left\{R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable rate for transmission of} \\ \text{entanglement over } \mathfrak{I} \text{ with informed decoder} \end{array}\right\}. \tag{23}$$

Definition 6. *A non-negative number* R *is said to be an achievable strong subspace transmission rate for* $\mathfrak{I}$ *with informed decoder if there is a sequence of* $(l, k_l)-$ *codes for* $\mathfrak{I}$ *with informed decoder such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$ *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} \min_{\psi\in S(\mathcal{F}_l)} F(|\psi\rangle\langle\psi|, \mathcal{R}^l_{\mathcal{N}} \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l(|\psi\rangle\langle\psi|)) = 1$.

Definition 7. *The strong subspace transmission capacity $Q_{s,ID}(\mathfrak{I})$ of $\mathfrak{I}$ with informed decoder is defined by*

$$Q_{s,ID}(\mathfrak{I}) := \sup\left\{ R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable strong subspace trans-} \\ \text{mission rate for } \mathfrak{I} \text{ with informed decoder} \end{array} \right\}. \tag{24}$$

Definition 8. *An entanglement-generating (l, k_l)-code for the compound channel $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ with informed decoder consists of a pair $(\{\mathcal{R}^l_{\mathcal{N}}\}_{\mathcal{N}\in\mathfrak{I}}, \varphi_l)$ where $\mathcal{R}^l \in \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l)$ with $k_l = \dim \mathcal{F}_l$ and φ_l is a pure state on $\mathcal{F}_l \otimes \mathcal{H}^{\otimes l}$.*

Definition 9. *$R \in \mathbb{R}_+$ is called an achievable entanglement generation rate for $\mathfrak{I}$ with informed decoder if there is a sequence of (l, k_l) entanglement-generating codes with*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} F(|\psi_l\rangle\langle\psi_l|, (id_{\mathcal{F}_l} \otimes \mathcal{R}^l_{\mathcal{N}} \circ \mathcal{N}^{\otimes l})(|\varphi_l\rangle\langle\varphi_l|)) = 1$ *where ψ_l denotes the standard maximally entangled state on $\mathcal{F}_l \otimes \mathcal{F}_l$ and $F(\cdot,\cdot)$ is the fidelity.*

Definition 10. *The entanglement-generating capacity $E(\mathfrak{I})$ of $\mathfrak{I}$ with inforemd decoder is then defined as*

$$E_{ID}(\mathfrak{I}) := \sup\left\{ R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable entanglement generation} \\ \text{rate for } \mathfrak{I} \text{ with informed decoder} \end{array} \right\}. \tag{25}$$

3.1.2 The informed encoder

Definition 11. *An (l, k_l)-*code *for $\mathfrak{I}$ with* informed encoder *is a pair $(\{\mathcal{P}^l_{\mathcal{N}} : \mathcal{N} \in \mathfrak{I}\}, \mathcal{R}^l)$ where:*

1. *$\mathcal{P}^l_{\mathcal{N}} : \mathcal{B}(\mathcal{F}_l) \to \mathcal{B}(\mathcal{H})^{\otimes l}$ is a CPTP map for each $\mathcal{N} \in \mathfrak{I}$ for some Hilbert space $\mathcal{F}_l$ with $k_l = \dim \mathcal{F}_l$. The maps $\mathcal{P}^l_{\mathcal{N}}$ are the encoding operations which we allow to depend on $\mathcal{N}$ since the encoder knows which channel is in use.*
2. *$\mathcal{R}^l : \mathcal{B}(\mathcal{K})^{\otimes l} \to \mathcal{B}(\mathcal{F}'_l)$ is a CPTP map where the Hilbert space $\mathcal{F}'_l$ satisfies $\mathcal{F}_l \subset \mathcal{F}'_l$.*

Definition 12. *A non-negative number R is called an achievable rate for entanglement transmission over $\mathfrak{I}$ with informed encoder if there is a sequence of (l, k_l)-codes such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l_{\mathcal{N}}) = 1$

holds.

Definition 13. *The* entanglement transmission capacity *$Q_{IE}(\mathfrak{I})$ of the compound channel $\mathfrak{I}$ with informed encoder is given by*

$$Q_{IE}(\mathfrak{I}) := \sup\left\{ R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is achievable for entanglement} \\ \text{transmission over } \mathfrak{I} \text{ with informed encoder} \end{array} \right\}. \tag{26}$$

Definition 14. *A non-negative number R is said to be an achievable strong subspace transmission rate for $\mathfrak{I}$ with informed encoder if there is a sequence of $(l, k_l)-$ codes for $\mathfrak{I}$ with informed encoder such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$ *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} \min_{\psi\in S(\mathcal{F}_l)} F(|\psi\rangle\langle\psi|, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l_{\mathcal{N}}(|\psi\rangle\langle\psi|)) = 1.$

Definition 15. *The strong subspace transmission capacity $Q_{s,ID}(\mathfrak{I})$ of $\mathfrak{I}$ with informed encoder is defined by*

$$Q_{s,IE}(\mathfrak{I}) := \sup\{R : R \text{ is an achievable strong subspace transmission rate for } \mathfrak{I} \text{ with informed encoder}\}. \tag{27}$$

Definition 16. *An entanglement-generating (l, k_l)-code for the compound channel $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ with informed encoder consists of a pair $(\mathcal{R}^l, \{\varphi_{l,\mathcal{N}}\}_{\mathcal{N}\in\mathfrak{I}})$ where $\mathcal{R}^l \in \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l)$ with $k_l = \dim \mathcal{F}_l$ and $\{\varphi_{l,\mathcal{N}}\}_{\mathcal{N}\in\mathfrak{I}}$ is a set of pure states on $\mathcal{F}_l \otimes \mathcal{H}^{\otimes l}$.*

Definition 17. *$R \in \mathbb{R}_+$ is called an achievable entanglement generation rate for $\mathfrak{I}$ with informed encoder if there is a sequence of (l, k_l) entanglement-generating codes with*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} F(|\psi_l\rangle\langle\psi_l|, (id_{\mathcal{F}_l} \otimes \mathcal{R}^l \circ \mathcal{N}^{\otimes l})(|\varphi_l\rangle\langle\varphi_l|)) = 1$ *where ψ_l denotes the standard maximally entangled state on $\mathcal{F}_l \otimes \mathcal{F}_l$ and $F(\cdot,\cdot)$ is the fidelity.*

Definition 18. *The entanglement-generating capacity $E_{IE}(\mathfrak{I})$ of $\mathfrak{I}$ with informed encoder is then defined as*

$$E(\mathfrak{I}) := \sup\left\{R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable entanglement} \\ \text{generation rate for } \mathfrak{I} \text{ with informed encoder} \end{array}\right\}. \tag{28}$$

3.1.3 The case of uninformed users

Codes and capacity for the compound channel $\mathfrak{I}$ with *uninformed users* are defined in a similar fashion. The only change is that we neither allow the encoding nor the recovery operations to depend on $\mathcal{N}$:

Definition 19. *An $(l, k_l)-$ code for $\mathfrak{I}$ with uninformed users is a pair $(\mathcal{P}^l, \mathcal{R}^l)$ of CPTP maps $\mathcal{P}^l \in \mathcal{C}(\mathcal{F}_l, \mathcal{H}^{\otimes l})$ where $\mathcal{F}_l$ is a Hilbert space with $k_l = \dim \mathcal{F}_l$ and $\mathcal{R}^l \in \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l')$ with $\mathcal{F}_l \subset \mathcal{F}_l'$.*

Definition 20. *A non-negative number R is called an achievable rate for transmission of entanglement over $\mathfrak{I}$ with uninformed users if there is a sequence of (l, k_l)-codes such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l) = 1$.

Definition 21. *The* capacity *$Q(\mathfrak{I})$ of the compound channel $\mathfrak{I}$ with uninformed users is given by*

$$Q(\mathfrak{I}) := \sup\left\{R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is achievable for transmission of} \\ \text{entanglement over } \mathfrak{I} \text{ with uninformed users} \end{array}\right\}. \tag{29}$$

Definition 22. *A non-negative number R is said to be an achievable strong subspace transmission rate for $\mathfrak{I}$ with uninformed users if there is a sequence of $(l, k_l)-$ codes for $\mathfrak{I}$ with uninformed users such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$ *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} \min_{\psi\in S(\mathcal{F}_l)} F(|\psi\rangle\langle\psi|, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l(|\psi\rangle\langle\psi|)) = 1$.

The strong subspace transmission capacity $Q_s(\mathfrak{I})$ of $\mathfrak{I}$ with uninformed users is defined by

$$Q_{s,IE}(\mathfrak{I}) := \sup\left\{R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable strong subspace trans-} \\ \text{mission rate for } \mathfrak{I} \text{ with uninformed users} \end{array}\right\}. \tag{30}$$

Definition 23. *An entanglement-generating (l, k_l)-code for the compound channel $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ with uninformed users consists of a pair $(\mathcal{R}^l, \varphi_l)$ where $\mathcal{R}^l \in \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l)$ with $k_l = \dim \mathcal{F}_l$ and φ_l is a pure state on $\mathcal{F}_l \otimes \mathcal{H}^{\otimes l}$.*

Definition 24. *$R \in \mathbb{R}_+$ is called an achievable entanglement generation rate for $\mathfrak{I}$ with uninformed users if there is a sequence of (l, k_l) entanglement-generating codes with*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, *and*
2. $\lim_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} F(|\psi_l\rangle\langle\psi_l|, (id_{\mathcal{F}_l} \otimes \mathcal{R}^l \circ \mathcal{N}^{\otimes l})(|\varphi_l\rangle\langle\varphi_l|)) = 1$ *where ψ_l denotes the standard maximally entangled state on $\mathcal{F}_l \otimes \mathcal{F}_l$ and $F(\cdot, \cdot)$ is the fidelity.*

Definition 25. *The entanglement-generating capacity $E(\mathfrak{I})$ of $\mathfrak{I}$ with uninformed users is then defined as*

$$E(\mathfrak{I}) := \sup \left\{ R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable entanglement generation} \\ \text{rate for } \mathfrak{I} \text{ with uninformed users} \end{array} \right\}. \tag{31}$$

Remark 26. *A first simple consequence of these definitions is given by the following relations among the capacities of $\mathfrak{I}$:*

$$\begin{aligned} Q(\mathfrak{I}) &\leq \min\{Q_{ID}(\mathfrak{I}), Q_{IE}(\mathfrak{I})\}, && (32)\\ Q_s(\mathfrak{I}) &\leq \min\{Q_{s,ID}(\mathfrak{I}), Q_{s,IE}(\mathfrak{I})\}, && (33)\\ E(\mathfrak{I}) &\leq \min\{E_{ID}(\mathfrak{I}), E_{IE}(\mathfrak{I})\}, && (34)\\ & && (35) \end{aligned}$$

3.1.4 Main result

With these definitions at our disposal, we are ready now to state the main result of the paper.

Theorem 27. *Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be an arbitrary set of quantum channels where $\mathcal{H}$ and $\mathcal{K}$ are finite dimensional Hilbert spaces.*

1. *The entanglement transmission capacities of $\mathfrak{I}$ satisfy*

$$Q(\mathfrak{I}) = Q_{ID}(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho, \mathcal{N}^{\otimes l}) \tag{36}$$

and

$$Q_{IE}(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \inf_{\mathcal{N}\in\mathfrak{I}} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{37}$$

2. *For the entanglement-generating capacities of $\mathfrak{I}$ we get*

$$E(\mathfrak{I}) = E_{ID}(\mathfrak{I}) = Q(\mathfrak{I}) \tag{38}$$

and

$$E_{IE}(\mathfrak{I}) = Q_{IE}(\mathfrak{I}). \tag{39}$$

3. *Strong subspace transmission is equivalent to entanglement transmission:*

$$Q_s(\mathfrak{I}) = Q_{s,ID}(\mathfrak{I}) = Q(\mathfrak{I}) \tag{40}$$

and

$$Q_{s,IE}(\mathfrak{I}) = Q_{IE}(\mathfrak{I}). \tag{41}$$

In the main part of the rest of chapter 3, we give a step-by-step proof of Theorem 27.

3.2 One-shot results

In this section we will establish the basic building blocks for the achievability parts of the coding theorems for compound channels with and without channel knowledge. The results are formulated as one-shot statements in order to simplify the notation.

3.2.1 One-shot coding result for a single channel

Before we turn our attention to quantum compound channels we will shortly describe a part of recent developments in coding theory for single (i.e. perfectly known) channels as given in [47] and [35]. Both approaches are based on a decoupling idea which is closely related to approximate error correction. In order to state this decoupling lemma we need some notational preparation.
Let $\rho \in \mathcal{S}(\mathcal{H})$ be given and consider any purification $\psi \in \mathcal{H}_a \otimes \mathcal{H}$, $\mathcal{H}_a = \mathcal{H}$, of ρ. According to Stinespring's representation theorem any $\mathcal{N} \in \mathcal{C}^{\downarrow}(\mathcal{H}, \mathcal{K})$ is given by

$$\mathcal{N}(\,\cdot\,) = \mathrm{tr}_{\mathcal{H}_e}((\mathbf{1}_{\mathcal{H}} \otimes p_e)v(\,\cdot\,)v^*), \tag{42}$$

where $\mathcal{H}_e$ is a suitable finite-dimensional Hilbert space, p_e is a projection onto a subspace of $\mathcal{H}_e$, and $v : \mathcal{H} \to \mathcal{K} \otimes \mathcal{H}_e$ is an isometry.
Let us define a pure state on $\mathcal{H}_a \otimes \mathcal{K} \otimes \mathcal{H}_e$ by the formula

$$\psi' := \frac{1}{\sqrt{\mathrm{tr}(\mathcal{N}(\pi_{\mathcal{F}}))}}(\mathbf{1}_{\mathcal{H}_a \otimes \mathcal{K}} \otimes p_e)(\mathbf{1}_{\mathcal{H}_a} \otimes v)\psi. \tag{43}$$

We set

$$\rho' := \mathrm{tr}_{\mathcal{H}_a \otimes \mathcal{H}_e}(|\psi'\rangle\langle\psi'|), \quad \rho'_{ae} := \mathrm{tr}_{\mathcal{K}}(|\psi'\rangle\langle\psi'|), \tag{44}$$

and

$$\rho_a := \mathrm{tr}_{\mathcal{K} \otimes \mathcal{H}_e}(|\psi'\rangle\langle\psi'|), \quad \rho'_e := \mathrm{tr}_{\mathcal{H}_a \otimes \mathcal{K}}(|\psi'\rangle\langle\psi'|). \tag{45}$$

The announced decoupling lemma can now be stated as follows.

Lemma 28 (Cf. [47],[35]). *For $\rho \in \mathcal{S}(\mathcal{H})$ and $\mathcal{N} \in \mathcal{C}^{\downarrow}(\mathcal{H}, \mathcal{K})$ there exists a recovery operation $\mathcal{R} \in \mathcal{C}(\mathcal{K}, \mathcal{H})$ with*

$$F_e(\rho, \mathcal{R} \circ \mathcal{N}) \geq w - ||w\rho'_{ae} - w\rho_a \otimes \rho'_e||_1, \tag{46}$$

where $w = tr(\mathcal{N}(\rho))$.

The striking implication of Lemma 28 is that if the so called quantum error $||\rho'_{ae} - \rho_a \otimes \rho'_e||_1$ for $\rho \in \mathcal{S}(\mathcal{H})$ and $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ is small then almost perfect error correction is possible via $\mathcal{R}$.
Lemma 28 was Klesse's [47] starting point for his highly interesting proof of the following theorem which is a one-shot version of the achievability part of the coding theorem. In the statement of the result we will use the following notation.

$$F_{c,e}(\rho, \mathcal{N}) := \max_{\mathcal{R} \in \mathcal{C}(\mathcal{K}, \mathcal{H})} F_e(\rho, \mathcal{R} \circ \mathcal{N}), \tag{47}$$

where $\rho \in \mathcal{S}(\mathcal{H})$ and $\mathcal{N} \in \mathcal{C}^{\downarrow}(\mathcal{H}, \mathcal{K})$.

Theorem 29 (Klesse [47]). *Let the Hilbert space $\mathcal{H}$ be given and consider subspaces $\mathcal{E} \subset \mathcal{G} \subset \mathcal{H}$ with $\dim \mathcal{E} = k$. Then for any $\mathcal{N} \in \mathcal{C}^{\downarrow}(\mathcal{H}, \mathcal{K})$ allowing a representation with n Kraus operators we have*

$$\int_{\mathfrak{U}(\mathcal{G})} F_{c,e}(u\pi_{\mathcal{E}}u^*, \mathcal{N})du \geq tr(\mathcal{N}(\pi_{\mathcal{G}})) - \sqrt{k \cdot n}||\mathcal{N}(\pi_{\mathcal{G}})||_2, \tag{48}$$

where $\mathfrak{U}(\mathcal{G})$ denotes the group of unitaries acting on $\mathcal{G}$ and du indicates that the integration is with respect to the Haar measure on $\mathfrak{U}(\mathcal{G})$.

We will indicate briefly how Klesse [47] derived the direct part of the coding theorem for memoryless quantum channels from Theorem 29. Let us choose for each $l \in \mathbb{N}$ subspaces $\mathcal{E}_l \subset \mathcal{G}^{\otimes l} \subset \mathcal{H}^{\otimes l}$ with

$$\dim \mathcal{E}_l =: k_l = 2^{l(I_c(\pi_{\mathcal{G}},\mathcal{N})-3\epsilon)}. \tag{49}$$

To given $\mathcal{N} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ and $\pi_{\mathcal{G}}$ Klesse constructed a reduced version $\mathcal{N}_l$ of $\mathcal{N}^{\otimes l}$ in such a way that $\mathcal{N}_l$ has a Kraus representation with $n_l \leq 2^{l(S_e(\pi_{\mathcal{G}},\mathcal{N})+\epsilon)}$ Kraus operators. Let $q_l \in \mathcal{B}(\mathcal{K}^{\otimes l})$ be the entropy-typical projection of the state $(\mathcal{N}(\pi_{\mathcal{G}}))^{\otimes l}$ and set $\mathcal{N}_l'(\cdot) := q_l \mathcal{N}_l(\cdot) q_l$. Then we have the following properties (some of which are stated once more for completeness)

1. $k_l = 2^{l(I_c(\pi_{\mathcal{G}},\mathcal{N})-3\epsilon)}$,
2. $\mathrm{tr}(\mathcal{N}_l'(\pi_{\mathcal{G}}^{\otimes l})) \geq 1 - o(l^0)$[1],
3. $n_l \leq 2^{l(S_e(\pi_{\mathcal{G}},\mathcal{N})+\epsilon)}$, and
4. $||\mathcal{N}_l'(\pi_{\mathcal{G}}^{\otimes l})||_2^2 \leq 2^{-l(S(\pi_{\mathcal{G}})-\epsilon)}$

An application of Theorem 29 to $\mathcal{N}_l'$ shows heuristically the existence of a unitary $u \in \mathfrak{U}(\mathcal{G}^{\otimes l})$ and a recovery operation $\mathcal{R}_l \in \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{H}^{\otimes l})$ with

$$F_e(u\pi_{\mathcal{E}_l}u^*, \mathcal{R}_l \circ \mathcal{N}_l') \geq 1 - o(l^0) - 2^{-\frac{l}{2}\epsilon}. \tag{50}$$

This in turn can be converted into

$$F_e(u\pi_{\mathcal{E}_l}u^*, \mathcal{R}_l \circ \mathcal{N}^{\otimes l}) \geq 1 - o(l^0), \tag{51}$$

which is the achievability of $I_c(\pi_{\mathcal{G}},\mathcal{N})$. The passage from $\pi_{\mathcal{G}}$ to arbitrary states ρ is then accomplished via the Bennett, Shor, Smolin, and Thapliyal Lemma from [12] and the rest is by regularization.

3.2.2 One-shot coding result for uninformed users

Our goal in this subsection is to establish a variant of Theorem 29 that works for finite sets of channels. Since the entanglement fidelity depends affinely on the channel it is easily seen that for each set $\mathfrak{I} = \{\mathcal{N}_1, \ldots, \mathcal{N}_N\}$ any good coding scheme with uninformed users is also good for the channel

$$\mathcal{N} := \frac{1}{N}\sum_{i=1}^{N} \mathcal{N}_i \tag{52}$$

and vice versa. Since it is easier to deal with a single channel and we do not loose anything if passing to averages we will formulate our next theorem for arithmetic averages of completely positive trace decreasing maps instead of the set $\{\mathcal{N}_1, \ldots, \mathcal{N}_N\}$.

Theorem 30 (One-Shot Result: Uninformed Users and Averaged Channel)**.** *Let the Hilbert space $\mathcal{H}$ be given and consider subspaces $\mathcal{E} \subset \mathcal{G} \subset \mathcal{H}$ with $\dim \mathcal{E} = k$. For any choice of $\mathcal{N}_1, \ldots \mathcal{N}_N \in \mathcal{C}^{\downarrow}(\mathcal{H},\mathcal{K})$ each allowing a representation with n_j Kraus operators, $j = 1, \ldots, N$, we set*

$$\mathcal{N} := \frac{1}{N}\sum_{j=1}^{N} \mathcal{N}_j, \tag{53}$$

[1] Here, $o(l^0)$ denotes simply a non-specified sequence tending to 0 as $l \to \infty$, i.e. we (ab)use the Bachmann-Landau little-o notation.

and and for any $u \in \mathfrak{U}(\mathcal{G})$

$$\mathcal{N}_u := \frac{1}{N}\sum_{j=1}^{N} \mathcal{N}_j \circ \mathcal{U}. \tag{54}$$

Then

$$\int_{\mathfrak{U}(\mathcal{G})} F_{c,e}(\pi_{\mathcal{E}}, \mathcal{N}_u) du \quad \geq \quad tr(\mathcal{N}(\pi_{\mathcal{G}})) - 2\sum_{j=1}^{N} \sqrt{kn_j} ||\mathcal{N}_j(\pi_{\mathcal{G}})||_2, \tag{55}$$

where the integration is with respect to the normalized Haar measure on $\mathfrak{U}(\mathcal{G})$.

Remark 31. *It is worth noting that the average in this theorem is no more over maximally mixed states like in Theorem 29, but rather over encoding operations.*

Proof.

Lemma 32. *Let* L *and* D *be* $N \times N$ *matrices with non-negative entries which satisfy*

$$L_{jl} \leq L_{jj}, \quad L_{jl} \leq L_{ll}, \tag{56}$$

and

$$D_{jl} \leq \max\{D_{jj}, D_{ll}\} \tag{57}$$

for all $j, l \in \{1, \ldots, N\}$. *Then*

$$\sum_{j,l=1}^{N} \frac{1}{N}\sqrt{L_{jl}D_{jl}} \leq 2\sum_{j=1}^{N} \sqrt{L_{jj}D_{jj}}. \tag{58}$$

Proof. Note that (57) implies

$$D_{jl} \leq D_{jj} + D_{ll}. \tag{59}$$

Therewith we obtain

$$\begin{aligned}
\sum_{j,l=1}^{N} \frac{1}{N}\sqrt{L_{jl}D_{jl}} &\leq \sum_{j,l=1}^{N} \frac{1}{N}\sqrt{L_{jl}(D_{jj} + D_{ll})} && (60)\\
&\leq \sum_{j,l=1}^{N} \frac{1}{N}\sqrt{L_{jj}D_{jj} + L_{ll}D_{ll}} && (61)\\
&\leq \sum_{j,l=1}^{N} \frac{1}{N}\left(\sqrt{L_{jj}D_{jj}} + \sqrt{L_{ll}D_{ll}}\right) && (62)\\
&= 2\sum_{j=1}^{N} \sqrt{L_{jj}D_{jj}}, && (63)
\end{aligned}$$

where in (60) we have used (59), in (61) we employed (56), and (62) holds because $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all non-negative real numbers a, b. □

Proof. We can assume without loss of generality that the numbering of the channels is chosen in such a way that $n_1 \leq n_2 \leq \ldots \leq n_N$ holds for the numbers of Kraus operators of the maps $\mathcal{N}_1, \ldots, \mathcal{N}_N$. From Lemma 28 we know that there is a recovery operation $\mathcal{R}$ such that

$$F_e(\pi_{\mathcal{F}}, \mathcal{R} \circ \mathcal{N}) \geq w - ||w\rho'_{ae} - w\rho_a \otimes \rho'_e||_1, \tag{64}$$

where we have used the notation introduced in the paragraph preceeding Lemma 28.
For each $j \in \{1,\dots,N\}$ let $\{a_{j,i}\}_{i=1}^{n_j}$ be the set of Kraus operators of $\mathcal{N}_j$. Let $\{f_1,\dots,f_N\}$ and $\{e_1,\dots,e_{n_N}\}$ be arbitrary orthonormal bases of $\mathbb{C}^N$ and $\mathbb{C}^{n_N}$. Let the projection p_e and unitary v in (42) be chosen in such a way that for each $\phi \in \mathcal{H}$ the relation

$$(\mathbf{1}_{\mathcal{K}} \otimes p_e)v(\phi\otimes) = \sum_{j=1}^{N}\sum_{i=1}^{n_j} \frac{1}{\sqrt{N}}(a_{j,i}\phi) \otimes e_i \otimes f_j, \tag{65}$$

holds. For a purification $\psi \in \mathcal{H}_a \otimes \mathcal{H}$ of the state $\pi_{\mathcal{F}}$ we consider a Schmidt representation

$$\psi = \frac{1}{\sqrt{k}} \sum_{m=1}^{k} h_m \otimes g_m, \tag{66}$$

with suitable orthonormal systems $\{h_1,\dots,h_k\}$ and $\{g_1,\dots,g_k\}$. A calculation identical to that performed by Klesse [47] shows that the states on the right hand side of (64) can be expressed with the help of representation (65) as

$$w\rho'_{ae} = \frac{1}{k}\sum_{j,l=1}^{N}\sum_{i,r=1}^{n_j,n_l}\sum_{s,t=1}^{k} \frac{\mathrm{tr}(a_{j,i}|g_s\rangle\langle g_t|a^*_{l,r})}{N}|x_{s,i,j}\rangle\langle x_{t,r,l}|, \tag{67}$$

with $x_{s,i,j} := h_s \otimes e_i \otimes f_j$, and

$$w\rho_a \otimes \rho'_e = \sum_{j,l=1}^{N}\sum_{i,r=1}^{n_j,n_l} \frac{\mathrm{tr}(a_{j,i}\pi_{\mathcal{F}}a^*_{l,r})}{kN}\rho_a \otimes |y_{i,j}\rangle\langle y_{r,l}|, \tag{68}$$

where $y_{i,j} := e_i \otimes f_j$.
If we perform the unitary conjugation induced by the unitary map $x_{s,i,j} = h_s \otimes e_i \otimes f_j \mapsto x'_{s,i,j} = g_s \otimes e_i \otimes f_j$ followed by the complex conjugation of the matrix elements with respect to the matrix units $\{|x'_{s,i,j}\rangle\langle x'_{t,k,l}|\}_{s,i,j,t,k,l}$ we obtain an anti-linear isometry I with respect to the metrics induced by the trace distances on the operator spaces under consideration. A calculation identical to that in [47] shows that under this isometry the sub-normalized states in (67) and (68) transform to

$$I(w\rho'_{ae}) = \frac{1}{kN}\sum_{j,l=1}^{N}\sum_{i,r=1}^{n_j,n_l} pa^*_{j,i}a_{l,r}p \otimes |y_{i,j}\rangle\langle y_{r,l}|, \tag{69}$$

and

$$I(w\rho_a \otimes \rho'_e) = \frac{1}{k}\sum_{j,l=1}^{N}\sum_{i,r=1}^{n_j,n_l} \frac{\mathrm{tr}(pa^*_{j,i}a_{l,r}p)}{kN}p \otimes |y_{i,j}\rangle\langle y_{r,l}|, \tag{70}$$

with $p = k\pi_{\mathcal{F}}$ and $y_{i,j} = e_i \otimes f_j$ for $j = 1,\dots,N$ and $i = 1,\dots,n_j$. In summary, using the isometry I, (69), and (70) the inequality (64) can be formulated as

$$F_e(\pi_{\mathcal{F}}, \mathcal{R}\circ\mathcal{N}) \geq w - ||D(p)||_1, \tag{71}$$

with $w = \mathrm{tr}(\mathcal{N}(\pi_{\mathcal{F}}))$ and

$$D(p) := \sum_{j,l=1}^{N}\frac{1}{N}\sum_{i,r=1}^{n_j,n_l} D_{(ij)(rl)}(p) \otimes |e_i\rangle\langle e_r| \otimes |f_j\rangle\langle f_l| \tag{72}$$

where

$$D_{(ij)(rl)}(p) := \frac{1}{k}\left(pa_{j,i}a_{l,r}^{*}p - \frac{1}{k}\mathrm{tr}(pa_{j,i}^{*}a_{l,r}p)p\right). \tag{73}$$

Let us define

$$D_{j,l}(p) := \sum_{i=1,k=1}^{n_j,n_l} D_{(ij)(kl)}(p)\otimes|e_i\rangle\langle e_k|\otimes|f_j\rangle\langle f_l|. \tag{74}$$

The triangle inequality for the trace norm yields

$$\begin{aligned} ||D(p)||_1 &\leq \sum_{j,l=1}^{N}\frac{1}{N}||D_{j,l}(p)||_1 && (75)\\ &\leq \sum_{j,l=1}^{N}\frac{1}{N}\sqrt{k\min\{n_j,n_l\}}||D_{j,l}(p)||_2, && (76)\\ &= \sum_{j,l=1}^{N}\frac{1}{N}\sqrt{k\min\{n_j,n_l\}||D_{j,l}(p)||_2^2}, && (77)\end{aligned}$$

where the second line is justified by the standard relation between the trace and Hilbert-Schmidt norm, $||a||_1 \leq \sqrt{d}||a||_2$, d being the number of non-zero singular values of a.
In the next step we will compute $||D_{j,l}(p)||_2^2$. A glance at (74) shows that

$$(D_{j,l}(p))^{*} = \sum_{i=1,k=1}^{n_j,n_l}(D_{(ij)(kl)}(p))^{*}\otimes|e_k\rangle\langle e_i|\otimes|f_l\rangle\langle f_j|, \tag{78}$$

and consequently we obtain

$$\begin{aligned} ||D_{j,l}(p)||_2^2 &= \mathrm{tr}((D_{j,l}(p))^{*}D_{j,l}(p)) && (79)\\ &= \sum_{i=1,r=1}^{n_j,n_l}\mathrm{tr}((D_{(ij)(kl)}(p))^{*}D_{(ij)(kl)}(p)) && (80)\\ &= \frac{1}{k^2}\sum_{i=1,r=1}^{n_j,n_l}\{\mathrm{tr}(p(a_{j,i}^{*}a_{l,r})^{*}pa_{j,i}^{*}a_{l,r}) - \frac{1}{k}|\mathrm{tr}(pa_{j,i}^{*}a_{l,r})|^2\}. && (81)\end{aligned}$$

Let U be a random variable taking values in $\mathfrak{U}(\mathcal{G})$ according to the Haar measure of $\mathfrak{U}(\mathcal{G})$. Then we can infer from (75) that

$$\mathbb{E}(||D(UpU^{*})||_1) \leq \sum_{j,l=1}^{N}\frac{1}{N}\sqrt{L_{jl}\mathbb{E}(||D_{j,l}(UpU^{*})||_2^2)}, \tag{82}$$

where we have used the concavity of the function $\sqrt{\cdot}$ and Jensen's inequality and, additionally, we abbreviated $k\min\{n_j,n_l\}$ by L_{jl}. Now, starting with (79) and arguing as Klesse [47] we obtain that

$$\begin{aligned} \mathbb{E}(||D_{j,l}(UpU^{*})||_2^2) &\leq \mathrm{tr}(\mathcal{N}_j(\pi_{\mathcal{G}})\mathcal{N}_l(\pi_{\mathcal{G}})) && (83)\\ &= \langle\mathcal{N}_j(\pi_{\mathcal{G}}),\mathcal{N}_l(\pi_{\mathcal{G}})\rangle_{HS}, && (84)\end{aligned}$$

where $\langle\,\cdot\,,\,\cdot\,\rangle_{HS}$ denotes the Hilbert-Schmidt inner product. Similarly

$$\mathbb{E}(\mathrm{tr}(\mathcal{N}(U\pi_{\mathcal{F}}U^{*}))) = \mathrm{tr}(\mathcal{N}(\pi_{\mathcal{G}})). \tag{85}$$

Now, using (64), (71), (79), (82), (83), and (85) we arrive at

$$\mathbb{E}(F_{c,e}(U\pi_{\mathcal{F}}U^*,\mathcal{N})) \quad \geq \quad \mathrm{tr}(\mathcal{N}(\pi_{\mathcal{G}})) - \sum_{j,l=1}^{N} \frac{1}{N}\sqrt{L_{jl}D_{jl}}, \tag{86}$$

where for $j,l \in \{1,\ldots,N\}$ we introduced the abbreviations

$$L_{jl} = k\min\{n_j, n_l\}, \tag{87}$$

and

$$D_{jl} := \langle \mathcal{N}_j(\pi_{\mathcal{G}}), \mathcal{N}_l(\pi_{\mathcal{G}})\rangle_{HS}. \tag{88}$$

It is obvious that

$$L_{jl} \leq L_{jj} \quad \text{and} \quad L_{jl} \leq L_{ll} \tag{89}$$

hold. Moreover, the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product justifies the following chain of inequalities

$$\begin{aligned} D_{jl} &= \langle \mathcal{N}_j(\pi_{\mathcal{G}}), \mathcal{N}_l(\pi_{\mathcal{G}})\rangle_{HS} && (90)\\ &\leq ||\mathcal{N}_j(\pi_{\mathcal{G}})||_2||\mathcal{N}_l(\pi_{\mathcal{G}})||_2 && (91)\\ &\leq \max\{||\mathcal{N}_j(\pi_{\mathcal{G}})||_2^2, ||\mathcal{N}_l(\pi_{\mathcal{G}})||_2^2\} && (92)\\ &= \max\{D_{jj}, D_{ll}\}. && (93)\end{aligned}$$

Therefore, an application of Lemma 32 allows us to conclude from (86) that

$$\mathbb{E}(F_{c,e}(U\pi_{\mathcal{F}}U^*,\mathcal{N})) \quad \geq \quad \mathrm{tr}(\mathcal{N}(\pi_{\mathcal{G}})) - 2\sum_{j=1}^{N} \sqrt{kn_j}||\mathcal{N}_j(\pi_{\mathcal{G}})||_2, \tag{94}$$

which is what we aimed to prove. □

□

3.2.3 One-shot coding result for informed encoder

We will focus now on the scenario where the sender or encoder knows which channel is in use. Consequently, the encoding operation can depend on the individual channel. The idea behind the next theorem is that we perform an independent, randomized selection of unitary encoders for each channel in the finite set $\mathfrak{I} = \{\mathcal{N}_1,\ldots,\mathcal{N}_N\}$. This explains why the averaging in (96) is with respect to products of Haar measures instead of averaging over one single Haar measure as in Theorem 30.

Theorem 33 (One-Shot Result: Informed Encoder and Averaged Channel)**.** *Let the finite-dimensional Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ be given. Consider subspaces $\mathcal{E},\mathcal{G}_1,\ldots,\mathcal{G}_N \subset \mathcal{H}$ with $\dim\mathcal{E} = k$ such that for all $i \in \{1,\ldots,N\}$ the dimension relation $k \leq \dim\mathcal{G}_i$ holds. Let $\mathcal{N}_1,\ldots\mathcal{N}_N \in \mathcal{C}^{\downarrow}(\mathcal{H},\mathcal{K})$ each allowing a representation with n_j Kraus operators, $j = 1,\ldots,N$. Let $\{v_i\}_{i=1}^N \subset \mathfrak{U}(\mathcal{H})$ be any fixed set of unitary operators such that $v_i\mathcal{E} \subset \mathcal{G}_i$ holds for every $i \in \{1,\ldots,N\}$. For an arbitrary set $\{u_i\}_{i=1}^N \subset \mathfrak{U}(\mathcal{H})$, define*

$$\mathcal{N}_{u_1,\ldots,u_N} := \frac{1}{N}\sum_{i=1}^{N} \mathcal{N}_i \circ \mathcal{U}_i \circ \mathcal{V}_i. \tag{95}$$

Then

$$\int_{\mathfrak{U}(\mathcal{G}_1)\times\ldots\times\mathfrak{U}(\mathcal{G}_N)} F_{c,e}(\pi_{\mathcal{E}},\mathcal{N}_{u_1,\ldots,u_N})du_1\ldots du_N \quad \geq \quad \sum_{j=1}^{N}\Big[\frac{1}{N}tr(\mathcal{N}_j(\pi_{\mathcal{G}_j})) - 2\sqrt{kn_j}||\mathcal{N}_j(\pi_{\mathcal{G}_j})||_2\Big], \tag{96}$$

where the integration is with respect to the product of the normalized Haar measures on $\mathfrak{U}(\mathcal{G}_1),\ldots,\mathfrak{U}(\mathcal{G}_N)$.

Proof. Our first step in the proof is to show briefly that $F_{c,e}(\pi_{\mathcal{E}}, \mathcal{N}_{u_1,\ldots,u_N})$ depends measurably on $(u_1, \ldots, u_N) \in \mathfrak{U}(\mathcal{G}_1) \times \ldots \times \mathfrak{U}(\mathcal{G}_N)$. For each recovery operation $\mathcal{R} \in \mathcal{C}(\mathcal{K}, \mathcal{H})$ we define a function $f_{\mathcal{R}} : \mathfrak{U}(\mathcal{G}_1) \times \ldots \times \mathfrak{U}(\mathcal{G}_N) \to [0, 1]$ by

$$f_{\mathcal{R}}(u_1, \ldots, u_N) := F_e(\pi_{\mathcal{E}}, \mathcal{R} \circ \mathcal{N}_{u_1,\ldots,u_N}). \tag{97}$$

Clearly, $f_{\mathcal{R}}$ is continuous for each fixed $\mathcal{R} \in \mathcal{C}(\mathcal{K}, \mathcal{H})$. Thus, the function

$$F_{c,e}(\pi_{\mathcal{E}}, \mathcal{N}_{u_1,\ldots,u_N}) = \max_{\mathcal{R} \in \mathcal{C}(\mathcal{K},\mathcal{H})} f_{\mathcal{R}}(u_1, \ldots, u_N) \tag{98}$$

is lower semicontinuous, and consequently measurable.
We turn now to the proof of inequality (96). From Lemma 28 we know that there is a recovery operation $\mathcal{R}$ such that

$$F_e(\pi_{\mathcal{E}}, \mathcal{R} \circ \mathcal{N}_{u_1,\ldots,u_N}) \geq w - ||w\rho'_{ae} - w\rho_a \otimes \rho'_e||_1, \tag{99}$$

where we have used the notation introduced in the paragraph preceding Lemma 28, and

$$w = w(u_1, \ldots, u_N) = \mathrm{tr}(\mathcal{N}_{u_1,\ldots,u_N}(\pi_{\mathcal{E}})). \tag{100}$$

For each $j \in \{1, \ldots, N\}$ let $\{b_{j,i}\}_{i=1}^{n_j}$ be the set of Kraus operators of $\mathcal{N}_j$. Clearly, for every set $u_1, \ldots, u_N$ of unitary matrices, $\mathcal{N}_j \circ \mathcal{U}_j \circ \mathcal{V}_j$ has Kraus operators $\{a_{j,i}\}_{i=1}^{n_j}$ given by $a_{j,i} = b_{j,i}u_jv_j$. Utilizing the very same calculation that was used in the proof of Theorem 30, which in turn is almost identical to the corresponding calculation in [47], we can reformulate inequality (99) as

$$F_e(\pi_{\mathcal{E}}, \mathcal{R} \circ \mathcal{N}_{u_1,\ldots,u_N}) \geq w - ||D(u_1, \ldots, u_N)||_1, \tag{101}$$

with $w = \mathrm{tr}(\mathcal{N}_{u_1,\ldots,u_N}(\pi_{\mathcal{E}}))$ and

$$D(u_1, \ldots, u_N) := \sum_{j,l=1}^{N} \frac{1}{N} \sum_{i,r=1}^{n_j,n_l} D_{(ij)(rl)}(u_j, u_l) \otimes |e_i\rangle\langle e_r| \otimes |f_j\rangle\langle f_l| \tag{102}$$

where

$$D_{(ij)(rl)}(u_j, u_l) := \frac{1}{k}\left(pa_{j,i}a^*_{l,r}p - \frac{1}{k}\mathrm{tr}(pa^*_{j,i}a_{l,r}p)p\right), \tag{103}$$

and $p := k\pi_{\mathcal{E}}$ is the projection onto $\mathcal{E}$. Let us define

$$D_{j,l}(u_j, u_l) := \sum_{i=1,k=1}^{n_j,n_l} D_{(ij)(kl)}(u_j, u_l) \otimes |e_i\rangle\langle e_k| \otimes |f_j\rangle\langle f_l|. \tag{104}$$

The triangle inequality for the trace norm yields

$$\begin{aligned} ||D(u_1, \ldots, u_N)||_1 &\leq \sum_{j,l=1}^{N} \frac{1}{N} ||D_{j,l}(u_j, u_l)||_1 && (105)\\ &\leq \sum_{j,l=1}^{N} \frac{1}{N} \sqrt{k \min\{n_j, n_l\}}\, ||D_{j,l}(u_j, u_l)||_2, && (106)\\ &= \sum_{j,l=1}^{N} \frac{1}{N} \sqrt{k \min\{n_j, n_l\}\, ||D_{j,l}(u_j, u_l)||_2^2}, && (107) \end{aligned}$$

where the second line follows from $||a||_1 \leq \sqrt{d}||a||_2$, d being the number of non-zero singular values of a. In the next step we will compute $||D_{j,l}(u_j, u_l)||_2^2$. We set $p_l := v_l p v_l^*$ which defines new projections $\{p_l\}_{l=1}^N$ with $\mathrm{supp}(p_l) \subset \mathcal{G}_l$ for every $l \in \{1, \ldots, N\}$. A glance at (104) shows that

$$(D_{j,l}(u_j, u_l))^* = \sum_{i=1,k=1}^{n_j,n_l} (D_{(ij)(kl)}(u_j, u_l))^* \otimes |e_k\rangle\langle e_i| \otimes |f_l\rangle\langle f_j|, \tag{108}$$

and consequently we obtain

$$\begin{aligned}
||D_{j,l}(u_j, u_l)||_2^2 &= \mathrm{tr}((D_{j,l}(u_j, u_l))^* D_{j,l}(u_j, u_l)) && (109)\\
&= \sum_{i=1,r=1}^{n_j,n_l} \mathrm{tr}((D_{(ij)(kl)}(u_j, u_l))^* D_{(ij)(kl)}(u_j, u_l)) && (110)\\
&= \frac{1}{k^2} \sum_{i=1,r=1}^{n_j,n_l} \{\mathrm{tr}(p(a_{j,i}^* a_{l,r})^* p a_{j,i}^* a_{l,r}) - \frac{1}{k}|\mathrm{tr}(p a_{j,i}^* a_{l,r})|^2\} && (111)\\
&= \frac{1}{k^2} \sum_{i=1,r=1}^{n_j,n_l} \{\mathrm{tr}(p_l u_l^* b_{l,r}^* b_{j,i} u_j p_j u_j^* b_{j,i}^* b_{l,r} u_l) - \frac{1}{k}|\mathrm{tr}(p v_j^* u_j^* b_{j,i}^* b_{l,r} u_l v_l)|^2\}. && (112)
\end{aligned}$$

It is apparent from the last two lines in (109) that $||D_{j,l}(u_j, u_l)||_2^2$ depends measurably on $(u_1, \ldots, u_N) \in \mathfrak{U}(\mathcal{G}_1) \times \ldots \times \mathfrak{U}(\mathcal{G}_N)$. Let $U_1, \ldots, U_N$ be independent random variables taking values in $\mathfrak{U}(\mathcal{G}_i)$ according to the normalized Haar measure on $\mathfrak{U}(\mathcal{G}_i)$ ($i \in \{1, \ldots, N\}$). Then using Jensen's inequality and abbreviating $L_{jl} := k \min\{n_j, n_l\}$ we can infer from (105) that

$$\mathbb{E}(||D(U_1, \ldots, U_N)||_1) \leq \sum_{j,l=1}^{N} \frac{1}{N} \sqrt{L_{jl} \mathbb{E}(||D_{j,l}(U_j, U_l)||_2^2)}. \tag{113}$$

Note that the expectations on the RHS of (113) are only with respect to pairs of random variables $U_1, \ldots, U_N$.
Our next goal is to upper-bound $\mathbb{E}(||D_{j,l}(U_j, U_l)||_2^2)$.
Case $j \neq l$: Since the last term in (109) is non-negative and the random variables U_j and U_l are independent we obtain the following chain of inequalities:

$$\begin{aligned}
\mathbb{E}(||D_{j,l}(U_j, U_l)||_2^2) &= \frac{1}{k^2} \sum_{i,r=1}^{n_j,n_l} \Big[\mathbb{E}\mathrm{tr}(p_l U_l^* b_{l,r}^* b_{j,i} U_j p_j U_j^* b_{j,i}^* b_{l,r} U_l) - \frac{1}{k}\mathbb{E}|\mathrm{tr}(p v_j^* U_j^* b_{j,i}^* b_{l,r} U_l v_l)|^2\Big] && (114)\\
&\leq \frac{1}{k^2} \sum_{i=1,r=1}^{n_j,n_l} \mathbb{E}\mathrm{tr}(p_l U_l^* b_{l,r}^* b_{j,i} U_j p_j U_j^* b_{j,i}^* b_{l,r} U_l) && (115)\\
&= \frac{1}{k^2} \sum_{i=1,r=1}^{n_j,n_l} \mathbb{E}\mathrm{tr}(U_l p_l U_l^* b_{l,r}^* b_{j,i} U_j p_j U_j^* b_{j,i}^* b_{l,r}) && (116)\\
&= \frac{1}{k^2} \sum_{i=1,r=1}^{n_j,n_l} \mathrm{tr}(\mathbb{E}(U_l p_l U_l^*) b_{l,r}^* b_{j,i} \mathbb{E}(U_j p_j U_j^*) b_{j,i}^* b_{l,r}) && (117)\\
&= \frac{1}{k^2} \sum_{i=1,r=1}^{n_j,n_l} \mathrm{tr}(k \cdot \pi_{\mathcal{G}_l} b_{l,r}^* b_{j,i} k \cdot \pi_{\mathcal{G}_j} b_{j,i}^* b_{l,r}) && (118)\\
&= \langle \mathcal{N}_j(\pi_{\mathcal{G}_j}), \mathcal{N}_l(\pi_{\mathcal{G}_l}) \rangle_{HS}, && (119)
\end{aligned}$$

where $\langle \cdot\,,\,\cdot \rangle_{HS}$ denotes the Hilbert-Schmidt inner product, and we used the fact that

$$\mathbb{E}(U_l p_l U_l^*) = k \cdot \pi_{\mathcal{G}_l} \quad \text{and} \quad \mathbb{E}(U_j p_j U_j^*) = k \cdot \pi_{\mathcal{G}_j}. \tag{120}$$

Case $j = l$: In this case we obtain

$$\begin{aligned}
\mathbb{E}(||D_{j,j}(U_j, U_j)||_2^2) &= \frac{1}{k^2} \sum_{i,r=1}^{n_j,n_j} \Big[\mathbb{E}\mathrm{tr}(p_j U_j^* b_{j,r}^* b_{j,i} U_j p_j U_j^* b_{j,i}^* b_{j,r} U_j) - \frac{1}{k} \mathbb{E}|\mathrm{tr}(p v_j^* U_j^* b_{j,i}^* b_{j,r} U_j v_j)|^2 \Big] && (121)\\
&= \frac{1}{k^2} \sum_{i=1,r=1}^{n_j,n_j} \mathbb{E}\mathrm{tr}(U_j p_j U_j^* b_{j,r}^* b_{j,i} U_j p_j U_j^* b_{j,i}^* b_{j,r}) - \frac{1}{k} \mathbb{E}|\mathrm{tr}(U_j p_j U_j^* b_{j,i}^* b_{j,r})|^2 \Big]. && (122)
\end{aligned}$$

Thus, the problem reduces to the evaluation of

$$\mathbb{E}\{b_{UpU^*}(x,y)\}, \qquad (x, y \in \mathcal{B}(\mathcal{H})) \tag{123}$$

where p is an orthogonal projection with $\mathrm{tr}(p) = k$ and

$$b_{UpU^*}(x,y) := \mathrm{tr}(UpU^* x^* UpU^* y) - \frac{1}{k}\mathrm{tr}(UpU^* x^*)\mathrm{tr}(UpU^* y), \tag{124}$$

for a Haar distributed random variable U with values in $\mathfrak{U}(\mathcal{G})$ where $\mathrm{supp}(p) \subset \mathcal{G} \subset \mathcal{H}$.
Here we can refer to [47] where the corresponding calculation is carried out via the theory of group invariants and explicit evaluations of appropriate integrals with respect to row-distributions of random unitary matrices. The result is

$$\mathbb{E}\{b_{UpU^*}(x,y)\} = \frac{k^2-1}{d^2-1}\mathrm{tr}(p_{\mathcal{G}} x^* p_{\mathcal{G}} y) + \frac{1-k^2}{d(d^2-1)}\mathrm{tr}(p_{\mathcal{G}} x^*)\mathrm{tr}(p_{\mathcal{G}} y), \tag{125}$$

for all $x, y \in \mathcal{B}(\mathcal{H})$ where $p_{\mathcal{G}}$ denotes the projection onto $\mathcal{G}$ with $\mathrm{tr}(p_{\mathcal{G}}) = d$. In Appendix 6 we will give an elementary derivation of (125) for the sake of completeness.
Inserting (125) with $x = y = b_{j,i}^* b_{j,r}$ into (121) yields with $d_j := \mathrm{tr}(p_{\mathcal{G}_j})$

$$\begin{aligned}
\mathbb{E}(||D_{j,j}(U_j, U_j)||_2^2) &= \frac{1-\frac{1}{k^2}}{d_j^2-1}\Big[\sum_{i=1,r=1}^{n_j,n_j} \mathrm{tr}(p_{\mathcal{G}_j} b_{j,r}^* b_{j,i} p_{\mathcal{G}_j} b_{j,i}^* b_{j,r}) - \frac{1}{d_j}|\mathrm{tr}((p_{\mathcal{G}_j} b_{j,i}^* b_{j,r})|^2 \Big] && (126)\\
&\le \frac{1-\frac{1}{k^2}}{d_j^2-1} \sum_{i=1,r=1}^{n_j,n_j} \mathrm{tr}(p_{\mathcal{G}_j} b_{j,r}^* b_{j,i} p_{\mathcal{G}_j} b_{j,i}^* b_{j,r}) && (127)\\
&\le \frac{1}{d_j^2} \sum_{i=1,r=1}^{n_j,n_j} \mathrm{tr}(p_{\mathcal{G}_j} b_{j,r}^* b_{j,i} p_{\mathcal{G}_j} b_{j,i}^* b_{j,r}) && (128)\\
&= \frac{1}{d_j^2} \sum_{i=1,r=1}^{n_j,n_j} \mathrm{tr}(b_{j,r} p_{\mathcal{G}_j} b_{j,r}^* b_{j,i} p_{\mathcal{G}_j} b_{j,i}^*) && (129)\\
&= \langle \mathcal{N}_j(\pi_{\mathcal{G}_j}), \mathcal{N}_j(\pi_{\mathcal{G}_j}) \rangle_{HS}. && (130)
\end{aligned}$$

Summarizing, we obtain

$$\mathbb{E}(||D_{j,j}(U_j, U_j)||_2^2) \le \langle \mathcal{N}_j(\pi_{\mathcal{G}_j}), \mathcal{N}_j(\pi_{\mathcal{G}_j}) \rangle_{HS} = ||\mathcal{N}_j(\pi_{\mathcal{G}_j})||_2^2. \tag{131}$$

Similarly

$$\mathbb{E}(\mathrm{tr}(\mathcal{N}_{U_1,\ldots,U_N}(\pi_{\mathcal{E}}))) = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}(\mathrm{tr}(\mathcal{N}_j(U_j\frac{1}{k}p_jU_j^*))) \quad (132)$$

$$= \frac{1}{N}\sum_{j=1}^{N}\mathrm{tr}(\mathcal{N}_j(\pi_{\mathcal{G}_j})). \quad (133)$$

(101), (105), (114), (131), and (133) show that

$$\mathbb{E}(F_{c,e}(\pi_{\mathcal{E}},\mathcal{N}_{U_1,\ldots,U_N})) \geq \frac{1}{N}\sum_{j=1}^{N}\mathrm{tr}(\mathcal{N}_j(\pi_{\mathcal{G}_j})) - \sum_{j,l=1}^{N}\frac{1}{N}\sqrt{L_{jl}D_{jl}}, \quad (134)$$

where for $j,l\in\{1,\ldots,N\}$ we introduced the abbreviation

$$D_{jl} := \langle\mathcal{N}_j(\pi_{\mathcal{G}_j}),\mathcal{N}_l(\pi_{\mathcal{G}_l})\rangle_{HS}, \quad (135)$$

and, as before,

$$L_{jl} = k\min\{n_j,n_l\}. \quad (136)$$

It is obvious that

$$L_{jl}\leq L_{jj} \quad \text{and} \quad L_{jl}\leq L_{ll} \quad (137)$$

hold. Moreover, the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product shows that

$$D_{jl} = \langle\mathcal{N}_j(\pi_{\mathcal{G}_j}),\mathcal{N}_l(\pi_{\mathcal{G}_l})\rangle_{HS} \quad (138)$$

$$\leq ||\mathcal{N}_j(\pi_{\mathcal{G}_j})||_2||\mathcal{N}_l(\pi_{\mathcal{G}_l})||_2 \quad (139)$$

$$\leq \max\{||\mathcal{N}_j(\pi_{\mathcal{G}_j})||_2^2,||\mathcal{N}_l(\pi_{\mathcal{G}_l})||_2^2\} \quad (140)$$

$$= \max\{D_{jj},D_{ll}\}. \quad (141)$$

Therefore, an application of Lemma 32 allows us to conclude from (134) that

$$\mathbb{E}(F_{c,e}(\pi_{\mathcal{E}},\mathcal{N}_{U_1,\ldots,U_N})) \geq \frac{1}{N}\sum_{j=1}^{N}\mathrm{tr}(\mathcal{N}_j(\pi_{\mathcal{G}_j})) \quad (142)$$

$$-2\sum_{j=1}^{N}\sqrt{kn_j}||\mathcal{N}_j(\pi_{\mathcal{G}_j})||_2, \quad (143)$$

and we are done. □

3.2.4 Entanglement fidelity

The purpose of this subsection is to develop a tool which will enable us to convert a special kind of recovery maps depending on the channel into such that are universal, at least for finite compound channels. Anticipating constructions in section 3.3 below the situation we will be faced with is as follows. For finite set $\mathfrak{I}=\{\mathcal{N}_1,\ldots,\mathcal{N}_N\}$ of channels, block length $l\in\mathbb{N}$, and small $\epsilon>0$ we will be able to find one single recovery map $\mathcal{R}^l$ and a unitary encoder $\mathcal{W}^l$ such that for each $i\in\{1,\ldots,N\}$

$$F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{Q}_{l,i}\circ\mathcal{N}_i^{\otimes l}\circ\mathcal{W}^l)\geq 1-\epsilon, \quad (144)$$

where $\mathcal{Q}_{l,i}(\cdot):=q_{l,i}(\cdot)q_{l,i}$ with suitable projections $q_{l,i}$ acting on $\mathcal{K}^{\otimes l}$. Thus we will effectively end up with the recovery maps $\mathcal{R}_i^l:=\mathcal{R}^l\circ\mathcal{Q}_{l,i}$. Consequently, it turns out that the decoder is *informed*. Lemma 34 below shows how to get rid of the maps $\mathcal{Q}_{l,i}$ ensuring the existence of a universal recovery map for the whole set $\mathfrak{I}$ while decreasing the entanglement fidelity only slightly.

Lemma 34. *Let $\rho \in \mathcal{S}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Let, for some other Hilbert space $\mathcal{K}$, $\mathcal{A} \in \mathcal{C}(\mathcal{H},\mathcal{K})$, $\mathcal{D} \in \mathcal{C}(\mathcal{K},\mathcal{H})$, $q \in \mathcal{B}(\mathcal{K})$ be an orthogonal projection.*

1. *Denoting by $\mathcal{Q}^\perp$ the completely positive map induced by $q^\perp := \mathbf{1}_\mathcal{K} - q$ we have*

$$F_e(\rho, \mathcal{D}\circ\mathcal{A}) \geq F_e(\rho, \mathcal{D}\circ\mathcal{Q}\circ\mathcal{A}) - 2\sqrt{F_e(\rho, \mathcal{D}\circ\mathcal{Q}\circ\mathcal{A})F_e(\rho, \mathcal{D}\circ\mathcal{Q}^\perp\circ\mathcal{A})}. \tag{145}$$

2. *If for some $\epsilon > 0$ the relation $F_e(\rho, \mathcal{D}\circ\mathcal{Q}\circ\mathcal{A}) \geq 1-\epsilon$ holds, then*

$$F_e(\rho, \mathcal{D}\circ\mathcal{Q}^\perp\circ\mathcal{A}) \leq \epsilon, \tag{146}$$

and (145) implies

$$F_e(\rho, \mathcal{D}\circ\mathcal{A}) \geq 1-\epsilon-2\sqrt{\epsilon} \geq 1-3\sqrt{\epsilon}. \tag{147}$$

3. *If for some $\epsilon > 0$ merely the relation $tr\{q\mathcal{A}(\rho)\} \geq 1-\epsilon$ holds then we can conclude that*

$$F_e(\rho, \mathcal{D}\circ\mathcal{A}) \geq F_e(\rho, \mathcal{D}\circ\mathcal{Q}\circ\mathcal{A}) - 2\sqrt{\epsilon}. \tag{148}$$

The following Lemma 35 contains two inequalities one of which will be needed in the proof of Lemma 34.

Lemma 35. *Let $\mathcal{D} \in \mathcal{C}(\mathcal{K},\mathcal{H})$ and $x_1 \perp x_2$, $z_1 \perp z_2$ be state vectors, $x_1, x_2 \in \mathcal{K}$, $z_1, z_2 \in \mathcal{H}$. Then*

$$|\langle z_1, \mathcal{D}(|x_1\rangle\langle x_2|)z_1\rangle| \leq \sqrt{|\langle z_1, \mathcal{D}(|x_1\rangle\langle x_1|)z_1\rangle| \cdot |\langle z_1, \mathcal{D}(|x_2\rangle\langle x_2|)z_1\rangle|} \leq 1, \tag{149}$$

and

$$|\langle z_1, \mathcal{D}(|x_1\rangle\langle x_2|)z_2\rangle| \leq \sqrt{|\langle z_1, \mathcal{D}(|x_1\rangle\langle x_1|)z_1\rangle| \cdot |\langle z_2, \mathcal{D}(|x_2\rangle\langle x_2|)z_2\rangle|} \leq 1. \tag{150}$$

We will utilize only (149) in the proof of Lemma 34. But the inequality (150) might prove useful in other context so that we state it here for completeness.

Proof of Lemma 35. Let $\dim\mathcal{H} = h$, $\dim\mathcal{K} = \kappa$. Extend $\{x_1, x_2\}$ to an orthonormal basis $\{x_1, x_2, \ldots, x_\kappa\}$ of $\mathcal{K}$ and $\{z_1, z_2\}$ to an orthonormal basis $\{z_1, z_2, \ldots, z_h\}$ on $\mathcal{H}$. Since $x_1 \perp x_2$ and $z_1 \perp z_2$, this can always be done. By the theorem of Choi [19], a linear map from $\mathcal{B}(\mathcal{H})$ to $\mathcal{B}(\mathcal{K})$ is completely positive if and only if its Choi matrix is positive. Write $\mathcal{D}(|x_i\rangle\langle x_j|) = \sum_{k,l=1}^h D_{kl}^{ij}|z_k\rangle\langle z_l|$. Then the Choi matrix of $\mathcal{D}$ is, with respect to the bases $\{x_1, \ldots, x_k\}$ and $\{z_1, \ldots, z_h\}$, written as

$$\mathrm{CHOI}(\mathcal{D}) = \sum_{i,j=1}^{\kappa} |x_i\rangle\langle x_j| \otimes \sum_{k,l=1}^{h} D_{kl}^{ij}|z_k\rangle\langle z_l|. \tag{151}$$

If $\mathrm{CHOI}(\mathcal{D})$ is positive, then all principal minors of $\mathrm{CHOI}(\mathcal{D})$ are positive (cf. Corollary 7.1.5 in [40]) and thus

$$|D_{kl}^{ij}| \leq \sqrt{|D_{kk}^{ii}| \cdot |D_{ll}^{jj}|} \tag{152}$$

for every suitable choice of i, j, k, l. Thus

$$\begin{aligned} |\langle z_1|\mathcal{D}(|x_1\rangle\langle x_2|)z_2\rangle| &= |D_{12}^{12}| && (153)\\ &\leq \sqrt{|D_{11}^{11}| \cdot |D_{22}^{22}|} && (154)\\ &= \sqrt{|\langle z_1, \mathcal{D}(|x_1\rangle\langle x_1|)z_1\rangle| \cdot |\langle z_2, \mathcal{D}(|x_2\rangle\langle x_2|)z_2\rangle|}, && (155) \end{aligned}$$

and similarly

$$|\langle z_1, \mathcal{D}(|x_1\rangle\langle x_2|)z_1\rangle| \leq \sqrt{|\langle z_1, \mathcal{D}(|x_1\rangle\langle x_1|)z_1\rangle| \cdot |\langle z_1, \mathcal{D}(|x_2\rangle\langle x_2|)z_1\rangle|}. \tag{156}$$

The fact that $\mathcal{D}$ is trace preserving gives us the estimate $\langle z_i, \mathcal{D}(|x_j\rangle\langle x_j|)z_i\rangle \leq 1$ (i, j suitably chosen) and we are done. □

Proof of Lemma 34. Let $\dim \mathcal{H} = h$, $\dim \mathcal{K} = \kappa$, $|\psi\rangle\langle\psi| \in \mathcal{H}_a \otimes \mathcal{H}$ be a purification of ρ (w.l.o.g. $\mathcal{H}_a = \mathcal{H}$). Set $\tilde{\mathcal{D}} := id_{\mathcal{H}_a} \otimes \mathcal{D}$, $\tilde{\mathcal{A}} := id_{\mathcal{H}_a} \otimes \mathcal{A}$, $\tilde{q} := \mathbf{1}_{\mathcal{H}_a} \otimes q$ and, as usual, $\tilde{q}^\perp$ the orthocomplement of $\tilde{q}$ within $\mathcal{H}_a \otimes \mathcal{K}$. Obviously,

$$\begin{aligned}
F_e(\rho, \mathcal{D} \circ \mathcal{A}) &= \langle \psi, \tilde{\mathcal{D}} \circ \tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\psi\rangle && (157)\\
&= \langle \psi, \tilde{\mathcal{D}}([\tilde{q} + \tilde{q}^\perp]\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|[\tilde{q} + \tilde{q}^\perp]))\psi\rangle && (158)\\
&= \langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q})\psi\rangle + \langle \psi, \tilde{\mathcal{D}}(\tilde{q}^\perp\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q}^\perp)\psi\rangle && (159)\\
&\quad + \langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q}^\perp)\psi\rangle + \langle \psi, \tilde{\mathcal{D}}(\tilde{q}^\perp\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q})\psi\rangle && (160)\\
&\geq \langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q})\psi\rangle + 2\Re\{\langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q}^\perp)\psi\rangle\} && (161)\\
&\geq \langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q})\psi\rangle - 2|\langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q}^\perp)\psi\rangle| && (162)\\
&= F_e(\rho, \mathcal{D} \circ \mathcal{Q} \circ \mathcal{A}) - 2|\langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q}^\perp)\psi\rangle|. && (163)
\end{aligned}$$

We establish a lower bound on the second term on the RHS of (163). Let

$$\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|) = \sum_{i=1}^{\kappa \cdot h} \lambda_i |a_i\rangle\langle a_i|, \tag{164}$$

where $\{a_1, \ldots, a_{\kappa \cdot h}\}$ are assumed to form an orthonormal basis. Now every a_i can be written as $a_i = \alpha_i x_i + \beta_i y_i$ where $x_i \in \mathrm{supp}(\tilde{q})$ and $y_i \in \mathrm{supp}(\tilde{q}^\perp)$, $i \in \{1, ..., \kappa \cdot h\}$, are state vectors and $\alpha_i, \beta_i \in \mathbb{C}$. Define $\sigma := \tilde{A}(|\psi\rangle\langle\psi|)$, then

$$\sigma = \sum_{j=1}^{\kappa \cdot h} \lambda_j (|\alpha_j|^2 |x_j\rangle\langle x_j| + \alpha_j \beta_j^* |x_j\rangle\langle y_j| + \beta_j \alpha_j^* |y_j\rangle\langle x_j| + |\beta_j|^2 |y_j\rangle\langle y_j|). \tag{165}$$

Set $X := |\langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q}^\perp)\psi\rangle|$. Then

$$\begin{aligned}
X &= |\langle \psi, \tilde{\mathcal{D}}(\tilde{q}\sigma q^\perp)\psi\rangle| && (166)\\
&\overset{\mathbf{a}}{=} |\sum_{i=1}^{\kappa \cdot h} \lambda_i \langle \psi, \tilde{\mathcal{D}}(\tilde{q}|a_i\rangle\langle a_i|\tilde{q}^\perp)\psi\rangle| && (167)\\
&= |\sum_{i=1}^{\kappa \cdot h} \lambda_i \alpha_i \beta_i^* \langle \psi, \tilde{\mathcal{D}}(|x_i\rangle\langle y_i|)\psi\rangle| && (168)\\
&\leq \sum_{i=1}^{\kappa \cdot h} |\lambda_i \alpha_i \beta_i^*| \cdot |\langle \psi, \tilde{\mathcal{D}}(|x_i\rangle\langle y_i|)\psi\rangle| && (169)\\
&\overset{\mathbf{b}}{\leq} \sum_{i=1}^{\kappa \cdot h} |\sqrt{\lambda_i |\langle \psi, \tilde{\mathcal{D}}(|x_i\rangle\langle x_i|)\psi\rangle}\alpha_i \sqrt{\lambda_i \langle \psi, \tilde{\mathcal{D}}(|y_i\rangle\langle y_i|)\psi\rangle}\beta_i^*| && (170)\\
&\overset{\mathbf{c}}{\leq} \sqrt{\sum_{i=1}^{\kappa \cdot h} \lambda_i |\alpha_i|^2 \langle \psi, \tilde{\mathcal{D}}(|x_i\rangle\langle x_i|)\psi\rangle \sum_{j=1}^{\kappa \cdot h} \lambda_j |\beta_j|^2 \langle \psi, \tilde{\mathcal{D}}(|y_j\rangle\langle y_j|)\psi\rangle}. && (171)
\end{aligned}$$

Here, **a** follows from using the convex decomposition of $\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)$, **b** from utilizing inequality (149) from Lemma 35 and **c** is an application of the Cauchy-Schwarz inequality.
Now, employing the representation (165) it is easily seen that

$$F_e(\rho, \mathcal{D} \circ \mathcal{Q} \circ \mathcal{A}) = \langle \psi, \tilde{\mathcal{D}}(\tilde{q}\tilde{\mathcal{A}}(|\psi\rangle\langle\psi|)\tilde{q})\psi\rangle = \sum_{i=1}^{\kappa \cdot h} \lambda_i |\alpha_i|^2 \langle \psi, \tilde{\mathcal{D}}(|x_i\rangle\langle , x_i|)\psi\rangle \tag{172}$$

and similarly

$$F_e(\rho, \mathcal{D} \circ \mathcal{Q}^\perp \circ \mathcal{A}) = \sum_{j=1}^{\kappa \cdot h} \lambda_j |\beta_j|^2 \langle \psi, \tilde{\mathcal{D}}(|y_j\rangle\langle y_j|)\psi\rangle. \tag{173}$$

The inequalities (173), (172), (171), and (163) yield

$$F_e(\rho, \mathcal{D} \circ \mathcal{A}) \quad \geq \quad F_e(\rho, \mathcal{D} \circ \mathcal{Q} \circ \mathcal{A}) - 2\sqrt{F_e(\rho, \mathcal{D} \circ \mathcal{Q} \circ \mathcal{A}) F_e(\rho, \mathcal{D} \circ \mathcal{Q}^\perp \circ \mathcal{A})} \tag{174}$$

which establishes (145).
Let us turn now to the other assertions stated in the lemma. Let $\mathrm{tr}\{q\mathcal{A}(\rho)\} \geq 1 - \epsilon$. This implies $\mathrm{tr}(q^\perp \mathcal{A}(\rho)) \leq \epsilon$. A direct calculation yields

$$\begin{aligned}
\mathrm{tr}(\tilde{q}^\perp \sigma) &= \mathrm{tr}_{\mathcal{H}_a}(\mathrm{tr}_{\mathcal{K}}((\mathbf{1}_{\mathcal{H}_a} \otimes q^\perp) id_{\mathcal{H}_a} \otimes \mathcal{A}(|\psi\rangle\langle\psi|))) && (175)\\
&= \mathrm{tr}_{\mathcal{K}}(q^\perp \mathcal{A}(\mathrm{tr}_{\mathcal{H}_a}(|\psi\rangle\langle\psi|))) && (176)\\
&= \mathrm{tr}_{\mathcal{K}}(q^\perp \mathcal{A}(\rho)) && (177)\\
&\leq \epsilon. && (178)
\end{aligned}$$

Using (165), we get the useful inequality

$$\begin{aligned}
\epsilon &\geq \mathrm{tr}(\tilde{q}^\perp \sigma) && (179)\\
&= \sum_{i=1}^{\kappa \cdot h} \lambda_i |\beta_i|^2 \mathrm{tr}(\tilde{q}^\perp |y_i\rangle\langle y_i|) && (180)\\
&= \sum_{i=1}^{\kappa \cdot h} \lambda_i |\beta_i|^2. && (181)
\end{aligned}$$

Using Lemma 35 and (181) we get

$$\begin{aligned}
X &\leq \sqrt{\sum_{i=1}^{\kappa \cdot h} \lambda_i |\alpha_i|^2 \sum_{j=1}^{\kappa \cdot h} \lambda_j |\beta_j|^2} && (182)\\
&\leq \sqrt{\epsilon}, && (183)
\end{aligned}$$

thus by equation (163) we have

$$F_e(\rho, \mathcal{D} \circ \mathcal{A}) \geq F_e(\rho, \mathcal{D} \circ \mathcal{Q} \circ \mathcal{A}) - 2\sqrt{\epsilon}. \tag{184}$$

□

3.3 Direct part of the coding theorem for finitely many channels

In the next two pages we briefly collect some well-known properties of frequency typical projections and reduced operations.

3.3.1 Typical projections

In this subsection we will collect some well-known results on typical projections.
Let $\rho \in \mathcal{S}(\mathcal{H})$ be a state and consider any diagonalization

$$\rho = \sum_{i=1}^{d} \lambda_i |e_i\rangle\langle e_i|, \tag{185}$$

where $d := \dim \mathcal{H}$. Using this representation the state $\rho^{\otimes l}$ can be written as

$$\rho^{\otimes l} = \sum_{x^l \in A^l} \lambda_{x^l} |e_{x^l}\rangle\langle e_{x^l}|, \tag{186}$$

with $A = \{1, \ldots, d\}$, $x^l := (x_1, \ldots, x_l) \in A^l$, $\lambda_{x^l} := \lambda_{x_1} \cdot \ldots \cdot \lambda_{x_l}$, and $e_{x^l} := e_{x_1} \otimes \ldots \otimes e_{x_l}$.
The frequency-typical set of eigenvalues of ρ is given by

$$T_{\delta,l} := \{x^l \in A^l : ||p_{x^l} - \lambda||_1 < \delta,\ p_{x^l} \ll \lambda\}, \tag{187}$$

where p_{x^l} denotes the empirical probability distribution on A generated by x^l, i.e.

$$p_{x^l}(x) := \frac{|\{j \in \{1, \ldots, l\} : x_j = x\}|}{l}, \tag{188}$$

λ is the probability distribution on A defined by the eigenvalues of ρ, and $p_{x^l} \ll \lambda$ means that $p_{x^l}(x) = 0$ whenever $\lambda_x = 0$.
The frequency-typical projection $q_{\delta,l}$ of ρ given by

$$q_{\delta,l} := \sum_{x^l \in T_{\delta,l}} |e_{x^l}\rangle\langle e_{x^l}| \tag{189}$$

has the following well-known properties:

Lemma 36. *There is a real number $c > 0$ such that for every Hilbert space $\mathcal{H}$ there exist functions $h : \mathbb{N} \to \mathbb{R}_+$, $\varphi : (0, 1/2) \to \mathbb{R}_+$ with $\lim_{l\to\infty} h(l) = 0$ and $\lim_{\delta\to 0} \varphi(\delta) = 0$ such that for any $\rho \in \mathcal{S}(\mathcal{H})$, $\delta \in (0, 1/2)$, $l \in \mathbb{N}$ there is an orthogonal projection $q_{\delta,l} \in \mathcal{B}(\mathcal{H})^{\otimes l}$ called frequency-typical projection that satisfies*

1. $tr(\rho^{\otimes l} q_{\delta,l}) \geq 1 - 2^{-l(c\delta^2 - h(l))}$,
2. $q_{\delta,l}\rho^{\otimes l} q_{\delta,l} \leq 2^{-l(S(\rho)-\varphi(\delta))} q_{\delta,l}$
3. $\eta_l(\delta) 2^{l(S(\rho)-\varphi(\delta))} \leq tr(q_{\delta,l}) \leq 2^{l(S(\rho)+\varphi(\delta))}$ *where*

$$\eta_l(\delta) := 1 - 2^{-l(c\delta^2 - h(l))}. \tag{190}$$

The inequality 2. implies

$$||q_{\delta,l}\rho^{\otimes l} q_{\delta,l}||_2^2 \leq 2^{-l(S(\rho)-\varphi(\delta))}. \tag{191}$$

Moreover, setting $d := \dim \mathcal{H}$, φ and h are given by

$$h(l) = \frac{d}{l} \log(l + 1) \quad \forall l \in \mathbb{N}, \quad \varphi(\delta) = -\delta \log \frac{\delta}{d}\ \forall \delta \in (0, 1/2). \tag{192}$$

The proof of the lemma is fairly standard and rests on purely classical reasoning. It combines the Bernstein-Sanov trick (cf. [64], sect. III.1) and the type counting methods from [20].

3.3.2 Typical kraus operators

According to Kraus' representation theorem we can find to any $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ a family of operators $a_1, \ldots, a_n \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ with $\sum_{i=1}^n a_i^* a_i = \mathbf{1}_{\mathcal{H}}$ and

$$\mathcal{N}(\rho) = \sum_{i=1}^n a_i \rho a_i^* \tag{193}$$

for all $\rho \in \mathcal{S}(\mathcal{H})$.
We fix the maximally mixed state $\pi_{\mathcal{G}}$ supported by the subspace $\mathcal{G}$ of $\mathcal{H}$. It is easily seen (cf. [56]) that the Kraus operators $a_1, \ldots, a_n$ of $\mathcal{N}$ can always be chosen such that

$$\mathrm{tr}(a_i \pi_{\mathcal{G}} a_j^*) = \delta_{ij} \mathrm{tr}(a_i \pi_{\mathcal{G}} a_i^*), \tag{194}$$

for all $i, j \in \{1, \ldots, n\}$. With this choice of Kraus operators we can define a probability distribution r on the set $B := \{1, \ldots, n\}$ by

$$r(i) := \mathrm{tr}(a_i \pi_{\mathcal{G}} a_i^*), \qquad (i \in B). \tag{195}$$

It is shown in [59] that the Shannon entropy of r is nothing else than the entropy exchange $S_e(\pi_{\mathcal{G}}, \mathcal{N})$, i.e.

$$H(r) = S_e(\pi_{\mathcal{G}}, \mathcal{N}). \tag{196}$$

In a similar vein as in the previous subsection we introduce the notion of frequency-typical subset for r, i.e we set

$$K_{\delta,l} := \{y^l \in B^l : ||p_{y^l} - r||_1 < \delta,\ p_{y^l} \ll r\} \tag{197}$$

with $\delta > 0$. With this we can introduce the notion of the reduced operation by setting

$$\mathcal{N}_{\delta,l}(\rho) := \sum_{y^l \in K_{\delta,l}} a_{y^l} \rho a_{y^l}^*, \tag{198}$$

where $a_{y^l} := a_{y_1} \otimes \ldots \otimes a_{y_l}$ and $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$. Moreover, we set

$$n_{\delta,l} := |K_{\delta,l}|, \tag{199}$$

which is the number of Kraus operators of the reduced operation $\mathcal{N}_{\delta,l}$. The properties of frequency-typical sets (cf. [20, 64])) lead immediately to

Lemma 37. *Let $\mathcal{H}, \mathcal{K}$ be finite dimensional Hilbert spaces. There are functions $\gamma : (0, 1/2) \to \mathbb{R}_+$, $h' : \mathbb{N} \to \mathbb{R}_+$ satisfying $\lim_{\delta \to 0} \gamma(\delta) = 0$ and $h'(l) \searrow 0$ such that for each $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$, $\delta \in (0, 1/2)$, $l \in \mathbb{N}$ and maximally mixed state $\pi_{\mathcal{G}}$ on some subspace $\mathcal{G} \subset \mathcal{H}$ there is an operation $\mathcal{N}_{\delta,l} \in \mathcal{C}^{\downarrow}(\mathcal{H}^{\otimes l}, \mathcal{K}^{\otimes l})$ called reduced operation with respect to $\mathcal{N}$ and $\pi_{\mathcal{G}}$ that satisfies*

1. *$tr(\mathcal{N}_{\delta,l}(\pi_{\mathcal{G}}^{\otimes l})) \geq 1 - 2^{-l(c'\delta^2 - h'(l))}$, with a universal positive constant $c' > 0$,*
2. *$\mathcal{N}_{\delta,l}$ has a Kraus representation with at most $n_{\delta,l} \leq 2^{l(S_e(\pi_{\mathcal{G}}, \mathcal{N}) + \gamma(\delta) + h'(l))}$ Kraus operators.*
3. *For every state $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$ and every two channels $\mathcal{I} \in \mathcal{C}^{\downarrow}(\mathcal{H}^{\otimes l}, \mathcal{H}^{\otimes l})$ and $\mathcal{L} \in \mathcal{C}^{\downarrow}(\mathcal{K}^{\otimes l}, \mathcal{H}^{\otimes l})$ the inequality $F_e(\rho, \mathcal{L} \circ \mathcal{N}_{\delta,l} \circ \mathcal{I}) \leq F_e(\rho, \mathcal{L} \circ \mathcal{N}^{\otimes l} \circ \mathcal{I})$ is fulfilled.*

Setting $d := \dim \mathcal{H}$ and $\kappa := \dim \mathcal{K}$, the function $h' : \mathbb{N} \to \mathbb{R}_+$ is given by $h'(l) = \frac{d \cdot \kappa}{l} \log(l+1)\ \forall l \in \mathbb{N}$ and γ by $\gamma(\delta) = -\delta \log \frac{\delta}{d \cdot \kappa},\ \forall \delta \in (0, 1/2)$.

3.3.3 The case of uninformed users

Let us consider a compound channel given by a finite set $\mathfrak{I} := \{\mathcal{N}_1, \ldots, \mathcal{N}_N\} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ and a subspace $\mathcal{G} \subset \mathcal{H}$. For every $l \in \mathbb{N}$, we choose a subspace $\mathcal{E}_l \subset \mathcal{G}^{\otimes l}$. As usual, $\pi_{\mathcal{E}_l}$ and $\pi_{\mathcal{G}}$ denote the maximally mixed states on $\mathcal{E}_l$, respectively $\mathcal{G}$ while $k_l := \dim \mathcal{E}_l$ gives the dimension of $\mathcal{E}_l$.
For $j \in \{1, \ldots, N\}$, $\delta \in (0, 1/2)$, $l \in \mathbb{N}$ and states $\mathcal{N}_j(\pi_{\mathcal{G}})$ let $q_{j,\delta,l} \in \mathcal{B}(\mathcal{K})^{\otimes l}$ be the frequency-typical projection of $\mathcal{N}_j(\pi_{\mathcal{G}})$ and $\mathcal{N}_{j,\delta,l}$ be the reduced operation associated with $\mathcal{N}_j$ and $\pi_{\mathcal{G}}$ as defined in Lemma 37.

These quantities enable us to define a new set of channels that is more adapted to our problem than the original one. We set for an arbitrary unitary operation $u^l \in \mathcal{B}(\mathcal{H}^{\otimes l})$

$$\hat{\mathcal{N}}^l_{j,u^l,\delta} := \mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_{j,\delta,l} \circ \mathcal{U}^l \tag{200}$$

and, accordingly,

$$\hat{\mathcal{N}}^l_{u^l,\delta} := \frac{1}{N} \sum_{j=1}^{N} \hat{\mathcal{N}}^l_{j,u^l,\delta}. \tag{201}$$

We will show the existence of good codes for the reduced channels $\mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_{j,\delta,l}$ in the limit of large $l \in \mathbb{N}$. An application of Lemma 34 and Lemma 37 will then show that these codes are also good for the original compound channel.
Let U^l be a random variable taking values in $\mathfrak{U}(\mathcal{G}^{\otimes l})$ which is distributed according to the Haar measure. Application of Theorem 30 yields

$$\mathbb{E}F_{c,e}(\pi_{\mathcal{E}_l}, \hat{\mathcal{N}}^l_{U^l,\delta}) \quad \geq \quad \mathrm{tr}(\hat{\mathcal{N}}^l_\delta(\pi^{\otimes l}_{\mathcal{G}})) - 2\sum_{j=1}^{N} \sqrt{k_l n_{j,\delta,l}} ||\hat{\mathcal{N}}^l_{j,\delta}(\pi^{\otimes l}_{\mathcal{G}})||_2, \tag{202}$$

where $n_{j,\delta,l}$ stands for the number of Kraus operators of the reduced operation $\mathcal{N}_{j,\delta,l}$ ($j \in \{1, \ldots, N\}$) and

$$\hat{\mathcal{N}}^l_{j,\delta} := \mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_{j,\delta,l}, \tag{203}$$

$$\hat{\mathcal{N}}^l_{\delta} := \frac{1}{N} \sum_{j=1}^{N} \hat{\mathcal{N}}^l_{j,\delta}. \tag{204}$$

Notice that $\mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_{j,\delta,l}$ trivially has a Kraus representation containing exactly $n_{j,\delta,l}$ elements. We will use inequality (202) in the proof of the following theorem.

Theorem 38 (Direct Part: Uninformed Users and $|\mathfrak{I}| < \infty$)**.** *Let $\mathfrak{I} = \{\mathcal{N}_1, ..., \mathcal{N}_N\} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a compound channel and $\pi_{\mathcal{G}}$ the maximally mixed state associated to a subspace $\mathcal{G} \subset \mathcal{H}$. Then*

$$Q(\mathfrak{I}) \geq \min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i). \tag{205}$$

Moreover, for any $R < \min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i)$ there is $c_R > 0$, $l_0 \in \mathbb{N}$ and at least one sequence of (l, k_l) codes such that both

$$\liminf_{l\to\infty} \frac{1}{l} \log(k_l) = R \qquad \textit{and} \qquad \min_{1\leq i\leq N} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) \geq 1 - 2^{-c_R l} \tag{206}$$

Proof. We show that for every $\epsilon > 0$ the number $\min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i) - \epsilon$ is an achievable rate for $\mathfrak{I}$.
1) If $\min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i) - \epsilon \leq 0$, there is nothing to prove.
2) Let $\min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i) - \epsilon > 0$.
Choose $\delta \in (0, 1/2)$ and $l_0 \in \mathbb{N}$ satisfying $\gamma(\delta) + \varphi(\delta) + h'(l_0) \leq \epsilon/2$ with functions γ, φ, h' from Lemma 36 and 37.
Now choose for every $l \in \mathbb{N}$ a subspace $\mathcal{E}_l \subset \mathcal{G}^{\otimes l}$ such that

$$\dim \mathcal{E}_l =: k_l = \lfloor 2^{l(\min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i) - \epsilon)} \rfloor. \tag{207}$$

By $S(\pi_{\mathcal{G}}) \geq I_c(\pi_{\mathcal{G}}, \mathcal{N}_j)$ (see [11]), this is always possible.
Obviously,

$$\min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i) - \epsilon - o(l^0) \leq \frac{1}{l} \log k_l \leq \min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i) - \epsilon. \tag{208}$$

We will now give lower bounds on the terms in (202), thereby making use of Lemma 36 and Lemma 37:

$$\mathrm{tr}(\hat{\mathcal{N}}_\delta^l(\pi_{\mathcal{G}}^{\otimes l})) \geq 1 - 2^{-l(c\delta^2 - h(l))} - 2^{-l(c'\delta^2 - h'(l))}. \tag{209}$$

A more detailed calculation can be found in [15] or [47]. Further, and additionally using the inequality $||A+B||_2^2 \geq ||A||_2^2 + ||B||_2^2$ valid for non-negative operators $A, B \in \mathcal{B}(\mathcal{K}^{\otimes l})$ (see [47]), we get the inequality

$$||\hat{\mathcal{N}}_{j,\delta}^l(\pi_{\mathcal{G}}^{\otimes l})||_2^2 \quad \leq \quad 2^{-l(S(\mathcal{N}_j(\pi_{\mathcal{G}})) - \varphi(\delta))}. \tag{210}$$

From (202), (209), (210) and our specific choice of k_l it follows that

$$\begin{aligned} \mathbb{E}F_{c,e}(\pi_{\mathcal{E}_l}, \hat{\mathcal{N}}_{U^l,\delta}^l) \quad &\geq \quad 1 - 2^{-l(c\delta^2 - h(l))} - 2^{-l(c'\delta^2 - h'(l))} \\ &\qquad -2\sum_{j=1}^N \sqrt{2^{l(\frac{1}{l}\log k_l + \gamma(\delta) + \varphi(\delta) + h'(l) - I_c(\pi_{\mathcal{G}}, \mathcal{N}_j)}} \qquad (211) \\ &\geq \quad 1 - 2^{-l(c\delta^2 - h(l))} - 2^{-l(c'\delta^2 - h'(l))} - 2N\sqrt{2^{-l(\epsilon - \gamma(\delta) - \varphi(\delta) - h'(l))}}. \qquad (212) \end{aligned}$$

Since $\epsilon - \gamma(\delta) - \varphi(\delta) - h'(l) \geq \varepsilon/2$ for every $l \geq l_0$, this shows the existence of at least one sequence of $(l, k_l)-$codes for $\mathfrak{I}$ with uninformed users and

$$\liminf_{l\to\infty} \frac{1}{l} \log k_l = \min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i) - \epsilon \tag{213}$$

as well as (using that entanglement fidelity is affine in the channel), for every $l \in \mathbb{N}$,

$$\min_{j \in \{1,\ldots,N\}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \hat{\mathcal{N}}_{j,\delta}^l \circ \mathcal{W}^l) \geq 1 - N\frac{1}{3}\epsilon_l \tag{214}$$

where $w^l \in \mathfrak{U}(\mathcal{G}^{\otimes l})$ $\forall l \in \mathbb{N}$ and

$$\epsilon_l = 3 \cdot (2^{-l(c\delta^2 - h(l))} + 2^{-l(c'\delta^2 - h'(l))} + 2N\sqrt{2^{-l(\epsilon - \gamma(\delta) - \varphi(\delta) - h'(l))}}). \tag{215}$$

Note that $\lim_{l\to\infty} \epsilon_l = 0$ exponentially fast, as can be seen from our choice of δ and l_0. We let $c_R > 0$ denote the largest real number such that

$$\varepsilon_l \leq 2^{-c_R l}. \tag{216}$$

For every $j \in \{1, \ldots, N\}$ and $l \in \mathbb{N}$ we thus have, by property 3. of Lemma 37, construction of $\hat{\mathcal{N}}_{j,w^j,\delta}^l$, and equation (214),

$$\begin{aligned} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_j^{\otimes l} \circ \mathcal{W}^l) \quad &\geq \quad F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_{j,\delta,l} \circ \mathcal{W}^l) \qquad (217) \\ &= \quad F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \hat{\mathcal{N}}_{j,w^j,\delta}^l) \qquad (218) \\ &\geq \quad 1 - N\frac{1}{3}\epsilon_l. \qquad (219) \end{aligned}$$

By the first two parts of Lemma 34, this immediately implies

$$\min_{\mathcal{N}_j \in \mathfrak{I}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_j^{\otimes l} \circ \mathcal{W}^l) \geq 1 - \sqrt{3N\epsilon_l} \quad \forall l \in \mathbb{N}. \tag{220}$$

Since $\epsilon > 0$ was arbitrary, we have shown that $\min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}}, \mathcal{N}_i)$ is an achievable rate. □

3.3.4 The informed encoder

In this subsection we shall prove the following Theorem:

Theorem 39 (Direct Part: Informed Encoder and $|\mathfrak{I}| < \infty$)**.** *For every finite compound channel* $\mathfrak{I} = \{\mathcal{N}_1, \ldots, \mathcal{N}_N\} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ *and any set* $\{\pi_{\mathcal{G}_1}, \ldots, \pi_{\mathcal{G}_N}\}$ *of maximally mixed states on subspaces* $\{\mathcal{G}_1, \ldots, \mathcal{G}_N\}$ *with* $\mathcal{G}_i \subset \mathcal{H}$ *for all* $i \in \{1, \ldots, N\}$ *we have*

$$Q_{IE}(\mathfrak{I}) \geq \min_{\mathcal{N}_i \in \mathfrak{I}} I_c(\pi_{\mathcal{G}_i}, \mathcal{N}_i). \tag{221}$$

Proof. Let a compound channel be given by a finite set $\mathfrak{I} := \{\mathcal{N}_1, \ldots, \mathcal{N}_N\} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ and let $\mathcal{G}_1, \ldots, \mathcal{G}_N$ be arbitrary subspaces of $\mathcal{H}$. We will prove that for every $\epsilon > 0$ the value

$$R(\epsilon) := \min_{1 \leq i \leq N} I_c(\pi_{\mathcal{G}_i}, \mathcal{N}_i) - \epsilon \tag{222}$$

is achievable. If $R(\epsilon) \leq 0$, there is nothing to prove. Hence we assume $R(\epsilon) > 0$. For every $l \in \mathbb{N}$ and all $i \in \{1, \ldots, N\}$ we choose the following. First, a subspace $\mathcal{E}_l \subset \mathcal{H}^{\otimes l}$ of dimension $k_l := \dim \mathcal{E}_l$ that satisfies $k_l \leq \dim \mathcal{G}_i^{\otimes l}$. Second, a set $\{v_1^l, \ldots, v_N^l\}$ of unitary operators with the property $v_i^l \mathcal{E}_l \subset \mathcal{G}_i^{\otimes l}$. Again, the maximally mixed states associated to the above mentioned subspaces are denoted by $\pi_{\mathcal{E}_l}$ on $\mathcal{E}_l$ and $\pi_{\mathcal{G}_i}$ on $\mathcal{G}_i$.
For $j \in \{1, \ldots, N\}$, $\delta \in (0, 1/2)$, $l \in \mathbb{N}$ and states $\mathcal{N}_j(\pi_{\mathcal{G}_j})$ let $q_{j,\delta,l} \in \mathcal{B}(\mathcal{K})^{\otimes l}$ be the frequency-typical projection of $\mathcal{N}_j(\pi_{\mathcal{G}_j})$ and $\mathcal{N}_{j,\delta,l}$ be the reduced operation associated with $\mathcal{N}_j$ and $\pi_{\mathcal{G}_j}$ as considered in Lemmas 36 and 37.
Let, for the moment, $l \in \mathbb{N}$ be fixed. We define a new set of channels that is more adapted to our problem than the original one. We set, for an arbitrary set $\{u_1^l, \ldots, u_N^l\}$ of unitary operators on $\mathcal{H}^{\otimes l}$

$$\tilde{\mathcal{N}}_{j,\delta}^l := \mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_{j,\delta,l}, \tag{223}$$

$$\hat{\mathcal{N}}_{j,u_j^l,\delta}^l := \tilde{\mathcal{N}}_{j,\delta}^l \circ \mathcal{U}_j^l \circ \mathcal{V}_j^l \tag{224}$$

and, accordingly,

$$\hat{\mathcal{N}}_{u_1^l,\ldots,u_N^l,\delta}^l := \frac{1}{N} \sum_{j=1}^N \hat{\mathcal{N}}_{j,u_j^l,\delta}^l. \tag{225}$$

we will first show the existence of good unitary encodings and recovery operation for $\{\tilde{\mathcal{N}}_{1,\delta}^l, \ldots, \tilde{\mathcal{N}}_{N,\delta}^l\}$. Like in the previous subsection, application of Lemma 34 will enable us to show the existence of reliable encodings and recovery operation for the original compound channel $\mathfrak{I}$.
Let $U_1^l, \ldots, U_N^l$ be independent random variables such that each U_i^l takes on values in $\mathfrak{U}(\mathcal{G}_i^{\otimes l})$ and is distributed according to the Haar measure on $\mathfrak{U}(\mathcal{G}_i^{\otimes l})$ $(i \in \{1, \ldots, N\})$. By Theorem 33 we get the lower bound

$$\mathbb{E} F_{c,e}(\pi_{\mathcal{E}_l}, \hat{\mathcal{N}}_{U_1^l,\ldots,U_N^l,\delta}^l) \geq \sum_{j=1}^N [\frac{1}{N} \mathrm{tr}(\tilde{\mathcal{N}}_{j,\delta}^l(\pi_{\mathcal{G}_j^{\otimes l}})) - 2\sqrt{k_l n_{j,\delta,l}} || \tilde{\mathcal{N}}_{j,\delta}^l(\pi_{\mathcal{G}_j^{\otimes l}}) ||_2], \tag{226}$$

where $n_{j,\delta,l}$ denotes the number of Kraus operators in the operations $\tilde{\mathcal{N}}_{j,\delta,l}$ $(j \in \{1, \ldots, N\})$. By Lemmas 36,37 for every $j \in \{1, \ldots, N\}$ the corresponding term in the above sum can be bounded from below through

$$\frac{1}{N} \mathrm{tr}(\tilde{\mathcal{N}}_{j,\delta}^l(\pi_{\mathcal{G}_j^{\otimes l}})) \geq \frac{1}{N}(1 - 2^{-l(c\delta^2 - h(l))} - 2^{-l(c'\delta^2 - h'(l))}) \tag{227}$$

and

$$-2\sqrt{k_l n_{j,\delta,l}}||\tilde{\mathcal{N}}^l_{j,\delta}(\pi_{\mathcal{G}_j^{\otimes l}})||_2 \geq -2\sqrt{k_l \cdot 2^{l(-\min_{1\leq j\leq N} I_c(\pi_{\mathcal{G}_j},\mathcal{N}_j)+\gamma(\delta)+\varphi(\delta)+h'(l))}}. \tag{228}$$

Set $k_l := \lfloor 2^{lR(\epsilon)} \rfloor$. Obviously, for any $j \in \{1,\ldots,N\}$,

$$k_l \cdot 2^{l(-\min_{1\leq j\leq N} I_c(\pi_{\mathcal{G}_j},\mathcal{N}_j))} \leq 2^{-l\epsilon}. \tag{229}$$

This implies

$$\mathbb{E}F_{c,e}(\pi_{\mathcal{E}_l},\hat{\mathcal{N}}^l_{U_1^l,\ldots,U_N^l,\delta}) \quad \geq \quad 1-2^{-l(c\delta^2-h(l))}-2^{-l(c'\delta^2-h'(l))}-2N\sqrt{2^{l(-\epsilon+\gamma(\delta)+\varphi(\delta)+h'(l))}}. \tag{230}$$

Now choosing both the approximation parameter δ and an integer $l_0 \in \mathbb{N}$ such that $-\epsilon+\gamma(\delta)+\varphi(\delta)+h'(l) < -\frac{1}{2}\epsilon$ holds for every $l \geq l_0$ and setting

$$\epsilon_l := 2^{-l(c\delta^2-h(l))}+2^{-l(c'\delta^2-h'(l))}+2N\sqrt{2^{l(-\epsilon+\gamma(\delta)+\varphi(\delta)+h'(l))}} \tag{231}$$

we see that

$$\mathbb{E}F_{c,e}(\pi_{\mathcal{E}_l},\hat{\mathcal{N}}^l_{U_1^l,\ldots,U_N^l,\delta}) \quad \geq \quad 1-\epsilon_l, \tag{232}$$

where again $\epsilon_l \searrow 0$ and our choice of δ and l_0 again shows that the speed of convergence is exponentially fast. Thus, there exist unitary operators $w_1^l,\ldots,w_N^l \subset \mathfrak{U}(\mathcal{H}^{\otimes l})$ and a recovery operation $\mathcal{R}^l$ such that, passing to the individual channels, we have for every $j \in \{1,\ldots,N\}$

$$F_e(\pi_{\mathcal{E}_l},\mathcal{R}^l \circ \mathcal{Q}_{j,\delta,l} \circ \mathcal{N}_{j,\delta,l} \circ \mathcal{W}_j^l) \geq 1-N\epsilon_l. \tag{233}$$

By property 3. of Lemma 37 and Lemma 34, we immediately see that

$$F_e(\pi_{\mathcal{E}_l},\mathcal{R}^l \circ \mathcal{N}_j^{\otimes l} \circ \mathcal{W}_j^l) \quad \geq \quad 1-3\sqrt{N\epsilon_l} \quad \forall j \in \{1,\ldots,N\} \tag{234}$$

is valid as well. We finally get the desired result: For every set $\{\pi_{\mathcal{G}_1},\ldots,\pi_{\mathcal{G}_N}\}$ of maximally mixed states on subspaces $\mathcal{G}_1,\ldots,\mathcal{G}_N \subset \mathcal{H}$ and every $\epsilon > 0$ there exists a sequence of (l,k_l) codes for $\mathfrak{I}$ with informed encoder with the properties

1. $\liminf_{l\to\infty} \frac{1}{l}\log k_l = \min_{\mathcal{N}_j\in\mathfrak{I}} I_c(\pi_{\mathcal{G}_j},\mathcal{N}_j)-\epsilon$,
2. $\min_{\mathcal{N}_j\in\mathfrak{I}} F_e(\pi_{\mathcal{E}_l},\mathcal{R}^l \circ \mathcal{N}_j^{\otimes l} \circ \mathcal{W}_j^l) \geq 1-3\sqrt{N\epsilon_l}$.

Since $\epsilon > 0$ was arbitrary and $\epsilon_l \searrow 0$, we are done. □

3.4 Finite approximations in the set of quantum channels

Our goal in this section is to discretize a given set of channels $\mathfrak{I} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ in such a way that the results derived so far for finite sets can be employed to derive general versions of coding theorems for compound channels.
The first concept we will need is that of a τ-net in the set $\mathcal{C}(\mathcal{H},\mathcal{K})$ and we will give an upper bound on the cardinality of the best τ-net in that set. Best τ-nets characterize the degree of compactness of $\mathcal{C}(\mathcal{H},\mathcal{K})$.
A τ-net in $\mathcal{C}(\mathcal{H},\mathcal{K})$ is a finite set $\{\mathcal{N}_i\}_{i=1}^N$ with the property that for each $\mathcal{N} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ there is at least one $i \in \{1,\ldots,N\}$ with $||\mathcal{N}-\mathcal{N}_i||_\diamond < \tau$. Existence of τ-nets in $\mathcal{C}(\mathcal{H},\mathcal{K})$ is guaranteed by the compactness of $\mathcal{C}(\mathcal{H},\mathcal{K})$. The next lemma contains a crude upper bound on the cardinality of minimal τ-nets.

Lemma 40. *For any $\tau \in (0,1]$ there is a $\tau-$net $\{\mathcal{N}_i\}_{i=1}^N$ in $\mathcal{C}(\mathcal{H},\mathcal{K})$ with $N \leq (\frac{3}{\tau})^{2(d\cdot d')^2}$, where $d = \dim \mathcal{H}$ and $d' = \dim \mathcal{K}$.*

Proof. The lemma is proved by simply imitating the proof of Lemma 2.6 in [55] where the corresponding result is shown for spheres in arbitrary finite dimensional normed spaces. We give the full argument for convenience.
Let $\{\mathcal{M}_i\}_{i=1}^M$ be an arbitrary subset of $\mathcal{C}(\mathcal{H},\mathcal{K})$ with the property that

$$||\mathcal{M}_i - \mathcal{M}_j||_\diamond \geq \tau, \tag{235}$$

for all $i \neq j$, $i,j \in \{1,\ldots,M\}$. We will establish an upper bound on the integer M now.
The open balls $B_\diamond(\mathcal{M}_i, \frac{\tau}{2})$, $i = 1,\ldots,M$, with centers at $\mathcal{M}_i$ and radii $\tau/2$ are mutually disjoint and are contained in the ball $B_\diamond(0, 1+\frac{\tau}{2})$ since $\mathcal{C}(\mathcal{H},\mathcal{K}) \subset S_\diamond$. So,

$$\bigcup_{i=1}^M B_\diamond(\mathcal{M}_i, \frac{\tau}{2}) \subset B_\diamond(0, 1+\frac{\tau}{2}). \tag{236}$$

Let μ be the Borel-Lebesgue measure (or equivalently the Haar measure) on $(\mathcal{B}(\mathcal{B}(\mathcal{H}),\mathcal{B}(\mathcal{K})), \Sigma_{Borel})$ where $\mathcal{B}(\mathcal{B}(\mathcal{H}),\mathcal{B}(\mathcal{K}))$ denotes the set of linear maps from $\mathcal{B}(\mathcal{H})$ to $\mathcal{B}(\mathcal{K})$ and Σ_{Borel} is the $\sigma-$algebra of Borel sets. Computing the volume of the sets in (236) we obtain

$$M \cdot \left(\frac{\tau}{2}\right)^{2(d\cdot d')^2} \mu(B_\diamond(0,1)) \leq \left(1+\frac{\tau}{2}\right)^{2(d\cdot d')^2} \mu(B_\diamond(0,1)), \tag{237}$$

where $B_\diamond(0,1)$ is the open unit ball with respect to the $\diamond-$norm and $2(d\cdot d')^2$ is the dimension of $\mathcal{B}(\mathcal{B}(\mathcal{H}),\mathcal{B}(\mathcal{K}))$ as a vector space over the field $\mathbb{R}$. This last inequality is equivalent to

$$M \leq \left(1+\frac{2}{\tau}\right)^{2(d\cdot d')^2}. \tag{238}$$

Now, let $\{\mathcal{N}_i\}_{i=1}^N$ be a *maximal* set in $\mathcal{C}(\mathcal{H},\mathcal{K})$ with the property that $||\mathcal{N}_i - \mathcal{N}_j||_\diamond \geq \tau$ for all $i \neq j$. Then, clearly, $\{\mathcal{N}_i\}_{i=1}^N$ is a $\tau-$net and (238) holds. Due to our assumption that $\tau \in (0,1]$ we obtain

$$N \leq \left(1+\frac{2}{\tau}\right)^{2(d\cdot d')^2} \leq \left(\frac{3}{\tau}\right)^{2(d\cdot d')^2} \tag{239}$$

and we are done. □

Let $\mathfrak{I} \subseteq \mathcal{C}(\mathcal{H},\mathcal{K})$ be an arbitrary set. Starting from a $\tau/2-$net $\mathfrak{N} := \{\mathcal{N}_i\}_{i=1}^N$ with $N \leq (\frac{6}{\tau})^{2(d\cdot d')^2}$ as in Lemma 40 we can build a $\tau/2-$net $\mathfrak{I}'_\tau$ that is adapted to the set $\mathfrak{I}$ given by

$$\mathfrak{I}'_\tau := \{\mathcal{N}_i \in \mathfrak{N} : \exists \mathcal{N} \in \mathfrak{I} \text{ with } ||\mathcal{N} - \mathcal{N}_i||_\diamond < \tau/2\}, \tag{240}$$

i.e. we select only those members of the $\tau/2$-net that are contained in the $\tau/2$-neighborhood of $\mathfrak{I}$. Let $\mathcal{T} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ be the useless channel given by $\mathcal{T}(\rho) := \frac{1}{\dim \mathcal{K}} \mathbf{1}_\mathcal{K}$, $\rho \in \mathcal{S}(\mathcal{H})$, and consider

$$\mathfrak{I}_\tau := \left\{(1-\frac{\tau}{2})\mathcal{N} + \frac{\tau}{2}\mathcal{T} : \mathcal{N} \in \mathfrak{I}'_\tau\right\}, \tag{241}$$

where $\mathfrak{I}'_\tau$ is defined in (240). For $\mathfrak{I} \subseteq \mathcal{C}(\mathcal{H},\mathcal{K})$ we set

$$I_c(\rho,\mathfrak{I}) := \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho,\mathcal{N}), \tag{242}$$

for $\rho \in \mathcal{S}(\mathcal{H})$. We list a few more or less obvious results in the following lemma that will be needed in the following.

Lemma 41. *Let $\mathfrak{I} \subseteq \mathcal{C}(\mathcal{H}, \mathcal{K})$. For each positive $\tau \leq \frac{1}{e}$ let $\mathfrak{I}_\tau$ be the finite set of channels defined in (241).*

1. $|\mathfrak{I}_\tau| \leq (\frac{6}{\tau})^{2(d \cdot d')^2}$ *with* $d = \dim \mathcal{H}$ *and* $d' = \dim \mathcal{K}$.

2. *For $\mathcal{N} \in \mathfrak{I}$ there is $\mathcal{N}_i \in \mathfrak{I}_\tau$ with*
$$||\mathcal{N}^{\otimes l} - \mathcal{N}_i^{\otimes l}||_\diamond < l\tau. \tag{243}$$
Consequently, for $\mathcal{N}$, $\mathcal{N}_i$, and any CPTP maps $\mathcal{P} : \mathcal{B}(\mathcal{F}) \to \mathcal{B}(\mathcal{H})^{\otimes l}$ and $\mathcal{R} : \mathcal{B}(\mathcal{K})^{\otimes l} \to \mathcal{B}(\mathcal{F}')$ the relation
$$|F_e(\rho, \mathcal{R} \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}) - F_e(\rho, \mathcal{R} \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P})| < l\tau \tag{244}$$
holds for all $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$ and $l \in \mathbb{N}$.

3. *For all $\rho \in \mathcal{S}(\mathcal{H})$ we have*
$$|I_c(\rho, \mathfrak{I}) - I_c(\rho, \mathfrak{I}_\tau)| \leq \tau + 2\tau \log \frac{d \cdot d'}{\tau}. \tag{245}$$

Proof. 1. This is clear from definition of $\mathfrak{I}_\tau$.
2. It is clear from construction of $\mathfrak{I}_\tau$ that there is at least one $\mathcal{N}_i \in \mathfrak{I}_\tau$ with
$$||\mathcal{N} - \mathcal{N}_i||_\diamond < \tau. \tag{246}$$
We know from [46, 58] that $||\mathcal{R}_1 \otimes \mathcal{R}_2||_\diamond = ||\mathcal{R}_1||_\diamond \cdot ||\mathcal{R}_2||_\diamond$ holds, i.e. $\diamond-$norm is multiplicative. The inequality (243) is easily seen using repeatedly the tensor identity
$$a_1 \otimes b_1 - a_2 \otimes b_2 = a_1 \otimes (b_1 - b_2) + (a_1 - a_2) \otimes b_2, \tag{247}$$
the multiplicativity of the $\diamond-$norm, and the fact that $||\mathcal{R}||_\diamond = 1$ for all CPTP maps.
Let $\psi \in \mathcal{H}^{\otimes l} \otimes \mathcal{H}^{\otimes l}$ be a purification of $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$. Let us denote the left hand side of (244) by $\triangle F_e$. By the definition of the entanglement fidelity we have
$$\triangle F_e = |\langle \psi, id_d^{\otimes l} \otimes (\mathcal{R} \circ (\mathcal{N}^{\otimes l} - \mathcal{N}_i^{\otimes l}) \circ \mathcal{P})(|\psi\rangle\langle\psi|)\psi\rangle|. \tag{248}$$
An application of the Cauchy-Schwarz inequality (writing $\|\cdot\|$ for the euclidean norm) shows that
$$\begin{aligned}
\triangle F_e &\leq \|id_d^{\otimes l} \otimes (\mathcal{R} \circ (\mathcal{N}^{\otimes l} - \mathcal{N}_i^{\otimes l}) \circ \mathcal{P})(|\psi\rangle\langle\psi|)\psi\| && (249)\\
&\leq \|id_d^{\otimes l} \otimes (\mathcal{R} \circ (\mathcal{N}^{\otimes l} - \mathcal{N}_i^{\otimes l}) \circ \mathcal{P})(|\psi\rangle\langle\psi|)\|_\infty && (250)\\
&\leq \|id_d^{\otimes l} \otimes (\mathcal{R} \circ (\mathcal{N}^{\otimes l} - \mathcal{N}_i^{\otimes l}) \circ \mathcal{P})(|\psi\rangle\langle\psi|)\|_1 && (251)\\
&= \|(id_d^{\otimes l} \otimes \mathcal{R}) \circ (id_d^{\otimes l} \otimes (\mathcal{N}^{\otimes l} - \mathcal{N}_i^{\otimes l})) \circ (id_d^{\otimes l} \otimes \mathcal{P}(|\psi\rangle\langle\psi|))\|_1 && (252)\\
&\leq \|\mathcal{R}\|_\diamond \|\mathcal{N}^{\otimes l} - \mathcal{N}_i^{\otimes l}\|_\diamond \|id_d^{\otimes l} \otimes \mathcal{P}(|\psi\rangle\langle\psi|)\|_1 && (253)\\
&< l\tau, && (254)
\end{aligned}$$
where we have used $\|\mathcal{R}\|_\diamond = 1$, $\||\psi\rangle\langle\psi|\|_1 = 1$, and (243).
3. The proof of (245) is based on Fannes inequality [30] and uses merely standard conclusions. So, we will confine ourselves to a brief outline of the argument. Fannes inequality states that $|S(\sigma_1) - S(\sigma_2)| \leq \tau \log d - \tau \log \tau$ for all density operators with $||\sigma_1 - \sigma_2||_1 \leq \tau \leq 1/e$. To the given τ we can always find an $\mathcal{N}' \in \mathfrak{I}$ with
$$I_c(\rho, \mathcal{N}') \leq I_c(\rho, \mathfrak{I}) + \tau. \tag{255}$$
On the other hand there is $\mathcal{N}_i \in \mathfrak{I}_\tau$ with $||\mathcal{N}' - \mathcal{N}_i||_\diamond < \tau$. This implies immediately
$$||\mathcal{N}'(\rho) - \mathcal{N}_i(\rho)||_1 < \tau, \tag{256}$$

and

$$||id_d \otimes \mathcal{N}'(|\psi\rangle\langle\psi|) - id_d \otimes \mathcal{N}_i(|\psi\rangle\langle\psi|)||_1 < \tau \tag{257}$$

by the definition of $\Diamond-$norm where $\psi \in \mathcal{H} \otimes \mathcal{H}$ is a purification of $\rho \in \mathcal{S}(\mathcal{H})$. Since

$$I_c(\rho, \mathcal{N}') = S(\mathcal{N}'(\rho)) - S(id_d \otimes \mathcal{N}'(|\psi\rangle\langle\psi|)) \tag{258}$$

with a similar relation for $\mathcal{N}_i$, an application of Fannes inequality leads to

$$I_c(\rho, \mathcal{N}_i) \leq I_c(\rho, \mathcal{N}') + 2(\tau \log(d \cdot d') - \tau \log \tau). \tag{259}$$

This and (255) show that

$$I_c(\rho, \mathfrak{I}_\tau) \leq I_c(\rho, \mathfrak{I}) + \tau + 2(\tau \log(d \cdot d') - \tau \log \tau). \tag{260}$$

The inequality

$$I_c(\rho, \mathfrak{I}) \leq I_c(\rho, \mathfrak{I}_\tau) + 2(\tau \log(d \cdot d') - \tau \log \tau) \tag{261}$$

is shown in a similar vein. □

3.4.1 The compound BSST-lemma

The following lemma is the compound analog of a result discovered by Bennett, Shor, Smolin, and Thapliyal in [12] (BSST lemma for short). We will use the following abbreviations: For any $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ and $l \in \mathbb{N}$ we set

$$\mathfrak{I}^{\otimes l} := \left\{ \mathcal{N}^{\otimes l} : \mathcal{N} \in \mathfrak{I} \right\}. \tag{262}$$

Recall also our earlier shortcut notation

$$I_c(\rho, \mathfrak{I}) = \inf_{\mathcal{N} \in \mathfrak{I}} I_c(\rho, \mathcal{N}) \tag{263}$$

for $\rho \in \mathcal{S}(\mathcal{H})$.

Lemma 42 (Compound BSST Lemma)**.** *Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be an arbitrary set of channels. For any $\rho \in \mathcal{S}(\mathcal{H})$ let $q_{\delta,l} \in \mathcal{B}(\mathcal{H}^{\otimes l})$ $(l \in \mathbb{N})$ be the frequency-typical projections of ρ and set*

$$\pi_{\delta,l} := \frac{q_{\delta,l}}{tr(q_{\delta,l})} \in \mathcal{S}(\mathcal{H}^{\otimes l}) \qquad (l \in \mathbb{N}). \tag{264}$$

Then there is a positive sequence $(\delta_l)_{l \in \mathbb{N}}$ satisfying $\lim_{l\to\infty} \delta_l = 0$ and a $\rho \in \mathcal{S}(\mathcal{H})$ with

$$\lim_{l\to\infty} \frac{1}{l} \inf_{\mathcal{N} \in \mathfrak{I}} I_c(\pi_{\delta_l,l}, \mathcal{N}^{\otimes l}) = \inf_{\mathcal{N} \in \mathfrak{I}} I_c(\rho, \mathcal{N}). \tag{265}$$

Proof. The proof is via reduction to Holevo's proof [39] of the BSST lemma for single channel supplemented by a discretization argument. We choose a decreasing sequence $(\tau_l)_{l \in \mathbb{N}}$, with $\tau_l > 0$, $\lim_{l\to\infty} l\tau_l = 0$, and consider the finite set of channels $\mathfrak{I}_{\tau_l} = \{\mathcal{N}_1, \ldots, \mathcal{N}_{N_{\tau_l}}\}$ defined in (241) associated to $\mathfrak{I}$.
By our construction of the set $\mathfrak{I}_{\tau_l}$ we know that

$$\mathcal{N}_i(\rho) \geq \frac{\tau_l}{d'2} \mathbf{1}_{\mathcal{K}} \tag{266}$$

holds for all $i \in \{1, \ldots, N_{\tau_l}\}$, which implies

$$\log \mathcal{N}_i(\rho) \geq \log \left(\frac{\tau_l}{d'2} \right) \mathbf{1}_{\mathcal{K}} \tag{267}$$

uniformly in $i \in \{1,\dots,N_{\tau_l}\}$. On the other hand let $\mathfrak{I}_{\tau_l,e} = \{\mathcal{E}_1,\dots\mathcal{E}_{N_{\tau_l}}\} \subset \mathcal{C}(\mathcal{H},\mathcal{H}_e)$ denote the complementary set of channels associated to $\mathfrak{I}_{\tau_l}$. We alter $\mathfrak{I}_{\tau_l,e}$ by mixing a part of useless channel $\mathcal{U}_e \in \mathcal{C}(\mathcal{H},\mathcal{H}_e)$ to each $\mathcal{E}_i \in \mathfrak{I}_{\tau_l,e}$, i.e. we set

$$\mathfrak{I}'_{\tau_l,e} := \left(1-\frac{\tau_l}{2}\right)\mathfrak{I}_{\tau_l,e} + \frac{\tau_l}{2}\mathcal{U}_e. \tag{268}$$

Note that then for each $\mathcal{E}'_i \in \mathfrak{I}'_{\tau_l,e}$

$$\mathcal{E}'_i(\rho) \geq \frac{\tau_l}{2\dim(\mathcal{H}_e)}\mathbf{1}_{\mathcal{H}_e}, \tag{269}$$

and consequently

$$\log \mathcal{E}'_i(\rho) \geq \log\left(\frac{\tau_l}{2\dim(\mathcal{H}_e)}\right)\mathbf{1}_{\mathcal{H}_e}. \tag{270}$$

Applying Holevo's argument from [39] to each channel from $\mathfrak{I}_{\tau_l}$ and $\mathfrak{I}'_{\tau_l,e}$ with our notation from lemma 36 and uniform bounds (267), (270) we obtain for each $i \in \{1,\dots,N_{\tau_l}\}$

$$\begin{aligned}\left|\frac{1}{l}S(\mathcal{T}_i^{\otimes l}(\pi_{\delta,l})) - S(\mathcal{T}_i(\rho))\right| &\leq -\frac{1}{l}\log\eta_l(\delta) + 2\varphi(\delta) \\ &\quad -d\delta\log\left(\frac{\tau_l}{2D}\right) &(271)\\ &=: \Theta_l(\delta,D) &(272)\end{aligned}$$

where $(\mathcal{T}_i,D) \in \{(\mathcal{N}_i,d'),(\mathcal{E}'_i,\dim\mathcal{H}_e)\}$.
Since for each $i \in \{1,\dots,N_{\tau_l}\}$

$$||\mathcal{E}_i - \mathcal{E}'_i||_\diamond \leq \tau_l, \tag{273}$$

we obtain

$$||\mathcal{E}_i^{\otimes l} - \mathcal{E}_i'^{\otimes l}||_\diamond \leq l\tau_l. \tag{274}$$

Hence choosing l sufficiently large we can ensure that $l\tau_l \leq \frac{1}{e}$ and an application of Fannes inequality shows that

$$|S(\mathcal{E}'_i(\rho)) - S(\mathcal{E}_i(\rho))| \leq \tau_l \log\frac{\dim\mathcal{H}_e}{\tau_l}, \tag{275}$$

and

$$\left|\frac{1}{l}S\left(\mathcal{E}_i'^{\otimes l}(\pi_{\delta,l})\right) - \frac{1}{l}S\left(\mathcal{E}_i^{\otimes l}(\pi_{\delta,l})\right)\right| \leq l\tau_l\log\frac{\dim\mathcal{H}_e}{l\tau_l}. \tag{276}$$

Inequalities (271), (275), and (276) show that

$$\begin{aligned}\left|\frac{1}{l}I_c(\pi_{\delta,l},\mathcal{N}_i^{\otimes l}) - I_c(\rho,\mathcal{N}_i)\right| &\leq \Theta_l(\delta,d') + \Theta_l(\delta,\dim\mathcal{H}_e) \\ &\quad +\tau_l\log\frac{\dim\mathcal{H}_e}{\tau_l} \\ &\quad +l\tau_l\log\frac{\dim\mathcal{H}_e}{l\tau_l} \\ &=: \Delta_l(\delta,d',\dim\mathcal{H}_e) &(277)\end{aligned}$$

for each $i \in \{1,\dots,N_{\tau_l}\}$. It is then easily seen utilizing (277) that

$$\left|\frac{1}{l}I_c(\pi_{\delta,l},\mathfrak{I}_{\tau_l}^{\otimes l}) - I_c(\rho,\mathfrak{I}_{\tau_l})\right| \leq \Delta_l(\delta,d',\dim\mathcal{H}_e). \tag{278}$$

Applying (245) to ρ, $\mathfrak{I}$ and $\pi_{\delta,l}$, $\mathfrak{I}^{\otimes l}$ we obtain

$$|I_c(\rho,\mathfrak{I}_{\tau_l})-I_c(\rho,\mathfrak{I})|\leq \tau_l+2\tau_l\log\frac{d}{\tau_l}, \tag{279}$$

and

$$\left|\frac{1}{l}I_c(\pi_{\delta,l},\mathfrak{I}_{\tau_l}^{\otimes l})-\frac{1}{l}I_c(\pi_{\delta,l},\mathfrak{I}^{\otimes l})\right|\leq \tau_l+2l\tau_l\log\frac{d}{l\tau_l}. \tag{280}$$

Using triangle inequality, (278), (279), and (280) we see that

$$\begin{aligned}\left|\frac{1}{l}I_c(\pi_{\delta,l},\mathfrak{I}^{\otimes l})-I_c(\rho,\mathfrak{I})\right| \quad &\leq \quad \Delta_l(\delta,d',\dim\mathcal{H}_e)\\ &\quad +\tau_l+2l\tau_l\log\frac{d}{l\tau_l}\\ &\quad +\tau_l+2\tau_l\log\frac{d}{\tau_l}.\end{aligned} \tag{281}$$

We are done now, since for any positive sequence $(\delta_l)_{l\in\mathbb{N}}$ with $\lim_{l\to\infty}\delta_l=0$, $\lim_{l\to\infty}\eta_l(\delta_l)=1$, and $\lim_{l\to\infty}\delta_l\log\tau_l=0$ we can conclude from (281) that

$$\lim_{l\to\infty}\frac{1}{l}\inf_{\mathcal{N}\in\mathfrak{I}}I_c(\pi_{\delta_l,l},\mathcal{N}^{\otimes l})=\inf_{\mathcal{N}\in\mathfrak{I}}I_c(\rho,\mathcal{N}). \tag{282}$$

holds. □

3.5 Direct parts of the coding theorems for general compound quantum channels

3.5.1 The case of informed decoder and uninformed users

The main step towards the direct part of the coding theorem for quantum compound channels with uninformed users is the following theorem.

Lemma 43. *Let $\mathfrak{I}\in\mathcal{C}(\mathcal{H},\mathcal{K})$ be an arbitrary compound channel and let $\pi_{\mathcal{G}}$ be the maximally mixed state associated with a subspace $\mathcal{G}\subset\mathcal{H}$. Then*

$$Q(\mathfrak{I})\geq\inf_{\mathcal{N}\in\mathfrak{I}}I_c(\pi_{\mathcal{G}},\mathcal{N}). \tag{283}$$

Proof. We consider two subspaces $\mathcal{E}_l,\mathcal{G}^{\otimes l}$ of $\mathcal{H}^{\otimes l}$ with $\mathcal{E}_l\subset\mathcal{G}^{\otimes l}\subset\mathcal{H}^{\otimes l}$. Let $k_l:=\dim\mathcal{E}_l$ and we denote as before the associated maximally mixed states on $\mathcal{E}_l$ and $\mathcal{G}$ by $\pi_{\mathcal{E}_l}$ and $\pi_{\mathcal{G}}$.
If $\inf_{\mathcal{N}\in\mathfrak{I}}I_c(\pi_{\mathcal{G}},\mathcal{N})\leq 0$ there is nothing to prove. Therefore we will suppose in the following that

$$\inf_{\mathcal{N}\in\mathfrak{I}}I_c(\pi_{\mathcal{G}},\mathcal{N})>0 \tag{284}$$

holds. We will show that for each $\varepsilon\in(0,\inf_{\mathcal{N}\in\mathfrak{I}}I_c(\pi_{\mathcal{G}},\mathcal{N}))$ the number

$$\inf_{\mathcal{N}\in\mathfrak{I}}I_c(\pi_{\mathcal{G}},\mathcal{N})-\varepsilon \tag{285}$$

is an achievable rate.
For each $l\in\mathbb{N}$ let us choose some $\tau_l>0$ with $\tau_l\leq\frac{1}{e}$, $\lim_{l\to\infty}l\tau_l=0$, and such that N_{τ_l} grows sub-exponentially with l. E.g. we may choose $\tau_l:=\min\{1/e,1/l^2\}$. We consider, for each $l\in\mathbb{N}$, the finite set of channels $\mathfrak{I}_{\tau_l}:=\{\mathcal{N}_1,\ldots,\mathcal{N}_{N_{\tau_l}}\}$ associated to $\mathfrak{I}$ given in (241) with the properties listed in Lemma

41. We can conclude from the proof of Theorem 38 that for each $l \in \mathbb{N}$ there is a subspace $\mathcal{F}_l \subset \mathcal{G}^{\otimes l}$ of dimension

$$k_l = \lfloor 2^{l(\min_{i\in\{1,\dots,N_\tau\}} I_c(\pi_\mathcal{G},\mathcal{N}_i)-\frac{\varepsilon}{2})} \rfloor, \tag{286}$$

a recovery operation $\mathcal{R}$, and a unitary encoder $\mathcal{W}^l$ such that

$$\min_{i\in\{1,\dots,N_{\tau_l}\}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}\circ\mathcal{N}_i^{\otimes l}\circ\mathcal{W}^l) \quad \geq \quad 1-\sqrt{N_{\tau_l}\epsilon_l} \tag{287}$$

where ϵ_l is defined in (215) (with the approximation parameter ε replaced by $\varepsilon/2$), and we have chosen $l, l_0 \in \mathbb{N}$ with $l \geq l_0$ large enough and $\delta > 0$ small enough to ensure that both

$$\min_{i\in\{1,\dots,N_{\tau_l}\}} I_c(\pi_\mathcal{G},\mathcal{N}_i) - \frac{\varepsilon}{2} > 0, \tag{288}$$

and

$$\frac{\varepsilon}{2} - \gamma(\delta) - \varphi(\delta) - h'(l_0) > \varepsilon/4 > 0. \tag{289}$$

By our construction of $\mathfrak{I}_{\tau_l}$ we can find to each $\mathcal{N} \in \mathfrak{I}$ at least one $\mathcal{N}_i \in \mathfrak{I}_{\tau_l}$ with

$$|F_e(\pi_{\mathcal{F}_l}, \mathcal{R}\circ\mathcal{N}_i^{\otimes l}\circ\mathcal{W}^l) - F_e(\pi_{\mathcal{F}_l}, \mathcal{R}\circ\mathcal{N}^{\otimes l}\circ\mathcal{W}^l)| \leq l\cdot\tau_l \tag{290}$$

according to Lemma 41. Moreover, by the last claim of Lemma 41 we obtain the following estimate on the dimension k_l of the subspace $\mathcal{F}_l$:

$$k_l \geq \lfloor 2^{l(\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_\mathcal{G},\mathcal{N})-\frac{\varepsilon}{2}-\tau_l-2\tau_l\log\frac{d}{\tau_l}} \rfloor. \tag{291}$$

The inequalities (287) and (290) show that

$$\inf_{\mathcal{N}\in\mathfrak{I}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}\circ\mathcal{N}^{\otimes l}\circ\mathcal{W}^l) \geq 1-\sqrt{3N_{\tau_l}\epsilon_l} - l\tau_l, \tag{292}$$

which in turn with (291) shows that $\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_\mathcal{G},\mathcal{N})$ is an achievable rate. □

In order to pass from the maximally mixed state $\pi_\mathcal{G}$ to an arbitrary one we will employ the compound generalization Lemma 42 of the Bennett, Shor, Smolin, and Thapliyal Lemma (BSST Lemma for short). With these preparations it is easy now to finish the proof of the direct part of the coding theorem for the quantum compound channel with uninformed users.
First notice that for each $k \in \mathbb{N}$

$$Q(\mathfrak{I}^{\otimes k}) = kQ(\mathfrak{I}) \tag{293}$$

holds. For any fixed $\rho \in \mathcal{S}(\mathcal{H}^{\otimes m})$ let $q_{\delta,l} \in \mathcal{B}(\mathcal{H}^{\otimes ml})$ be the frequency-typical projection of ρ and set $\pi_{\delta,l} = \frac{q_{\delta,l}}{\mathrm{tr}(q_{\delta,l})}$. Lemma 43 implies that for any $\delta \in (0, 1/2)$ we have

$$Q(\mathfrak{I}^{\otimes ml}) \geq I_c(\pi_{\delta,l}, \mathfrak{I}^{\otimes ml}), \tag{294}$$

for all $m, l \in \mathbb{N}$. Utilizing (293), (294) and Lemma 42 we arrive at

$$\begin{aligned} Q(\mathfrak{I}) &= \frac{1}{m}\lim_{l\to\infty}\frac{1}{l}Q(\mathfrak{I}^{\otimes ml}) && (295)\\ &\geq \frac{1}{m}\lim_{l\to\infty}\frac{1}{l}\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\delta_l,l}, (\mathcal{N}^{\otimes m})^{\otimes l}) && (296)\\ &= \frac{1}{m} I_c(\rho, \mathfrak{I}^{\otimes m}). && (297) \end{aligned}$$

From (295) and since $Q_{ID}(\mathfrak{I}) \geq Q(\mathfrak{I})$ trivially holds we get without further ado the direct part of the coding theorem.

Theorem 44 (Direct Part: Informed Decoder and Uninformed Users)**.** *Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H},\mathcal{K})$ be an arbitrary set. Then*

$$Q_{ID}(\mathfrak{I}) \geq Q(\mathfrak{I}) \geq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho,\mathcal{N}^{\otimes l}). \tag{298}$$

Remark 45. *It is quite easy to see that the limit in (298) exists. Indeed it holds that*

$$\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l+k})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho,\mathcal{N}^{\otimes l+k}) \quad \geq \quad \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho,\mathcal{N}^{\otimes l}) + \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes k})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho,\mathcal{N}^{\otimes k}) \tag{299}$$

which implies the existence of the limit via standard arguments.

3.5.2 The informed encoder

The main result of this subsection will rely on an appropriate variant of the BSST Lemma. To this end we first recall Holevo's version of that result. For $\delta > 0$, $l \in \mathbb{N}$, and $\rho \in \mathcal{S}(\mathcal{H})$ let $q_{\delta,l} \in \mathcal{B}(\mathcal{H}^{\otimes l})$ denote the frequency typical projection of $\rho^{\otimes l}$. Set

$$\pi_{\delta,l} = \pi_{\delta,l}(\rho) := \frac{q_{\delta,l}}{\mathrm{tr}(q_{\delta,l})}. \tag{300}$$

Moreover, let

$$\lambda_{\min}(\rho) := \min\{\lambda \in \sigma(\rho) : \lambda > 0\}, \tag{301}$$

where $\sigma(\rho)$ stands for the spectrum of the density operator ρ.

Lemma 46 (BSST Lemma [12], [39])**.** *For any $\delta \in (0, \frac{1}{2\dim\mathcal{H}})$, any $\mathcal{N} \in \mathcal{C}(\mathcal{H},\mathcal{K})$, and every $\rho \in \mathcal{S}(\mathcal{H})$ with associated state $\pi_{\delta,l} = \pi_{\delta,l}(\rho) \in \mathcal{S}(\mathcal{H}^{\otimes l})$ we have*

$$\left|\frac{1}{l} S(\mathcal{N}^{\otimes l}(\pi_{\delta,l})) - S(\mathcal{N}(\rho))\right| \leq \theta_l(\delta, \lambda_{\min}(\rho), \lambda_{\min}(\mathcal{N}(\rho))) \tag{302}$$

where

$$\begin{aligned} \theta_l(\delta, \lambda_{\min}(\rho), \lambda_{\min}(\mathcal{N}(\rho))) \quad &= \quad \frac{\dim\mathcal{H}}{l}\log(l+1) - \dim\mathcal{H}\cdot\delta\log\delta \\ &\quad - \dim\mathcal{H}\cdot\delta\cdot(\log\lambda_{\min}(\rho) + \log\lambda_{\min}(\mathcal{N}(\rho))). \end{aligned} \tag{303}$$

Before we present our extended version of BSST Lemma we introduce some notation. For $t \in (0, \frac{1}{e})$ and any set $\mathfrak{I} \subset \mathcal{C}(\mathcal{H},\mathcal{K})$ let us define

$$\mathfrak{I}^{(t)} := \{\mathcal{N}^{(t)} = (1-t)\mathcal{N} + t\mathcal{T}_{\mathcal{K}} : \mathcal{N} \in \mathfrak{I}\} = (1-t)\mathfrak{I} + t\mathcal{T}_{\mathcal{K}}, \tag{304}$$

where $\mathcal{T} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ is given by $\mathcal{T}_{\mathcal{K}}(x) := \frac{\mathrm{tr}(x)}{\dim\mathcal{K}}\mathbf{1}_{\mathcal{K}}$.
On the other hand, to each $\mathcal{N} \in \mathfrak{I} \subset \mathcal{C}(\mathcal{H},\mathcal{K})$ we can associate a complementary channel $\mathcal{N}_c \in \mathcal{C}(\mathcal{H},\mathcal{H}_e)$ where we assume w.l.o.g. that $\mathcal{H}_e = \mathbb{C}^{\dim\mathcal{H}\cdot\dim\mathcal{K}}$. Let $\mathfrak{I}' \subset \mathcal{C}(\mathcal{H},\mathcal{H}_e)$ denote the set of channels complementary to $\mathfrak{I}$ and set

$$\mathfrak{I}'^{(t)} := (\mathfrak{I}')^{(t)} = \{\mathcal{N}_c^{(t)} = (1-t)\mathcal{N}_c + t\mathcal{T}_{\mathcal{H}_e} : \mathcal{N}_c \in \mathfrak{I}'\} = (1-t)\mathfrak{I}' + t\mathcal{T}_{\mathcal{H}_e}, \tag{305}$$

where $\mathcal{T}_{\mathcal{H}_e} \in \mathcal{C}(\mathcal{H},\mathcal{H}_e)$ is the defined in a similar way as $\mathcal{T}_{\mathcal{K}}$. Finally, for $\mathcal{N} \in \mathfrak{I}$ let

$$\rho_{\mathcal{N}} := \arg\max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho,\mathcal{N}), \tag{306}$$

and for $t \in (0, \frac{1}{e})$, $\delta > 0$, and $l \in \mathbb{N}$ define

$$\pi^{(t)}_{\delta,l,\mathcal{N}} := \pi_{\delta,l}\left(\rho^{(t)}_{\mathcal{N}}\right), \tag{307}$$

where we have used the notation from (300) and

$$\rho^{(t)}_{\mathcal{N}} := (1-t)\rho_{\mathcal{N}} + \frac{t}{\dim \mathcal{H}}\mathbf{1}_{\mathcal{H}}. \tag{308}$$

Lemma 47 (Uniform BSST-Lemma)**.** *1. Let $l \in \mathbb{N}$, $t \in (0, \frac{1}{l\cdot e})$, and $\delta \in (0, \frac{1}{2\dim \mathcal{H}})$. Then with the notation introduced in the preceding paragraph we have*

$$\left|\frac{1}{l}\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi^{(t)}_{\delta,l,\mathcal{N}}, \mathcal{N}^{\otimes l}) - \inf_{\mathcal{N}\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho,\mathcal{N})\right| \leq \Delta_l(\delta,t), \tag{309}$$

with

$$\begin{aligned}\Delta_l(\delta,t) &= 2\theta_l\left(\delta, \frac{t}{\dim\mathcal{H}}, \frac{t}{\dim\mathcal{K}}\right) + 2\theta_l\left(\delta, \frac{t}{\dim\mathcal{H}}, \frac{t}{\dim\mathcal{H}_e}\right) &(310)\\ &\quad -4t\log\frac{t}{\dim\mathcal{K}\cdot\dim\mathcal{H}_e} - 2lt\log\frac{lt}{\dim\mathcal{K}\cdot\mathcal{H}_e}, &(311)\\ & &(312)\end{aligned}$$

where $\theta_l\left(\delta, \frac{t}{\dim\mathcal{H}}, \frac{t}{\dim\mathcal{K}}\right)$ and $\theta_l\left(\delta, \frac{t}{\dim\mathcal{H}}, \frac{t}{\dim\mathcal{H}_e}\right)$ are from Lemma 46.
2. Consequently, choosing suitable positive sequences $(\delta_l)_{l\in\mathbb{N}}$, $(t_l)_{l\in\mathbb{N}}$ with

1. *$\lim_{l\to\infty}\delta_l = 0 = \lim_{l\to\infty} lt_l$, and*
2. *$\lim_{l\to\infty}\delta_l \log t_l = 0$*

we see that for $\nu_l := \Delta_l(\delta_l, t_l)$

$$\left|\frac{1}{l}\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi^{(t_l)}_{\delta_l,l,\mathcal{N}}, \mathcal{N}^{\otimes l}) - \inf_{\mathcal{N}\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho,\mathcal{N})\right| \leq \nu_l \tag{313}$$

holds with $\lim_{l\to\infty}\nu_l = 0$.

Proof. Our proof strategy is to reduce the claim to the BSST Lemma 46. Let $t > 0$ be small enough to ensure that $l\cdot t \in (0, \frac{1}{e})$ and let $\delta \in (0, \frac{1}{2\dim\mathcal{H}})$ be given. From (304) and (305) we obtain that

$$\lambda_{\min}(\mathcal{N}^{(t)}(\rho)) \geq \frac{t}{\dim\mathcal{K}}, \qquad \lambda_{\min}(\mathcal{N}^{(t)}_c(\rho)) \geq \frac{t}{\dim\mathcal{H}_e} \qquad \forall\ \rho\in\mathcal{S}(\mathcal{H}), \tag{314}$$

and (308) yields that

$$\lambda_{\min}(\rho^{(t)}_{\mathcal{N}}) \geq \frac{t}{\dim\mathcal{H}} \tag{315}$$

for all $\mathcal{N}\in\mathfrak{I}$. The bounds (314) and (315) along with Lemma 46 show that

$$\left|\frac{1}{l}S((\mathcal{N}^{(t)})^{\otimes l}(\pi^{(t)}_{\delta,l,\mathcal{N}})) - S(\mathcal{N}^{(t)}(\rho^{(t)}_{\mathcal{N}}))\right| \leq \theta_l\left(\delta, \frac{t}{\dim\mathcal{H}}, \frac{t}{\dim\mathcal{K}}\right), \tag{316}$$

and

$$\left|\frac{1}{l}S((\mathcal{N}^{(t)}_c)^{\otimes l}(\pi^{(t)}_{\delta,l,\mathcal{N}})) - S(\mathcal{N}^{(t)}_c(\rho^{(t)}_{\mathcal{N}}))\right| \leq \theta_l\left(\delta, \frac{t}{\dim\mathcal{H}}, \frac{t}{\dim\mathcal{H}_e}\right). \tag{317}$$

On the other hand, by definition we have

$$||\mathcal{N}^{(t)} - \mathcal{N}||_\diamond \leq t, \qquad ||(\mathcal{N}^{(t)})^{\otimes l} - \mathcal{N}^{\otimes l}||_\diamond \leq l \cdot t, \tag{318}$$

and similarly

$$||\mathcal{N}_c^{(t)} - \mathcal{N}_c||_\diamond \leq t, \qquad ||(\mathcal{N}_c^{(t)})^{\otimes l} - \mathcal{N}_c^{\otimes l}||_\diamond \leq l \cdot t, \tag{319}$$

for all $\mathcal{N} \in \mathfrak{I}$. Since $l \cdot t \in (0, \frac{1}{e})$ we obtain from this by Fannes inequality

$$|S(\mathcal{N}^{(t)}(\rho_{\mathcal{N}}^{(t)})) - S(\mathcal{N}(\rho_{\mathcal{N}}^{(t)}))| \leq -t \log \frac{t}{\dim \mathcal{K}}, \tag{320}$$

$$|S(\mathcal{N}_c^{(t)}(\rho_{\mathcal{N}}^{(t)})) - S(\mathcal{N}_c(\rho_{\mathcal{N}}^{(t)}))| \leq -t \log \frac{t}{\dim \mathcal{H}_e} \tag{321}$$

and

$$\left|\frac{1}{l} S((\mathcal{N}^{(t)})^{\otimes l}(\pi_{\delta,l,\mathcal{N}}^{(t)})) - \frac{1}{l} S(\mathcal{N}^{\otimes l}(\pi_{\delta,l,\mathcal{N}}^{(t)}))\right| \leq -l \cdot t \log \frac{l \cdot t}{\dim \mathcal{K}}, \tag{322}$$

as well as

$$\left|\frac{1}{l} S((\mathcal{N}_c^{(t)})^{\otimes l}(\pi_{\delta,l,\mathcal{N}}^{(t)})) - \frac{1}{l} S(\mathcal{N}_c^{\otimes l}(\pi_{\delta,l,\mathcal{N}}^{(t)}))\right| \leq -l \cdot t \log \frac{l \cdot t}{\dim \mathcal{H}_e}, \tag{323}$$

for all $\mathcal{N} \in \mathfrak{I}$. Since

$$I_c(\rho_{\mathcal{N}}^{(t)}, \mathcal{N}) = S(\mathcal{N}(\rho_{\mathcal{N}}^{(t)})) - S(\mathcal{N}_c(\rho_{\mathcal{N}}^{(t)})) \tag{324}$$

and

$$I_c(\pi_{\delta,l,\mathcal{N}}^{(t)}, \mathcal{N}^{\otimes l}) = S(\mathcal{N}^{\otimes l}(\pi_{\delta,l,\mathcal{N}}^{(t)})) - S(\mathcal{N}_c^{\otimes l}(\pi_{\delta,l,\mathcal{N}}^{(t)})), \tag{325}$$

the inequalities (316),(317), (320), (321), (322), (323) and triangle inequality show that *uniformly* in $\mathcal{N} \in \mathfrak{I}$ we have

$$\begin{aligned}\left|\frac{1}{l} I_c(\pi_{\delta,l,\mathcal{N}}^{(t)}, \mathcal{N}^{\otimes l}) - I_c(\rho_{\mathcal{N}}^{(t)}, \mathcal{N})\right| \quad &\leq \quad \theta_l\left(\delta, \frac{t}{\dim \mathcal{H}}, \frac{t}{\dim \mathcal{K}}\right) + \theta_l\left(\delta, \frac{t}{\dim \mathcal{H}}, \frac{t}{\dim \mathcal{H}_e}\right) \\ &\quad -t \log \frac{t}{\dim \mathcal{K} \cdot \dim \mathcal{H}_e} - l \cdot t \log \frac{l \cdot t}{\dim \mathcal{K} \cdot \mathcal{H}_e}. \end{aligned} \tag{326}$$

Now, by (308) we have

$$||\rho_{\mathcal{N}}^{(t)} - \rho_{\mathcal{N}}||_1 \leq t \tag{327}$$

which implies

$$||\mathcal{N}(\rho_{\mathcal{N}}^{(t)}) - \mathcal{N}(\rho_{\mathcal{N}})||_1 \leq t, \qquad ||\mathcal{N}_c(\rho_{\mathcal{N}}^{(t)}) - \mathcal{N}_c(\rho_{\mathcal{N}})||_1 \leq t, \tag{328}$$

since the trace distance of two states can only decrease after applying a trace preserving completely positive map to both states. Thus Fannes inequality leads us to the conclusion that

$$\left|I_c(\rho_{\mathcal{N}}^{(t)}, \mathcal{N}) - I_c(\rho_{\mathcal{N}}, \mathcal{N})\right| \leq -t \log \frac{t}{\dim \mathcal{K} \cdot \dim \mathcal{H}_e}. \tag{329}$$

This and (326) shows that uniformly in $\mathcal{N} \in \mathfrak{I}$

$$\begin{aligned}\left|\frac{1}{l} I_c(\pi_{\delta,l,\mathcal{N}}^{(t)}, \mathcal{N}^{\otimes l}) - I_c(\rho_{\mathcal{N}}, \mathcal{N})\right| \quad &\leq \quad \theta_l\left(\delta, \frac{t}{\dim \mathcal{H}}, \frac{t}{\dim \mathcal{K}}\right) + \theta_l\left(\delta, \frac{t}{\dim \mathcal{H}}, \frac{t}{\dim \mathcal{H}_e}\right) \\ &\quad -2t \log \frac{t}{\dim \mathcal{K} \cdot \dim \mathcal{H}_e} - l \cdot t \log \frac{l \cdot t}{\dim \mathcal{K} \cdot \mathcal{H}_e} \qquad (330) \\ &=: \quad \frac{\Delta_l(\delta, t)}{2}. \qquad (331)\end{aligned}$$

Finally, it is clear from the uniform estimate in (330) that

$$\left|\frac{1}{l}\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi^{(t)}_{\delta,l,\mathcal{N}},\mathcal{N}^{\otimes l}) - \inf_{\mathcal{N}\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho,\mathcal{N})\right| = \left|\frac{1}{l}\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi^{(t)}_{\delta,l,\mathcal{N}},\mathcal{N}^{\otimes l}) - \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho_{\mathcal{N}},\mathcal{N})\right| \quad (332)$$

$$\leq \Delta_l(\delta,t), \quad (333)$$

which concludes the proof. □

Lemma 47 and Theorem 39 easily imply the following result.

Lemma 48. *Let $\mathfrak{I}\subset\mathcal{C}(\mathcal{H},\mathcal{K})$ be an arbitrary set of quantum channels. Then*

$$Q_{IE}(\mathfrak{I}) \geq \inf_{\mathcal{N}\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho,\mathcal{N}). \quad (334)$$

Proof. Take any set $\{\pi_{\mathcal{G}_\mathcal{N}}\}_{\mathcal{N}\in\mathfrak{I}}$ of maximally mixed states on subspaces $\mathcal{G}_\mathcal{N}\subset\mathcal{H}$. In a first step we will show that this yields

$$Q_{IE}(\mathfrak{I}) \geq \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}_\mathcal{N}},\mathcal{N}). \quad (335)$$

Notice that we can assume w.l.o.g. that $\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}_\mathcal{N}},\mathcal{N})>0$.
Denote, for every $\tau>0$, by $\mathfrak{I}_\tau$ a τ-net for $\mathfrak{I}$ as given in (241) of cardinality $N_\tau := |\mathfrak{I}_\tau| \leq (\frac{6}{\tau})^{2(d\cdot d')^2}$, where d,d' are the dimensions of $\mathcal{H},\mathcal{K}$. Starting from this set $\mathfrak{I}_\tau$ it is easy to construct a finite set $\mathfrak{I}^\circ_\tau$ with the following properties:

1. $\mathfrak{I}^\circ_\tau\subset\mathfrak{I}$,
2. $|\mathfrak{I}^\circ_\tau| \leq (\frac{6}{\tau})^{2(d\cdot d')^2}$, and
3. to each $\mathcal{N}\in\mathfrak{I}$ there is at least one $\mathcal{N}'\in\mathfrak{I}^\circ_\tau$ with $||\mathcal{N}-\mathcal{N}'||_\diamond \leq 2\tau$.

Let $(\tau_l)_{l\in\mathbb{N}}$ be defined by $\tau_l := \frac{1}{l^2}$ and consider the sets $\mathfrak{I}^\circ_{\tau_l}$, $l\in\mathbb{N}$.
Take any $\eta\in(0,\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}_\mathcal{N}},\mathcal{N}))$ and set

$$R(\eta) := \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}_\mathcal{N}},\mathcal{N}) - \eta, \quad (336)$$

and

$$R_l(\eta) := \min_{\mathcal{N}\in\mathfrak{I}^\circ_{\tau_l}} I_c(\pi_{\mathcal{G}_\mathcal{N}},\mathcal{N}) - \eta. \quad (337)$$

Then for every $l\in\mathbb{N}$,

$$R_l(\eta) \geq R(\eta) \quad (338)$$

since $\mathfrak{I}^\circ_{\tau_l}\subset\mathfrak{I}$.
Fix some $\delta'\in(0,1/2)$ such that $\gamma(\delta')+\varphi(\delta')<\eta/4$. For every $l\in\mathbb{N}$, choose a subspace $\mathcal{E}_l\subset\mathcal{H}^{\otimes l}$ of dimension

$$k_l(\eta) := \dim\mathcal{E}_l = \lfloor 2^{lR_l(\eta)}\rfloor. \quad (339)$$

The proof of Theorem 39 then shows the existence of a recovery operation $\mathcal{R}^l$ and for each $\mathcal{N}'\in\mathfrak{I}^\circ_{\tau_l}$ a unitary encoder $\mathcal{W}^l_{\mathcal{N}'}$ such that for each $l\in\mathbb{N}$

$$F_e(\pi_{\mathcal{E}_l},\mathcal{R}^l\circ\mathcal{N}'^{\otimes l}\circ\mathcal{W}^l_{\mathcal{N}'}) \geq 1-3\sqrt{N_{\tau_l}\cdot\varepsilon_l} \qquad \forall\,\mathcal{N}'\in\mathfrak{I}^\circ_{\tau_l}, \quad (340)$$

where $\varepsilon_l := 2^{-l(c\delta'^2-h(l))}+2^{-l(c'\delta'^2-h'(l))}+2N_{\tau_l}\sqrt{2^{l(-\frac{3\eta}{4}+h'(l))}})$. From Lemma 41 along with the properties of $\mathfrak{I}^\circ_{\tau_l}$ and our specific choice of $(\tau_l)_{l\in\mathbb{N}}$ it follows that there exist unitary encodings $\mathcal{W}^l_\mathcal{N}$ (for every $l\in\mathbb{N}$ and each $\mathcal{N}\in\mathfrak{I}$), such that

$$F_e(\pi_{\mathcal{E}_l},\mathcal{R}^l\circ\mathcal{N}^{\otimes l}\circ\mathcal{W}^l_\mathcal{N}) \geq 1-3\sqrt{N_{\tau_l}\cdot\varepsilon_l}-\frac{2}{l}\ \forall l\in\mathbb{N},\ \mathcal{N}\in\mathfrak{I}. \quad (341)$$

Clearly, $\lim_{l\to\infty} F_e(\pi_{\mathcal{E}_l}, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{W}^l_{\mathcal{N}}) = 1$ and (338) implies for each $\eta \in (0, \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}_\mathcal{N}}, \mathcal{N}))$ that

$$\liminf_{l\to\infty} \frac{1}{l} \log k_l(\eta) = \liminf_{l\to\infty} \frac{1}{l} \log \dim \mathcal{E}_l \geq R(\eta). \tag{342}$$

Consequently $\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}_\mathcal{N}}, \mathcal{N})$ is achievable.
We proceed by repeated application of the inequality

$$Q_{IE}(\mathfrak{I}) \geq \frac{1}{l} Q_{IE}(\mathfrak{I}^{\otimes l}) \ (\forall l \in \mathbb{N}). \tag{343}$$

From (335) and (343) we get that for each $l \in \mathbb{N}$ and every set $\{\pi^l_\mathcal{N}\}_{\mathcal{N}\in\mathfrak{I}}$ of maximally mixed states on subspaces of $\mathcal{H}^{\otimes l}$,

$$Q_{IE}(\mathfrak{I}) \geq \frac{1}{l} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi^l_\mathcal{N}, \mathcal{N}^{\otimes l}). \tag{344}$$

We now make a specific choice of the states $\pi^l_\mathcal{N}$, namely, for every $\mathcal{N} \in \mathfrak{I}$ and $l \in \mathbb{N}$, set $\pi^l_\mathcal{N} := \pi^{(t_l)}_{\delta_l,l,\mathcal{N}}$ with $\pi^{(t_l)}_{\delta_l,l,\mathcal{N}}$ taken from the second part of Lemma 47. By an application of the second part of Lemma 47 it follows

$$\begin{aligned}
Q_{IE}(\mathfrak{I}) &\geq \lim_{l\to\infty} \frac{1}{l} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi^l_\mathcal{N}, \mathcal{N}^{\otimes l}) && (345)\\
&\geq \lim_{l\to\infty} (\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho_\mathcal{N}, \mathcal{N}) - \nu_l) && (346)\\
&= \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho_\mathcal{N}, \mathcal{N}) && (347)\\
&= \inf_{\mathcal{N}\in\mathfrak{I}} \max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{N}). && (348)
\end{aligned}$$

□

Employing inequality (343) one more time we obtain from Lemma 48 applied to $\mathfrak{I}^{\otimes l}$

$$\begin{aligned}
Q_{IE}(\mathfrak{I}) &\geq \frac{1}{l} Q_{IE}(\mathfrak{I}^{\otimes l}) && (349)\\
&\geq \frac{1}{l} \inf_{\mathcal{N}\in\mathfrak{I}} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}). && (350)
\end{aligned}$$

Consequently we obtain the desired achievability result.

Theorem 49 (Direct Part: Informed Encoder)**.** *For any $\mathfrak{I} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ we have*

$$Q_{IE}(\mathfrak{I}) \geq \lim_{l\to\infty} \frac{1}{l} \inf_{\mathcal{N}\in\mathfrak{I}} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{351}$$

Remark 50. *Note that the limit in (351) exists. Indeed, set*

$$C_l(\mathcal{N}) := \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{352}$$

Then it is clear that

$$C_{l+k}(\mathcal{N}) \geq C_l(\mathcal{N}) + C_k(\mathcal{N}) \tag{353}$$

and consequently

$$\begin{aligned}
\inf_{\mathcal{N}\in\mathfrak{I}} C_{l+k}(\mathcal{N}) &\geq \inf_{\mathcal{N}\in\mathfrak{I}} (C_l(\mathcal{N}) + C_k(\mathcal{N})) && (354)\\
&\geq \inf_{\mathcal{N}\in\mathfrak{I}} C_l(\mathcal{N}) + \inf_{\mathcal{N}\in\mathfrak{I}} C_k(\mathcal{N}), && (355)
\end{aligned}$$

which implies the existence of the limit in (351).

3.6 Converse parts of the coding theorems for general quantum compound channels

In this section we prove the converse parts of the coding theorems for general quantum compound channels in the three different settings concerned with entanglement transmission that are treated in this paper. The proofs deviate from the usual approach due to our more general definitions of codes.

3.6.1 Converse for informed decoder and uninformed users

We first prove the converse part in the case of a finite compound channel, then use a recent result [51] that gives a more convenient estimate for the difference in coherent information of two nearby channels in order to pass on to the general case.
For the converse part in the case of a finite compound channel we need the following lemma that is due to Alicki and Fannes [8]:

Lemma 51 (Cf. [8]). *For two states $\sigma, \rho \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ with trace distance $f = \|\sigma - \rho\|_1$,*

$$|\Delta S(\rho) - \Delta S(\sigma)| \leq 4f \log \dim \mathcal{H}_1 + 2\eta(f) + 2\eta(1-f) \tag{356}$$

where

$$\Delta S(\,\cdot\,) := S(\mathrm{tr}_{\mathcal{H}_1}[\,\cdot\,]) - S(\,\cdot\,) \tag{357}$$

and $\eta : [0,1] \to \mathbb{R}$ is given by the formula $\eta(x) := -x \log x$.

We shall now embark on the proof of the following theorem.

Theorem 52 (Converse Part: Informed Decoder, Uninformed Users, $|\mathfrak{I}| < \infty$). *Let $\mathfrak{I} = \{\mathcal{N}_1, \ldots, \mathcal{N}_N\} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a finite compound channel. The capacities $Q_{ID}(\mathfrak{I})$ and $Q(\mathfrak{I})$ of $\mathfrak{I}$ with informed decoder and uninformed users are bounded from above by*

$$Q(\mathfrak{I}) \leq Q_{ID}(\mathfrak{I}) \leq \lim_{l\to\infty} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})} \min_{\mathcal{N}_i \in \mathfrak{I}} \frac{1}{l} I_c(\rho, \mathcal{N}_i^{\otimes l}). \tag{358}$$

Proof. The inequality $Q(\mathfrak{I}) \leq Q(\mathfrak{I})_{ID}$ is obvious from the definition of codes. We give a proof for the second inequality. Let for arbitrary $l \in \mathbb{N}$ an (l, k_l) code for a compound channel $\mathfrak{I} = \{\mathcal{N}_1, \ldots, \mathcal{N}_N\}$ with informed decoder and the property $\min_{1\leq i\leq N} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_i^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) \geq 1 - \epsilon_l$ be given, where $\epsilon_l \in [0,1]$. Let $|\psi_l\rangle\langle\psi_l| \in \mathcal{S}(\mathcal{E}_l \otimes \mathcal{F}_l)$ be a purification of $\pi_{\mathcal{F}_l}$ where $\mathcal{E}_l$ is just a copy of $\mathcal{F}_l$. We use the abbreviation $\mathcal{D}^l := \frac{1}{N}\sum_{i=1}^N \mathcal{R}_i^l \circ \mathcal{N}_i^{\otimes l}$. Obviously, the above code then satisfies

$$\begin{aligned} \langle \psi_l, id_{\mathcal{E}_l} \otimes \mathcal{D}^l(id_{\mathcal{E}_l} \otimes \mathcal{P}^l(|\psi_l\rangle\langle\psi_l|))\psi_l\rangle &= \frac{1}{N}\sum_{i=1}^N F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_i^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) && (359)\\ &\geq 1 - \epsilon_l. && (360) \end{aligned}$$

Let $\sigma_{\mathcal{P}^l} := id_{\mathcal{E}_l} \otimes \mathcal{P}^l(|\psi^l\rangle\langle\psi^l|)$ and consider any convex decomposition $\sigma_{\mathcal{P}^l} = \sum_{i=1}^{(\dim \mathcal{F}_l)^2} \lambda_i |e_i\rangle\langle e_i|$ of $\sigma_{\mathcal{P}^l}$ into pure states $|e_i\rangle\langle e_i| \in \mathcal{S}(\mathcal{F}_l \otimes \mathcal{H}^{\otimes l})$. By (360) there is at least one $i \in \{1, \ldots, (\dim \mathcal{F}_l)^2\}$ such that

$$\langle \psi_l, id_{\mathcal{E}_l} \otimes \mathcal{D}^l(|e_i\rangle\langle e_i|)\psi_l\rangle \geq 1 - \epsilon_l \tag{361}$$

holds. Without loss of generality, $i = 1$. Turning back to the individual channels, we get

$$\langle \psi_l, id_{\mathcal{E}_l} \otimes \mathcal{R}_i^l \circ \mathcal{N}_i^{\otimes l}(|e_1\rangle\langle e_1|)\psi_l\rangle \geq 1 - N\epsilon_l \quad \forall i \in \{1, \ldots, N\}. \tag{362}$$

We define the state $\rho^l := \mathrm{tr}_{\mathcal{E}_l}(|e_1\rangle\langle e_1|) \in \mathcal{S}(\mathcal{H}^{\otimes l})$ and note that $|e_1\rangle\langle e_1|$ is a purification of ρ^l. Application of recovery operation and individual channels to ρ^l now defines the states $\sigma_k^l := id_{\mathcal{E}_l} \otimes \mathcal{R}_k^l \circ \mathcal{N}_k^{\otimes l}(|e_1\rangle\langle e_1|)$ ($k \in \{1, \ldots, N\}$) which have independently of k the property

$$F(\psi^l, \sigma_k^l) = \langle \psi_l, id_{\mathcal{E}_l} \otimes \mathcal{R}_k^l \circ \mathcal{N}_k^{\otimes l}(|e_i\rangle\langle e_i|)\psi_l\rangle \geq 1 - N\epsilon_l, \tag{363}$$

which by application of the well-known estimate $\|a - b\|_1 \leq \sqrt{1 - F(a,b)}$ for $a, b \in \mathcal{S}(\mathcal{E}_l \otimes \mathcal{F}_l')$ (see [32]), yields

$$\|\psi^l - \sigma_k^l\|_1 \leq \sqrt{N\varepsilon_l}. \tag{364}$$

Now, set $c_3 := \max\{-x\log(x) : x \in [0,1]\}$. This puts us into position for an application of Lemma 51, which together with the data processing inequality for coherent information [60] for small enough ε_l (or, alternatively, large enough l) establishes the following chain of inequalities for every $k \in \{1, \ldots, N\}$:

$$\begin{aligned}
\log\dim\mathcal{F}_l &= S(\pi_{\mathcal{F}_l}) && (365)\\
&= \Delta S(|\psi^l\rangle\langle\psi^l|) && (366)\\
&\leq \Delta S(\sigma_k^l) + 4c_3 + 4\log(\dim\mathcal{F}_l^2)\sqrt{N\epsilon_l} && (367)\\
&= S(\mathrm{tr}_{\mathcal{E}_l}(id_{\mathcal{E}_l} \otimes \mathcal{R}_k^l \circ \mathcal{N}^{\otimes l}(|e_1\rangle\langle e_1|)) - S(id_{\mathcal{E}_l} \otimes \mathcal{R}_k^l \circ \mathcal{N}_k^{\otimes l}(|e_1\rangle\langle e_1|)) \\
&\quad +4c_3 + 8\log(\dim\mathcal{F}_l)\sqrt{N\epsilon_l} && (368)\\
&= I_c(\rho^l, \mathcal{R}_k^l \circ \mathcal{N}_k^{\otimes l}) + 4c_3 + 8\log(\dim\mathcal{F}_l)\sqrt{N\epsilon_l} && (369)\\
&\leq I_c(\rho^l, \mathcal{N}_k^{\otimes l}) + 4c_3 + 8\log(\dim\mathcal{F}_l)\sqrt{N\epsilon_l}. && (370)
\end{aligned}$$

Thus,

$$\begin{aligned}
\log\dim\mathcal{F}_l &\leq \min_{k\in\{1,\ldots,N\}} I_c(\rho^l, \mathcal{N}_k^{\otimes l}) + 4c_3 + 8\log(\dim\mathcal{F}_l)\sqrt{N\epsilon_l} && (371)\\
&\leq \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{1\leq k\leq N} I_c(\rho, \mathcal{N}_k^{\otimes l}) + 4c_3 + 8\log(\dim\mathcal{F}_l)\sqrt{N\epsilon_l}. && (372)
\end{aligned}$$

Let a sequence of (l, k_l) codes for $\mathfrak{I}$ with informed decoder be given such that $\liminf_{l\to\infty} \frac{1}{l}\log\dim\mathcal{F}_l = R \in \mathbb{R}$ and $\lim_{l\to\infty}\epsilon_l = 0$. Then by (372) we get

$$\begin{aligned}
R &= \liminf_{l\to\infty} \frac{1}{l}\log\dim\mathcal{F}_l && (373)\\
&\leq \liminf_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{k\in\{1,\ldots,N\}} I_c(\rho, \mathcal{N}_k^{\otimes l}) + \liminf_{l\to\infty}\frac{1}{l}4c_3 + \liminf_{l\to\infty}\frac{1}{l}16\log(\dim\mathcal{F}_l)\sqrt{2N\epsilon_l} && (374)\\
&= \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{k\in\{1,\ldots,N\}} I_c(\rho, \mathcal{N}_k^{\otimes l}), && (375)
\end{aligned}$$

□

Let us now focus on the general case. We shall prove the following theorem:

Theorem 53 (Converse Part: Informed Decoder, Uninformed Users)**.** *Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a compound channel. The capacities $Q_{ID}(\mathfrak{I})$ and $Q(\mathfrak{I})$ for $\mathfrak{I}$ with informed decoder and with uninformed users are bounded from above by*

$$Q(\mathfrak{I}) \leq Q_{ID}(\mathfrak{I}) \leq \lim_{l\to\infty} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} \frac{1}{l} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{376}$$

For the proof of this theorem, we will make use of the following Lemma:

Lemma 54 (Cf. [51]). *Let $\mathcal{N}, \mathcal{N}_i \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ and $d_{\mathcal{K}} = \dim \mathcal{K}$. Let $\mathcal{H}_r$ be an additional Hilbert space , $l \in \mathbb{N}$ and $\phi \in \mathcal{S}(\mathcal{H}_r \otimes \mathcal{H}^{\otimes l})$. If $||\mathcal{N} - \mathcal{N}_i||_\diamond \leq \epsilon$, then*

$$|S(id_{\mathcal{H}_r} \otimes \mathcal{N}^{\otimes l}(\phi)) - S(id_{\mathcal{H}_r} \otimes \mathcal{N}_i^{\otimes l}(\phi))| \leq l(4\epsilon \log(d_{\mathcal{K}}) + 2h(\epsilon)). \tag{377}$$

Here, $h(\cdot)$ denotes the binary entropy.

This result immediately implies the following Lemma:

Lemma 55. *Let $\mathcal{H}, \mathcal{K}$ be finite dimensional Hilbert spaces. There is a function $\nu : [0,1] \to \mathbb{R}_+$ with $\lim_{x \to 0} \nu(x) = 0$ such that for every $\mathfrak{I}, \mathfrak{I}' \subseteq \mathcal{C}(\mathcal{H}, \mathcal{K})$ with $D_\diamond(\mathfrak{I}, \mathfrak{I}') \leq \tau \leq 1/2$ and every $l \in \mathbb{N}$ we have the estimates*

1.
$$|\frac{1}{l} I_c(\rho, \mathfrak{I}^{\otimes l}) - \frac{1}{l} I_c(\rho, \mathfrak{I}'^{\otimes l})| \leq \nu(2\tau) \quad \forall \rho \in \mathcal{S}(\mathcal{H}^{\otimes l}) \tag{378}$$

2.
$$|\frac{1}{l} \inf_{\mathcal{N} \in \mathfrak{I}} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}) - \frac{1}{l} \inf_{\mathcal{N}' \in \mathfrak{I}'} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}'^{\otimes l})| \leq \nu(2\tau) \tag{379}$$

The function ν is given by $\nu(x) = x + 8x \log(d_{\mathcal{K}}) + 4h(x)$. Again, $h(\cdot)$ denotes the binary entropy.

Proof of Theorem 53. Again, the first inequality is easily seen to be true from the very definition of codes in the two cases, so we concentrate on the second. Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a compound channel and let for every $l \in \mathbb{N}$ an (l, k_l) code for $\mathfrak{I}$ with informed decoder be given such that $\liminf_{l \to \infty} \frac{1}{l} \log k_l = R$, and $\lim_{l \to \infty} \inf_{\mathcal{N} \in \mathfrak{I}} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_{\mathcal{N}}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l) = 1$ hold.
Take any $0 < \tau \leq 1/2$. Then it is easily seen that starting with a $\frac{\tau}{2}$-net in $\mathcal{C}(\mathcal{H}, \mathcal{K})$ we can find a set $\mathfrak{I}'_\tau = \{\mathcal{N}_1, \ldots, \mathcal{N}_{N_\tau}\} \subset \mathfrak{I}$ with $|N_\tau| \leq (\frac{6}{\tau})^{2(\dim \mathcal{H} \cdot \dim \mathcal{K})^2}$ such that for each $\mathcal{N} \in \mathfrak{I}$ there is $\mathcal{N}_i \in \mathfrak{I}'_\tau$ with

$$||\mathcal{N} - \mathcal{N}_i||_\diamond \leq \tau. \tag{380}$$

Clearly, the above sequence of codes satisfies for each $i \in \{1, \ldots, N_\tau\}$

1. $\liminf_{l \to \infty} \frac{1}{l} \log k_l = R$, and
2. $\lim_{l \to \infty} \min_{\mathcal{N}_i \in \mathfrak{I}_\tau} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) = 1$.

From Theorem 52 it is immediately clear then, that

$$R \leq \lim_{l \to \infty} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})} \min_{\mathcal{N}_i \in \mathfrak{I}'_\tau} \frac{1}{l} I_c(\rho, \mathcal{N}_i^{\otimes l}) \tag{381}$$

and from the first estimate in Lemma 55 we get by noting that $D_\diamond(\mathfrak{I}, \mathfrak{I}'_\tau) \leq \tau$ holds

$$R \leq \lim_{l \to \infty} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N} \in \mathfrak{I}} \frac{1}{l} I_c(\rho, \mathcal{N}^{\otimes l}) + \nu(2\tau). \tag{382}$$

Taking the limit $\tau \to 0$ proves the theorem.

3.6.2 The informed encoder

The case of an informed encoder can be treated in the same manner as the other two cases. We will just state the theorem and very briefly indicate the central ideas of the proof.

Theorem 56 (Converse Part: Informed Encoder)**.** *Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a compound channel. The capacity $Q_{IE}(\mathfrak{I})$ for $\mathfrak{I}$ with informed encoder is bounded from above by*

$$Q_{IE}(\mathfrak{I}) \leq \lim_{l\to\infty} \frac{1}{l} \inf_{\mathcal{N}\in\mathfrak{I}} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{383}$$

Proof. The proof of this theorem is a trivial modification of the one for Theorem 53. Again, the first part of the proof is the converse in the finite case, while the second part uses the second estimate in Lemma 55.
For the proof in the finite case note the following: due to the data processing inequality, the structure of the proof is entirely independent from the decoder. A change from an informed decoder to an uninformed decoder does not change our estimate. The only important change is that there will be a whole set $\{e^1_{i_1}, \ldots, e^N_{i_N}\}$ of vector states satisfying equation (361), one for each channel in $\mathfrak{I}$. This causes the state ρ^l in equation (370) to depend on the channel. □

3.7 Continuity of compound capacity

This section is devoted to a question that has been answered only recently in [51] for single-channel capacities, namely that of continuity of capacities of quantum channels.
The question is relevant not only from a mathematical point of view, but might also have a strong impact on applications. It seems a hard task in general to compute the regularized capacity formulas obtained so far for quantum channels. There are, however, cases where the regularized capacity formula can be reduced to a one-shot quantity (see for example [22] and references therein) that can be calculated using standard optimization techniques.
Knowing that capacity is a continuous quantity one could raise the question how close an arbitrary (compound) channel is to a (compound) channel with one-shot capacity and thereby get an estimate on arbitrary capacities.
We will now state the main result of this section.

Theorem 57 (Continuity of Compound Capacity)**.** *The compound capacities $Q(\,\cdot\,)$, $Q_{ID}(\,\cdot\,)$ and $Q_{IE}(\,\cdot\,)$ are continuous. To be more precise, let $\mathfrak{I}, \mathfrak{I}' \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be two compound channels with $D_\diamondsuit(\mathfrak{I}, \mathfrak{I}') \leq \epsilon \leq 1/2$. Then*

$$|Q(\mathfrak{I}) - Q(\mathfrak{I}')| = |Q_{ID}(\mathfrak{I}) - Q_{ID}(\mathfrak{I}')| \leq \nu(2\epsilon), \tag{384}$$

$$|Q_{IE}(\mathfrak{I}) - Q_{IE}(\mathfrak{I}')| \leq \nu(2\epsilon), \tag{385}$$

where the function ν is taken from Lemma 55.

Remark 58. *Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$. Then $D(\mathfrak{I}, \bar{\mathfrak{I}}) = 0$, implying that the three different capacities of $\mathfrak{I}$ coincide with those for $\bar{\mathfrak{I}}$. We may thus define the equivalence relation $\mathfrak{I} \sim \mathfrak{I}' \Leftrightarrow \bar{\mathfrak{I}} = \bar{\mathfrak{I}}'$ and even use $D_\diamondsuit$ as a metric on the set of equivalence classes without losing any information about our channels.*

Proof. Let $D_\diamondsuit(\mathfrak{I}, \mathfrak{I}') \leq \epsilon$. By the first estimate in Lemma 55 and the capacity formula $Q_{ID}(\mathfrak{I}) = Q(\mathfrak{I}) =$

$\lim_{l\to\infty}\frac{1}{l}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathfrak{I}^{\otimes l})$ we get

$$\begin{aligned}
|Q(\mathfrak{I})-Q(\mathfrak{I}')| &= |Q_{ID}(\mathfrak{I})-Q_{ID}(\mathfrak{I}')| && (386)\\
&= |\lim_{l\to\infty}\frac{1}{l}[\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathfrak{I}^{\otimes l})-\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathfrak{I}'^{\otimes l})]| && (387)\\
&= \lim_{l\to\infty}|\frac{1}{l}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathfrak{I}^{\otimes l})-\frac{1}{l}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathfrak{I}'^{\otimes l})| && (388)\\
&\leq \lim_{l\to\infty}\nu(2\epsilon) && (389)\\
&= \nu(2\epsilon). && (390)
\end{aligned}$$

For the proof in the case of an informed encoder let us first note that $Q_{IE}(\mathfrak{I}) = \lim_{l\to\infty}\inf_{\mathcal{N}\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathcal{N}^{\otimes l})$ holds. The second estimate in Lemma 55 justifies the following inequality:

$$\begin{aligned}
|Q_{IE}(\mathfrak{I})-Q_{IE}(\mathfrak{I}')| &= |\lim_{l\to\infty}\frac{1}{l}[\inf_{\mathcal{N}\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathcal{N}^{\otimes l})-\inf_{\mathcal{N}'\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathcal{N}'^{\otimes l})]| && (391)\\
&= \lim_{l\to\infty}|\frac{1}{l}\inf_{\mathcal{N}\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathcal{N}^{\otimes l})-\frac{1}{l}\inf_{\mathcal{N}'\in\mathfrak{I}}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho,\mathcal{N}'^{\otimes l})| && (392)\\
&\leq \lim_{l\to\infty}\nu(2\epsilon) && (393)\\
&= \nu(2\epsilon). && (394)
\end{aligned}$$

□

3.8 Entanglement-generating capacity of compound channels

In this last section we will apply the results obtained so far to derive the second statement of Theorem 27. Namely, we will determine the entanglement-generating capacity of quantum compound channels. We give the definitions of codes and capacity only for the most interesting case of uninformed users because there is no doubt that the reader will easily guess the definitions in the remaining cases. Nevertheless, we will state the coding result in all three cases.

Recall from the proof of Theorem 43 that to each subspace $\mathcal{G}\subset\mathcal{H}$ and $\epsilon>0$ we always can find a subspace $\mathcal{F}_l\subset\mathcal{G}^{\otimes l}\subset\mathcal{H}^{\otimes l}$, a recovery operation $\mathcal{R}^l\in\mathcal{C}(\mathcal{K}^{\otimes l},\mathcal{F}_l)$, and a unitary operation $\mathcal{U}^l\in\mathcal{C}(\mathcal{H}^{\otimes l},\mathcal{H}^{\otimes l})$ with

$$k_l=\dim\mathcal{F}_l\geq\lfloor 2^{l(\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}},\mathcal{N})-\frac{\epsilon}{2}-o(l^0))}\rfloor, \tag{395}$$

and

$$\inf_{\mathcal{N}\in\mathfrak{I}} F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}^{\otimes l}\circ\mathcal{U}^l)=1-o(l^0). \tag{396}$$

Notice that the maximally entangled state ψ_l in $\mathcal{F}_l\otimes\mathcal{F}_l$ purifies the maximally mixed state $\pi_{\mathcal{F}_l}$on $\mathcal{F}_l$ and defining $|\varphi_l\rangle\langle\varphi_l|:=\mathcal{U}^l(|\psi_l\rangle\langle\psi_l|)$, the relation (396) can be rewritten as

$$\inf_{\mathcal{N}\in\mathfrak{I}} F(|\psi_l\rangle\langle\psi_l|,id_{\mathcal{F}_l}\otimes\mathcal{R}^l\circ\mathcal{N}^{\otimes l}(|\varphi_l\rangle\langle\varphi_l|))=1-o(l^0). \tag{397}$$

This together with (395) shows that

$$E(\mathfrak{I})\geq\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\pi_{\mathcal{G}},\mathcal{N}). \tag{398}$$

Thus, using the compound BSST Lemma 42 and arguing as in the proof of Theorem 44, we can conclude that

$$E(\mathfrak{I})\geq Q(\mathfrak{I})=\lim_{l\to\infty}\frac{1}{l}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})}\inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho,\mathcal{N}^{\otimes l}). \tag{399}$$

Since $E(\mathfrak{I}) \leq E_{ID}(\mathfrak{I})$ holds it suffices to show

$$E_{ID}(\mathfrak{I}) \leq Q(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho, \mathcal{N}^{\otimes l}) \tag{400}$$

in order to establish the coding theorem for $E_{ID}(\mathfrak{I})$ and $E(\mathfrak{I})$ simultaneously.
The proof of (400) relies on Lemma 51 and the data processing inequality. Indeed, let $R \in \mathbb{R}_+$ be an achievable entanglement generation rate for $\mathfrak{I}$ with informed decoder and let $((\mathcal{R}^l_{\mathcal{N}})_{\mathcal{N}\in\mathfrak{I}}, \varphi_l)_{l\in\mathbb{N}}$ be a corresponding sequence of (l, k_l)-codes, i.e we have

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, and
2. $\inf_{\mathcal{N}\in\mathfrak{I}} F(|\psi_l\rangle\langle\psi_l|, (id_{\mathcal{F}_l} \otimes \mathcal{R}^l_{\mathcal{N}} \circ \mathcal{N}^{\otimes l})(|\varphi_l\rangle\langle\varphi_l|)) = 1 - \epsilon_l$ where $\lim_{l\to\infty} \epsilon_l = 0$ and ψ_l denotes the standard maximally entangled state on $\mathcal{F}_l \otimes \mathcal{F}_l$ with Schmidt rank k_l.

Set $\rho^l := \mathrm{tr}_{\mathcal{F}_l}(|\varphi_l\rangle\langle\varphi_l|)$ and

$$\sigma^l_{\mathcal{N}} := id_{\mathcal{F}_l} \otimes \mathcal{R}^l_{\mathcal{N}} \circ \mathcal{N}^{\otimes l}(|\varphi_l\rangle\langle\varphi_l|). \tag{401}$$

Then (remembering that $c_3 = \max\{-x\log(x) : x \in [0,1]\}$) the data processing inequality and Lemma 51 imply for each $\mathcal{N} \in \mathfrak{I}$

$$\begin{aligned} I_c(\rho^l, \mathcal{N}^{\otimes l}) &\geq I_c(\rho^l, \mathcal{R}^l_{\mathcal{N}} \circ \mathcal{N}^{\otimes l}) && (402)\\ &= \Delta(\sigma^l_{\mathcal{N}}) && (403)\\ &\geq \Delta(|\psi_l\rangle\langle\psi_l|) - 4c_3 - 8\log(k_l)\sqrt{\epsilon_l} && (404)\\ &= \log k_l - 4c_3 - 16\log(k_l)\sqrt{2\epsilon_l}. && (405) \end{aligned}$$

Consequently, for every $l \in \mathbb{N}$,

$$(1 - 8\sqrt{\epsilon_l})\frac{1}{l}\log k_l \leq \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho, \mathcal{N}^{\otimes l}) + c_3/l \tag{406}$$

and we end up with

$$R \leq \liminf_{l\to\infty} \frac{1}{l}\log k_l \leq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho, \mathcal{N}^{\otimes l}), \tag{407}$$

which implies (400). The expression for $E_{IE}(\mathfrak{I})$ is obtained in a similar fashion. We summarize the results in the following theorem.

Theorem 59 (Entanglement-Generating Capacities of $\mathfrak{I}$)**.** *For arbitrary compound channels $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ we have*

$$E(\mathfrak{I}) = E_{ID}(\mathfrak{I}) = Q(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho, \mathcal{N}^{\otimes l}), \tag{408}$$

and

$$E_{IE}(\mathfrak{I}) = Q_{IE}(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \inf_{\mathcal{N}\in\mathfrak{I}} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{409}$$

3.9 Equivalence of strong subspace and entanglement transmission

We will now use results from convex high-dimensional geometry to show that every sequence of asymptotically perfect (random) codes for entanglement transmission for $\mathfrak{I}$ yields another sequence of (random) codes that guarantees asymptotically perfect strong subspace transmission.
First, we state the following theorem which is the complex version of a theorem that can essentially be picked up from [55], Theorem 2.4 and Remark 2.7:

Theorem 60. *For $\delta, \Theta > 0$ and an integer n let $k(\delta, \Theta, n) = \lfloor \delta^2(n-1)/(2\log(4/\Theta)) \rfloor$. Let $f : S(\mathbb{C}^n) \to \mathbb{R}$ be a continuous function and ν_k the uniform measure induced on the Grassmannian $G_{n,k} := \{G \subset \mathbb{C}^n : G \text{ is subspace and } \dim G = k\}$ by the normalized Haar measure on the unitary group on $\mathbb{C}^n$ then, for all $\delta, \Theta > 0$, the measure of the set $E_k \subset G_{n,k}$ of all subspaces $E \subset \mathbb{C}^n$ satisfying the three conditions*

1. $\dim E = k(\delta, \Theta, n)$
2. *There is a $\Theta-$net N in $S(E) = S(\mathbb{C}^n) \bigcap E$ such that $|f(x) - M_f| \leq \omega_f(\delta)$ for all $x \in N$*
3. $|f(x) - M_f| \leq \omega_f(\delta) + \omega_f(\Theta)$ *for all $x \in S(E)$*

satisfies $\nu_k(E_k) \geq 1 - \sqrt{2/\pi} e^{-\delta^2(n-1)/2}$.

Here, $S(\mathbb{C}^n)$ is the unit sphere in $\mathbb{C}^n$, $\omega_f(\delta) := \sup\{|f(x) - f(y)| : D(x,y) \leq \delta\}$ is the modulus of continuity, D the geodesic metric on $S(\mathbb{C}^n)$ and M_f the median of f, which is the number such that with ν the Haar measure on $S(\mathbb{C}^n)$ both $\nu(\{x : f(x) \leq M_f\}) \geq 1/2$ and $\nu(\{x : f(x) \geq M_f\}) \geq 1/2$ hold.

Remark 61. *The proof of Theorem 60 uses the identification $\mathbb{C}^n \simeq \mathbb{R}^{2n}$ under the map $\sum_{i=1}^n z_i e_i \mapsto \sum_{i=1}^n (\mathfrak{Re}\{z_i\} e_i + \mathfrak{Im}\{z_i\} e_{i+n})$, where $\{e_1, \ldots, e_n\}$ and $\{e_1, \ldots, e_{2n}\}$ denote the standard bases in $\mathbb{C}^n$ and $\mathbb{R}^{2n}$.*

Second, we use the following lemma which first appeared in [42]:

Lemma 62. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{H})$, $\mathcal{G}$ a d-dimensional subspace of $\mathcal{H}$ and $\phi \in \mathcal{G}$ with euclidean norm $\|\phi\| = 1$. Then*

$$\int_{\mathfrak{U}(\mathcal{G})} \langle U\phi, \mathcal{N}(U|\phi\rangle\langle\phi|U^*)U\phi\rangle dU = \frac{1}{d+1}(d \cdot F_e(\pi_{\mathcal{G}}, \mathcal{N}) + 1). \tag{410}$$

Third, we need a well behaving relation between the median and the expectation of a function $f : S(\mathbb{C}^n) \to \mathbb{R}$. This is given by Proposition 14.3.3 taken from [54]:

Lemma 63. *Let $f : S(\mathbb{C}^n) \to \mathbb{R}$ be Lipschitz with constant one (w.r.t. the geodesic metric). Then*

$$|M_f - \mathbb{E}(f)| \leq \frac{12}{\sqrt{2(n-1)}}. \tag{411}$$

Remark 64. *Obviously, this implies $|M_f - \mathbb{E}(f)| \leq \frac{12 \cdot L}{\sqrt{2(n-1)}}$ for Lipschitz functions with constant $L \in \mathbb{R}_+$.*

The function that we will apply Lemma 63 to is given by the following:

Lemma 65. *Let $\Lambda \in \mathcal{C}(\mathcal{H}, \mathcal{H})$. Define $f_\Lambda : S(\mathcal{H}) \to \mathbb{R}$ by*

$$f_\Lambda(x) := \langle x, \Lambda(|x\rangle\langle x|)x\rangle, \ x \in S(\mathcal{H}). \tag{412}$$

Then f_Λ is Lipschitz with constant $L = 4$ (w.r.t. the geodesic metric).

Proof. Let $x, y \in S(\mathcal{H})$. Then by Hölder's inequality,

$$\begin{aligned}
|f_\Lambda(x) - f_\Lambda(y)| &= |\mathrm{tr}(|x\rangle\langle x|\Lambda(|x\rangle\langle x|)) - \mathrm{tr}(|y\rangle\langle y|\Lambda(|y\rangle\langle y|))| && (413)\\
&= |\mathrm{tr}(|x\rangle\langle x|\Lambda(|x\rangle\langle x| - |y\rangle\langle y|))| + |\mathrm{tr}((|x\rangle\langle x| - |y\rangle\langle y|)\Lambda(|y\rangle\langle y|))| && (414)\\
&\leq \| \ |x\rangle\langle x| \ \|_\infty \cdot \|\Lambda(|x\rangle\langle x| - |y\rangle\langle y|)\|_1 + \| \ |x\rangle\langle x| - |y\rangle\langle y| \ \|_1 \cdot \|\Lambda(|y\rangle\langle y|)\|_\infty && (415)\\
&\leq \|\Lambda(|x\rangle\langle x| - |y\rangle\langle y|)\|_1 + \||x\rangle\langle x| - |y\rangle\langle y|\|_1 && (416)\\
&\leq 2\| \ |x\rangle\langle x| - |y\rangle\langle y| \ \|_1. && (417)
\end{aligned}$$

It further holds, with $\|\cdot\|$ denoting the euclidean norm,

$$\| \ |x\rangle\langle x| - |y\rangle\langle y| \ \|_1 \leq 2\|x - y\| \leq 2D(x,y). \tag{418}$$

□

We now state the main ingredient of this section.

Lemma 66. *Let $(\mathcal{I}^l)_{l\in\mathbb{N}}$ be a sequence of finite sets of channels such that for every $l \in \mathbb{N}$ the following, mostly technical, assumptions hold:*

1. *(w.l.o.g.) $\mathcal{I}^l$ is indexed such that $\mathcal{I}^l = \{\mathcal{N}_1, \dots, \mathcal{N}_{|\mathcal{I}^l|}\}$*
2. *There is a set of measures $\{\mu_{l,i}\}_{i=1}^{|\mathcal{I}^l|}$ s.t. each $\mu_{l,i}$ is a measure on $(\mathcal{C}(\mathcal{F}_l, \mathcal{H}^{\otimes l}) \times \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l'), \sigma_l)$ where $\mathcal{F}_l, \mathcal{F}_l'$ are Hilbert spaces, $\mathcal{F}_l \subset \mathcal{F}_l'$*
3. *The sigma-algebra σ_l is chosen such that the function $(\mathcal{P}_l, \mathcal{R}_l) \mapsto F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_l \circ \mathcal{N} \circ \mathcal{P}_l)$ is measurable w.r.t. σ_l for every $\mathcal{N} \in \mathcal{I}^l$.*
4. *All singleton sets are contained in σ_l. An example of such a sigma-algebra σ_l is given by the product of sigma-algebras of Borel sets induced on $\mathcal{C}(\mathcal{F}_l, \mathcal{H})$ and $\mathcal{C}(\mathcal{K}, \mathcal{F}_l')$ by the standard topologies of the ambient spaces.*

Additionally, we require that

A1 $\min_{1\le i\le|\mathcal{I}^l|} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i \circ \mathcal{P}^l) d\mu_{l,i}(\mathcal{R}^l, \mathcal{P}^l) = 1 - f_l$,
where $(f_l)_{l\in\mathbb{N}}$ is any sequence of real numbers in the interval $[0, 1]$.

Let $(\varepsilon_l)_{l\in\mathbb{N}}$ be a sequence with $\varepsilon_l \in (0, 1]$ $\forall l \in \mathbb{N}$ satisfying

A2 *There is $\hat{l} \in \mathbb{N}$ such that $|\mathcal{I}^l|\sqrt{2/\pi}e^{-\varepsilon_l^2(k_l-1)/128} < 1$ and $k_l \ge 2$ hold for all $l \ge \hat{l}$ (where, as usual, $k_l := \dim \mathcal{F}_l$).*

Then for any $l \ge \hat{l}$ there is a subspace $\hat{\mathcal{F}}_l \subset \mathcal{F}_l$ with the properties

P1 $\dim \hat{\mathcal{F}}_l = \lfloor \frac{\varepsilon_l^2}{256 \log(32/\varepsilon_l)} \cdot k_l \rfloor$,

P2 $\min_{1\le i\le|\mathcal{I}^l|} \min_{\phi\in S(\hat{\mathcal{F}}_l)} \int \langle \phi, \mathcal{R}^l \circ \mathcal{N}_i \circ \mathcal{P}^l(|\phi\rangle\langle\phi|)\phi\rangle d\mu_{l,i}(\mathcal{R}^l, \mathcal{P}^l) \ge 1 - f_l - \frac{4\cdot 12}{\sqrt{2(k_l-1)}} - \varepsilon_l$.

Proof. Let $l \in \mathbb{N}$. For an arbitrary $i \in \{1, \dots, |\mathcal{I}^l|\}$ define $f_{l,i} : S(\mathcal{F}_l) \to \mathbb{R}$ by

$$f_{l,i}(\phi) := \int \langle \phi, \mathcal{R}^l \circ \mathcal{N}_i \circ \mathcal{P}^l(|\phi\rangle\langle\phi|)\phi\rangle d\mu_{l,i}(\mathcal{R}^l, \mathcal{P}^l) \quad (\phi \in S(\mathcal{F}_l)). \tag{419}$$

Since $f_{l,i}$ is an affine combination of functions with Lipschitz-constant $L = 4$, it is itself Lipschitz with $L = 4$.
Also, by the Theorem of Fubini, Lemma 62 and our assumption *A1* we have

$$\begin{aligned}
\mathbb{E}(f_{l,i}) &= \int_{\mathfrak{U}(\mathcal{F}_l)} f_{l,i}(U\phi) dU && (420)\\
&= \int_{\mathfrak{U}(\mathcal{F}_l)} [\int \langle U\phi, \mathcal{R}^l \circ \mathcal{N}_i \circ \mathcal{P}^l(|U\phi\rangle\langle U\phi|)U\phi\rangle d\mu_{l,i}(\mathcal{R}^l, \mathcal{P}^l)] dU && (421)\\
&= \int \int_{\mathfrak{U}(\mathcal{F}_l)} [\langle U\phi, \mathcal{R}^l \circ \mathcal{N}_i \circ \mathcal{P}^l(|U\phi\rangle\langle U\phi|)U\phi\rangle dU] d\mu_{l,i}(\mathcal{R}^l, \mathcal{P}^l) && (422)\\
&= \int \frac{k_l F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i \circ \mathcal{P}^l) + 1}{k_l + 1} d\mu_{l,i}(\mathcal{R}^l, \mathcal{P}^l) && (423)\\
&\ge \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i \circ \mathcal{P}^l) d\mu_{l,i}(\mathcal{R}^l, \mathcal{P}^l) && (424)\\
&= 1 - f_l. && (425)
\end{aligned}$$

By Lemma 63 and Lemma 65 we now get a good lower bound on the median of f_{s^l}:

$$M_{f_{l,i}} \geq \mathbb{E}(f_{l,i}) - \frac{4 \cdot 12}{\sqrt{2(k_l - 1)}} \tag{426}$$

$$\geq 1 - f_l - \frac{4 \cdot 12}{\sqrt{2(k_l - 1)}}. \tag{427}$$

We now apply Theorem 60 with $n = k_l$ and $\delta = \Theta = \varepsilon_l/8$ to f_{s^l}. Then $k(\varepsilon_l/8, \varepsilon_l/8, k_l) \geq \lfloor \frac{\varepsilon_l^2}{256 \log(32/\varepsilon_l)} k_l \rfloor$ holds due to the second estimate in *A2*.
We set $k_l' := \lfloor \frac{\varepsilon_l^2}{256 \log(32/\varepsilon_l)} k_l \rfloor$.
Since the fact that $f_{l,i}$ is 4-Lipschitz implies $\omega_{f_{l,i}}(\delta) \leq 4\delta$ we get the following:

$$\nu_k(\{E \in G_{k_l, k(\varepsilon_l/8, \varepsilon_l/8, k_l)} : |f_{l,i}(\phi) - M_{f_{s^l}}| \leq \varepsilon_l \ \forall \phi \in S(E)\}) \geq 1 - \sqrt{2/\pi} e^{-\varepsilon_l^2 (k_l - 1)/128}. \tag{428}$$

The last inequality is valid for each choice of $1 \leq i \leq |\mathcal{I}^l|$, so we can conclude that

$$\nu_k(\{E \in G_{k_l, k(\frac{\varepsilon_l}{8}, \frac{\varepsilon_l}{8}, k_l)} : |f_{l,i}(\phi) - M_{f_{l,i}}| \leq \varepsilon_l \ \forall \phi \in S(E), \ 1 \leq i \leq |I^l|\}) \geq 1 - |\mathcal{I}^l| \sqrt{\frac{2}{\pi}} e^{-\frac{\varepsilon_l^2 (k_l - 1)}{128}}. \tag{429}$$

Thus for all $l \geq \hat{l}$ we have

$$\nu_k(\{E \in G_{k_l, k(\varepsilon_l/8, \varepsilon_l/8, k_l)} : |f_{l,i}(\phi) - M_{f_{l,i}}| \leq \varepsilon_l \ \forall \phi \in S(E), \ \forall 1 \leq i \leq |\mathcal{I}^l|\}) > 0 \tag{430}$$

by assumption *A2*, implying the existence of a subspace $E \subset \mathcal{F}_l$ of dimension $\dim E = k(\varepsilon_l/8, \varepsilon_l/8, k_l)$ such that

$$|f_{l,i}(\phi) - M_{f_{l,i}}| \leq \varepsilon_l \ \forall \ \phi \in S(E), \ i \in \{1, \ldots, |\mathcal{I}^l|\}. \tag{431}$$

Let $\hat{\mathcal{F}}_l \subset E$ be any subspace of dimension k_l'. Then *P1* holds and equation (427) together with (431) establishes *P2*:

$$f_{l,i}(\phi) \geq 1 - f_l - \frac{4 \cdot 12}{\sqrt{2(k_l - 1)}} - \varepsilon_l \ \ \forall \ \phi \in S(\hat{\mathcal{F}}_l), \ \forall 1 \leq i \leq |\mathcal{I}^l|. \tag{432}$$

□

We turn to an application of Lemma 66.

Lemma 67. *[Third statement in Theorem 27] Let $\mathfrak{I}$ be a compound channel. Then*

$$1. \ \ Q_{ID}(\mathfrak{I}) = Q_{s,ID}(\mathfrak{I}), \qquad 2. \ \ Q_{IE}(\mathfrak{I}) = Q_{s,IE}(\mathfrak{I}), \qquad 3. \ \ Q(\mathfrak{I}) = Q_s(\mathfrak{I}). \tag{433}$$

Proof. We first show that the l.h.s. is upper bounded by the r.h.s. in all three of the above cases. This is trivially true in case that the l.h.s. equals zero, so we concentrate on the case where the l.h.s. is strictly greater than zero.
We then uniformly describe the codes in the three settings by probability measures. But first, a few approximations have to be carried out. For every number $l \in \mathbb{N}$, we let N_{τ_l}' denote a τ_l-net with $\tau_l := 2^{-lR/4}$ and cardinality bounded by $|N_{\tau_l}'| \leq (3/\tau_l)^{2(dd')^2}$ in $\mathcal{C}(\mathcal{H}, \mathcal{K})$. Set $N_{\tau_l} := N_{\tau_l}' \bigcap \mathfrak{I}$. Then

$$D_\diamond(N_{\tau_l}, \mathfrak{I}^{\otimes l}) \leq 2\tau_l \tag{434}$$

and, what is more, multiplicativity of $\|\cdot\|_\diamond$ also grants us, for all $\mathcal{N}\in\mathfrak{I}$, $\mathcal{R}^l\in\mathcal{C}(\mathcal{K}^{\otimes l},\mathcal{F}_l')$ and $\mathcal{P}^l\in\mathcal{C}(\mathcal{F}_l,\mathcal{H}^{\otimes l})$, the estimate

$$\min_{\mathcal{N}_i\in N_{\tau_l}}\max_{x\in S(\mathcal{F}_l)}\langle x,[\mathcal{R}^l\circ(\mathcal{N}_i^{\otimes l}-\mathcal{N}^{\otimes l})\circ\mathcal{P}^l](|x\rangle\langle x|)x\rangle\leq l2\tau_l. \tag{435}$$

We set $\mathcal{I}^l:=N_{\tau_l}$ and take an arbitrary sequence of sets of measures $(\{\mu_{l,i}\}_{i=1}^{|N_{\tau_l}|})_{l\in\mathbb{N}}$ satisfying all of the four technical assumptions in Lemma 66.
Additionally, let the sequence $(f_l)_{l\in\mathbb{N}}$ in Lemma 66, *A1* be such that $f_l\searrow 0$. The dimensions of the subspaces $\mathcal{F}_l$ shall satisfy $\liminf_{l\to\infty}\frac{1}{l}\log k_l\geq R>0$ and we set $\varepsilon_l:=2^{-lR/4}$.
A short calculation reveals that $|N_{\tau_l}|\leq 2^{lR(dd')^2/2}$, so *A2* is clearly satisfied for some fixed $\hat{l}\in\mathbb{N}$ and it follows

$$f_{l,i}(\phi)\geq 1-f_l-\frac{4\cdot 12}{\sqrt{2(k_l-1)}}-\varepsilon_l\quad\forall\, l\geq\hat{l},\ \phi\in S(\hat{\mathcal{F}}_l),\quad 1\leq i\leq|N_{\tau_l}|. \tag{436}$$

Under application of (435), this translates to

$$\inf_{\mathcal{N}\in\mathfrak{I}}\min_{\phi\in\hat{\mathcal{F}}_l}\int\langle\phi,\mathcal{R}^l\circ\mathcal{N}_i^{\otimes l}\circ\mathcal{P}^l(|\phi\rangle\langle\phi|)\phi\rangle d\mu_{l,i}(\mathcal{R}^l,\mathcal{P}^l)\geq 1-f_l-\frac{4\cdot 12}{\sqrt{2(k_l-1)}}-\varepsilon_l-l2\tau_l\quad\forall\, l\geq\hat{l}. \tag{437}$$

Additionally, we obviously have

$$\liminf_{l\to\infty}\frac{1}{l}\log\dim\hat{\mathcal{F}}_l=\liminf_{l\to\infty}\frac{1}{l}\log\dim\mathcal{F}_l\geq R. \tag{438}$$

Let $R>0$ be achievable for transmission of entanglement over $\mathfrak{I}$ with informed decoder. Thus, there exists a sequence of (l,k_l) codes for $\mathfrak{I}$ with informed decoder such that 1. $\liminf_{l\to\infty}\frac{1}{l}\log k_l\geq R$, and 2. $\lim_{l\to\infty}\inf_{\mathcal{N}\in\mathfrak{I}}F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l_{\mathcal{N}}\circ\mathcal{N}^{\otimes l}\circ\mathcal{P}^l)=1$
Define $\mathcal{R}_i^l:=\mathcal{R}^l_{\mathcal{N}_i}$ and $\mu_{l,i}:=\delta_{(\mathcal{R}_i^l,\mathcal{P}^l)}$, where δ_a as usually denotes the point measure at the point a.
Our above discussion shows the existence of a sequence $(\hat{\mathcal{F}}_l)_{l\in\mathbb{N}}$ of subspaces $\hat{\mathcal{F}}_l\subset\mathcal{F}_l$ such that (compare equations (437) and (438))

$$\min_{1\leq i\leq N}\min_{\phi\in\hat{\mathcal{F}}_l}\langle\phi,\mathcal{R}_i^l\circ\mathcal{N}_i^{\otimes l}\circ\mathcal{P}^l(|\phi\rangle\langle\phi|)\phi\rangle\geq 1-f_l-\frac{4\cdot 12}{\sqrt{2(k_l-1)}}-\varepsilon_l-2l\tau_l\quad\forall\, l\geq\hat{l}. \tag{439}$$

and

$$\liminf_{l\to\infty}\frac{1}{l}\log\dim\hat{\mathcal{F}}_l=\liminf_{l\to\infty}\frac{1}{l}\log\dim\mathcal{F}_l\geq R. \tag{440}$$

This shows that not only is R an achievable rate for strong subspace transmission over $\mathfrak{I}$ with informed decoder, but moreover, if $f_l\to 0$ exponentially fast, then the convergence on the r.h.s. of (439) is exponentially fast as well.
The cases of informed encoder and of uninformed users can be handled by setting either $\mu_{l,i}:=\delta_{(\mathcal{R}^l,\mathcal{P}_i^l)}$ or $\mu_{l,i}=\mu_l:=\delta_{(\mathcal{R}^l,\mathcal{P}^l)}$. Conclusions we drew about the speed of convergence for the informed decoder remain valid for the other two cases as well.
That the r.h.s. is upper bounded by the l.h.s. in all the three cases of Lemma 67 follows from an application of ([10], Theorem 2). □

Having proven Lemma 67, we are also finally finished with the proof of Theorem 27.

3.10 A symmetrizability condition for compound quantum channels

The notion of symmetrizability stems from the theory of classical arbitrarily varying channels (AVCs). A more detailed description of the topic can be found in chapter 4. In this section, we give an example of the

impact that the symmetrizability criteria (see section 4.6) for arbitrarily varying quantum channels could have on compound quantum channels. We concentrate on classical message transmission with average error criterion and entanglement transmission and abstain from defining other notions of symmetrizablity for compound quantum channels than those that are necessary for this specific task.
This is mostly due to limited belief in the relevance of these notions in our focus, the field of entanglement and strong subspace transmission or entanglement generation - they seem to be designed to either cope with classical message transmission only (see subsections 4.6.1, 4.6.1) or are too restrictive to be necessary and sufficient at the same time (see subsection 4.6.3).
Let us further restrict to the case of uninformed users for a *finite* compound quantum channel $\mathfrak{I} = \{\mathcal{N}_1, \dots, \mathcal{N}_N\}$. The capacity for transmission of classical messages using the average error probability criterion can be defined as follows.

Definition 68. *An (l, M_l)-(deterministic) code for message transmission is a family of pairs $\mathfrak{C}_l = (\rho_i, D_i)_{i=1}^{M_l}$ where $\rho_1, \dots, \rho_{M_l} \in \mathcal{S}(\mathcal{H}^{\otimes l})$, and positive semi-definite operators $D_1, \dots, D_{M_l} \in \mathcal{B}(\mathcal{K}^{\otimes l})$ satisfying $\sum_{i=1}^{M_l} D_i = \mathbf{1}_{\mathcal{K}^{\otimes l}}$.*
The worst-case average probability of error of a code $\mathfrak{C}_l$ is given by

$$\bar{P}_{e,l}(\mathfrak{I}) := \max_{1 \leq j \leq N} \bar{P}_e(\mathfrak{C}_l, j), \tag{441}$$

where for $1 \leq j \leq N$ we set

$$\bar{P}_e(\mathfrak{C}_l, j) := \frac{1}{M_l} \sum_{i=1}^{M_l} \left(1 - \mathrm{tr}(\mathcal{N}_j(\rho_i) D_i)\right). \tag{442}$$

The achievable rates and the classical deterministic capacity $C_{\mathrm{compound}}(\mathfrak{I})$ of $\mathfrak{I}$, with respect to the error criterion given in (442), are then defined in the usual way.

Definition 69. *Let $l \in \mathbb{N}$. The compound channel $\mathfrak{I}$ is called l-symmetrizable, if for each finite set $\{\rho_1, \dots, \rho_K\} \subset \mathcal{S}(\mathcal{H}^{\otimes l})$, $K \in \mathbb{N}$, there is a map $p : \{\rho_1, \dots, \rho_K\} \to \mathfrak{P}(\{1, \dots, N\})$, $\rho_i \mapsto p[\rho_i]$ such that for all $i, j \in \{1, \dots, K\}$*

$$\sum_{m=1}^{N} p[\rho_i](m) \mathcal{N}_m^{\otimes l}(\rho_j) = \sum_{m=1}^{N} p[\rho_j](m) \mathcal{N}_m^{\otimes l}(\rho_i) \tag{443}$$

holds. We call $\mathfrak{I}$ symmetrizable if it is l-symmetrizable for all $l \in \mathbb{N}$.

Given these definitions, we can state the following theorem.

Theorem 70. *The finite compound channel $\mathfrak{I}$ is symmetrizable if and only if $C_{\mathrm{compound}}(\mathfrak{I}) = 0$. If $\mathfrak{I}$ is symmetrizable, then also $Q(\mathfrak{I}) = 0$.*

Remark 71. *If $\mathfrak{I}$ is l-symmetrizable, it is not at all clear that it is $(l+1)$-symmetrizable as well (this follows from the fact that only channels of the form $\mathcal{N}_i^{\otimes l}$ enter the definition (compare Definition 104, where this incompatibility is in principle due to non-separability of some quantum states). Even if a compound channel $\{\mathcal{N}_1, \mathcal{N}_2\}$ satisfies $\lambda\mathcal{N}_1 + (1-\lambda)\mathcal{N}_2 = \mathcal{T}$ for some $\lambda \in (0,1)$ and $\mathcal{T}(\cdot) := \pi_{\mathcal{K}}\mathrm{tr}(\cdot)$, which implies 1-symmetrizability, it is not automatically clear that it is 2-symmetrizable.*
It is also clear that the case $Q(\mathfrak{I}) = 0$ and $C_{\mathrm{compound}}(\mathfrak{I}) > 0$ can occur - just let $\mathfrak{I}$ consist of one single channel and let this channel be entanglement breaking.

Proof. The proof is an immediate consequence of the proof of Theorem 105, which simply has to be read with $\mathbf{S}^l$ replaced by $\{1, \dots, N\}$ (for every $l \in \mathbb{N}$). □

The topic will be further discussed in Chapter 5.

4 The arbitrarily varying quantum channel

This chapter evolves around the task of entanglement transmission over an unknown channel in the presence of a third party (called the adversary), which is enabled to choose the channel from a given set of memoryless but non-stationary channels without informing the legitimate sender and receiver about the particular choice that he made. This channel model is called arbitrarily varying quantum channel (AVQC).
A quantum version of Ahlswede's dichotomy for classical arbitrarily varying channels is derived. This includes a regularized formula for the common randomness-assisted capacity for entanglement transmission of an AVQC. Quite surprisingly and in contrast to the classical analog of the problem involving the maximal and average error probability, it turns out that the capacity for entanglement transmission of an AVQC always equals its strong subspace transmission capacity.
These results are accompanied by different notions of symmetrizability (zero-capacity conditions) as well as by conditions for an AVQC to have a capacity described by a single-letter formula. In the final part of the paper the capacity of the erasure-AVQC is computed and some light shed on the connection between AVQCs and zero-error capacities. Additionally, it is shown by entirely elementary and operational arguments motivated by the theory of AVQCs that the quantum, classical, and entanglement-assisted zero-error capacities of quantum channels are generically zero and are discontinuous at every positivity point.
In the last part of this chapter a quick sidestep to the entanglement generation capacities of an AVQC is made. Some results are proven to hold by trivial modifications of the proofs for entanglement transmission, only the question whether $G_{\det}(\mathfrak{I}) > \mathcal{A}_{\det}(\mathfrak{I})$ can occur for an AVQC $\mathfrak{I}$ is left open.

4.1 Basic definitions and main results

We now introduce the quantities that we will be dealing with in the rest of the paper: Arbitrarily varying quantum channels and codes for transmission of entanglement and subspaces. Since they will be of importance for our derandomization arguments, we will also include definitions of the capacities for message transmission with average and maximal error probability criterion.
Our most basic object is the arbitrarily varying quantum channel (AVQC). It is generated by a set $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ of CPTP maps with input Hilbert space $\mathcal{H}$ and output Hilbert space $\mathcal{K}$ and given by the family of CPTP maps $\{\mathcal{N}_{s^l} : \mathcal{B}(\mathcal{H})^{\otimes l} \to \mathcal{B}(\mathcal{K})^{\otimes l}\}_{l\in\mathbb{N}, s^l\in\mathbf{S}^l}$, where

$$\mathcal{N}_{s^l} := \mathcal{N}_{s_1} \otimes \ldots \otimes \mathcal{N}_{s_l} \qquad (s^l \in \mathbf{S}^l). \tag{444}$$

Thus, even in the case of a finite set $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$, showing the existence of reliable codes for the AVQC determined by $\mathfrak{I}$ is a non-trivial task: For each block length $l \in \mathbb{N}$ we have to deal with $|\mathfrak{I}|^l$, i.e. exponentially many, memoryless partly non-stationary quantum channels simultaneously.

In order to relieve ourselves from the burden of complicated notation we will simply write $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ for the AVQC. If instead $\mathfrak{I}$ shall denote a compound channel, this will be stated explicitly to avoid severe notational collisions.

4.1.1 Entanglement transmission

For the rest of this subsection, let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an AVQC.

Definition 72. *An* $(l, k_l)-$random entanglement transmission code *for* $\mathfrak{I}$ *is a probability measure* μ_l *on* $(\mathcal{C}(\mathcal{F}_l, \mathcal{H}^{\otimes l}) \times \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l'), \sigma_l)$*, where* $\mathcal{F}_l, \mathcal{F}_l'$ *are Hilbert spaces,* $\dim \mathcal{F}_l = k_l$*,* $\mathcal{F}_l \subset \mathcal{F}_l'$ *and the sigma-algebra* σ_l *is chosen such that the function* $(\mathcal{P}_l, \mathcal{R}_l) \mapsto F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_l \circ \mathcal{N}_{s^l} \circ \mathcal{P}_l)$ *is measurable w.r.t.* σ_l *for every* $s^l \in \mathbf{S}^l$*.*
Moreover, we assume that σ_l *contains all singleton sets. An example of such a sigma-algebra* σ_l *is given*

by the product of sigma-algebras of Borel sets induced on $\mathcal{C}(\mathcal{F}_l, \mathcal{H})$ and $\mathcal{C}(\mathcal{K}, \mathcal{F}_l')$ by the standard topologies of the ambient spaces.

Definition 73. *A non-negative number R is said to be an achievable entanglement transmission rate for the AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ with random codes if there is a sequence of $(l, k_l)-$random entanglement transmission codes such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$ *and*
2. $\lim_{l\to\infty} \inf_{s^l\in\mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) = 1.$

The random entanglement transmission capacity $\mathcal{A}_{\text{random}}(\mathfrak{I})$ of $\mathfrak{I}$ is defined by

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) := \sup \left\{ R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable entanglement trans-} \\ \text{mission rate for } \mathfrak{I} \text{ with random codes} \end{array} \right\}. \tag{445}$$

Having defined random codes and random code capacity for entanglement transmission we are in the position to introduce their deterministic counterparts: An $(l, k_l)-$code for entanglement transmission over $\mathfrak{I}$ is an $(l, k_l)-$random code for $\mathfrak{I}$ with $\mu_l(\{(\mathcal{P}^l, \mathcal{R}^l)\}) = 1$ for some encoder-decoder pair $(\mathcal{P}^l, \mathcal{R}^l)$[2] and $\mu_l(A) = 0$ for any $A \in \sigma_l$ with $(\mathcal{P}^l, \mathcal{R}^l) \notin A$. We will refer to such measures as point measures in what follows.

Definition 74. *A non-negative number R is a deterministically achievable entanglement transmission rate for the AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ if it is achievable in the sense of Definition 73 for random codes with* point measures μ_l.
The deterministic entanglement transmission capacity $\mathcal{A}_{\text{det}}(\mathfrak{I})$ of $\mathfrak{I}$ is given by

$$\mathcal{A}_{\text{det}}(\mathfrak{I}) := \sup\{R : R \text{ is a deterministically achievable entanglement transmission rate for } \mathfrak{I}\}. \tag{446}$$

Finally, we shall need the notion of the classical deterministic capacity $C_{\text{det}}(\mathfrak{I})$ of the AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ with *average* error criterion.

Definition 75. *An (l, M_l)-(deterministic) code for message transmission is a family of pairs $\mathfrak{C}_l = (\rho_i, D_i)_{i=1}^{M_l}$ where $\rho_1, \ldots, \rho_{M_l} \in \mathcal{S}(\mathcal{H}^{\otimes l})$, and positive semi-definite operators $D_1, \ldots, D_{M_l} \in \mathcal{B}(\mathcal{K}^{\otimes l})$ satisfying $\sum_{i=1}^{M_l} D_i = \mathbf{1}_{\mathcal{K}^{\otimes l}}$.*
The worst-case average probability of error of a code $\mathfrak{C}_l$ is given by

$$\bar{P}_{e,l}(\mathfrak{I}) := \sup_{s^l\in\mathbf{S}^l} \bar{P}_e(\mathfrak{C}_l, s^l), \tag{447}$$

where for $s^l \in \mathbf{S}^l$ we set

$$\bar{P}_e(\mathfrak{C}_l, s^l) := \frac{1}{M_l} \sum_{i=1}^{M_l} \left(1 - \operatorname{tr}(\mathcal{N}_{s^l}(\rho_i) D_i)\right). \tag{448}$$

The achievable rates and the classical deterministic capacity $C_{det}(\mathfrak{I})$ of $\mathfrak{I}$, with respect to the error criterion given in (447), are then defined in the usual way.

For any AVQC (finite or infinite), the compound quantum channel generated by the set $\operatorname{conv}(\mathfrak{I})$ (cf. [16] for the relevant definition) shall play the crucial role in our derivation of the coding results below. In the relevant cases we will have $|\mathfrak{I}| > 1$ and, therefore, $\operatorname{conv}(\mathfrak{I})$ will be *infinite*.
Our main result, a quantum version of Ahlswede's dichotomy for finite AVQCs, goes as follows:

[2] This explains our requirement on σ_l to contain all singleton sets.

Theorem 76. *Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an AVQC.*

1. *With conv($\mathfrak{I}$) denoting the convex hull of $\mathfrak{I}$ we have*

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{449}$$

2. *Either $C_{\text{det}}(\mathfrak{I}) = 0$ or else $\mathcal{A}_{\text{det}}(\mathfrak{I}) = \mathcal{A}_{\text{random}}(\mathfrak{I})$.*

Proof. The claim made in (449) follows from Theorem 92 and Corollary 100.
The proof that $C_{\text{det}}(\mathfrak{I}) > 0$ implies $\mathcal{A}_{\text{det}}(\mathfrak{I}) = \mathcal{A}_{\text{random}}(\mathfrak{I})$ requires a derandomization argument which is presented in section 4.5. □

We conclude this subsection with some explaining remarks:
1. Coherent information depends continuously on the state, therefore $\rho \mapsto \inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l})$ is upper semicontinuous and thus $\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l})$ exists due to the compactness of $\mathcal{S}(\mathcal{H}^{\otimes l})$. The limit in (449) exists due to superadditivity of the sequence $(\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}))_{l\in\mathbb{N}}$.
2. It is clear that $\mathcal{A}_{\text{det}}(\mathfrak{I}) \leq C_{\text{det}}(\mathfrak{I})$, so that $C_{\text{det}}(\mathfrak{I}) = 0$ implies $\mathcal{A}_{\text{det}}(\mathfrak{I}) = 0$. Therefore, Theorem 76 gives a regularized formula for $\mathcal{A}_{\text{det}}(\mathfrak{I})$ in form of (449), and the question remains when $C_{\text{det}}(\mathfrak{I}) = 0$ happens. We derive a non-single-letter necessary and sufficient condition for the latter in section 4.6.
3. Continuous dependence of the coherent information on the channel reveals that for each $l \in \mathbb{N}$ and $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$

$$\inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}) = \min_{\mathcal{N}\in\overline{\text{conv}(\mathfrak{I})}} I_c(\rho, \mathcal{N}^{\otimes l}), \tag{450}$$

4.1.2 Strong subspace transmission

Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an AVQC. An $(l, k_l)-$*random strong subspace transmission code* for $\mathfrak{I}$ is a probability measure μ_l on $(\mathcal{C}(\mathcal{F}_l, \mathcal{H}^{\otimes l}) \times \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l'), \sigma_l)$, where $\mathcal{F}_l$, $\mathcal{F}_l'$ are Hilbert spaces, $\dim \mathcal{F}_l = k_l$, $\mathcal{F}_l \subset \mathcal{F}_l'$ and the sigma-algebra σ_l is chosen such that the function $(\mathcal{P}^l, \mathcal{R}^l) \mapsto F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l)$ is measurable w.r.t. σ_l for every $s^l \in \mathbf{S}^l$. Again, we assume that σ_l contains all singleton sets.

Definition 77. *A non-negative number R is said to be an achievable strong subspace transmission rate for the AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ with random codes if there is a sequence of $(l, k_l)-$random strong subspace transmission codes such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$ *and*
2. $\lim_{l\to\infty} \inf_{s^l\in\mathbf{S}^l} \min_{\psi\in S(\mathcal{F}_l)} \int F(|\psi\rangle\langle\psi|, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l(|\psi\rangle\langle\psi|)) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) = 1.$

The random strong subspace transmission capacity $\mathcal{A}_{\text{s,random}}(\mathfrak{I})$ of $\mathfrak{I}$ is defined by

$$\mathcal{A}_{\text{s,random}}(\mathfrak{I}) := \sup \left\{ R \in \mathbb{R}_+ : \begin{array}{l} R \text{ is an achievable strong subspace trans-} \\ \text{mission rate for } \mathfrak{I} \text{ with random codes} \end{array} \right\}. \tag{451}$$

As before we also define deterministic codes: A *deterministic* $(l, k_l)-$strong subspace transmission code for $\mathfrak{I}$ is an $(l, k_l)-$random strong subspace transmission code for $\mathfrak{I}$ with $\mu_l(\{(\mathcal{P}^l, \mathcal{R}^l)\}) = 1$ for some encoder-decoder pair $(\mathcal{P}^l, \mathcal{R}^l)$ and $\mu_l(A) = 0$ for any $A \in \sigma_l$ with $(\mathcal{P}^l, \mathcal{R}^l) \notin A$. We will refer to such measures as point measures in what follows.

Definition 78. *A non-negative number R is a deterministically achievable strong subspace transmission rate for the AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ if it is achievable in the sense of Definition 77 for random codes with* point measures μ_l.
The deterministic capacity $\mathcal{A}_{\text{s,det}}(\mathfrak{I})$ for strong subspace transmission over an AVQC $\mathfrak{I}$ is given by

$$\mathcal{A}_{\text{s,det}}(\mathfrak{I}) := \sup\{R : R \text{ is a deterministically achievable rate for } \mathfrak{I}\}. \tag{452}$$

If we want to transmit classical messages, then the error criterion that is most closely related to strong subspace transmission is that of maximal error probability. It leads to the notion of classical deterministic capacity with maximal error:

Definition 79. *Let $\mathfrak{C}_l$ be an (l, M_l)-(deterministic) code for message transmission as given in Definition 75. The worst-case maximal probability of error of the code $\mathfrak{C}_l$ is given by*

$$P_{e,l}(\mathfrak{I}) := \sup_{s^l \in \mathbf{S}^l} P_e(\mathfrak{C}_l, s^l), \tag{453}$$

where for $s^l \in \mathbf{S}^l$ we set

$$P_e(\mathfrak{C}_l, s^l) := \max_{i \in M_l} \left(1 - \mathrm{tr}(\mathcal{N}_{s^l}(\rho_i) D_i)\right). \tag{454}$$

The achievable rates and the classical deterministic capacity $C_{\mathrm{det,max}}(\mathfrak{I})$ of $\mathfrak{I}$, with respect to the error criterion given in (453), are then defined in the usual way.

The perhaps surprising result is that the strong subspace transmission capacity of a (finite) AVQC always equals its entanglement transmission capacity:

Theorem 80. *For every AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s \in \mathbf{S}}$ we have the equalities*

$$\mathcal{A}_{\mathrm{s,random}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{random}}(\mathfrak{I}), \tag{455}$$

$$\mathcal{A}_{\mathrm{s,det}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{det}}(\mathfrak{I}). \tag{456}$$

4.1.3 Zero-error capacities

In this subsection we only give definitions of zero-error capacities. Through the ideas of [1] these capacities are connected to arbitrarily varying channels, though this connection is not as strong as in the classical setting.
Results concerning these capacities are stated in subsections 4.8.2 and 4.8.3.

Definition 81. *An (l, k) zero-error quantum code (QC for short) $(\mathcal{F}, \mathcal{P}, \mathcal{R})$ for $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ consists of a Hilbert space $\mathcal{F}$, $\mathcal{P} \in \mathcal{C}(\mathcal{F}, \mathcal{H}^{\otimes l})$, $\mathcal{R} \in \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F})$ with $\dim \mathcal{F} = k$ such that*

$$\min_{x \in \mathcal{F}, ||x||=1} \langle x, \mathcal{R} \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}(|x\rangle\langle x|)x\rangle = 1. \tag{457}$$

For fixed block length $l \in \mathbb{N}$ define

$$k(l, \mathcal{N}) := \max\{\dim \mathcal{F} : \exists (l, k) \text{ zero-error QC for } \mathcal{N}\}. \tag{458}$$

The zero-error quantum capacity $Q_0(\mathcal{N})$ of $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ is then defined by

$$Q_0(\mathcal{N}) := \lim_{l \to \infty} \frac{1}{l} \log k(l, \mathcal{N}). \tag{459}$$

The existence of the limit follows from standard arguments based on Fekete's Lemma, which can be picked up in [31].

Next we pass to the zero-error classical capacities of quantum channels.

Definition 82. *Let $\sigma_{\mathcal{F}\mathcal{F}'}$ be a bipartite state on $\mathcal{F}\otimes\mathcal{F}'$ where $\mathcal{F}'$ denotes a unitary copy of the Hilbert space $\mathcal{F}$. An (l,M) entanglement assisted code (ea-code for short) $(\sigma_{\mathcal{F}\mathcal{F}'},\{\mathcal{P}_m,D_m\}_{m=1}^M)$ consists of a bipartite state $\sigma_{\mathcal{F}\mathcal{F}'}$, $\mathcal{P}_m\in\mathcal{C}(\mathcal{F},\mathcal{H}^{\otimes l})$, $m=1,\ldots,M$, and a POVM $\{D_m\}_{m=1}^M$ on $\mathcal{F}'\otimes\mathcal{K}^{\otimes l}$. A given (l,M) entanglement assisted code $(\sigma_{\mathcal{F}\mathcal{F}'},\{\mathcal{P}_m,D_m\}_{m=1}^M)$ is a zero-error code for $\mathcal{N}\in\mathcal{C}(\mathcal{H},\mathcal{K})$ if*

$$\mathrm{tr}((\mathcal{N}^{\otimes l}\circ\mathcal{P}_m\otimes id_{\mathcal{F}'})(\sigma_{\mathcal{F}\mathcal{F}'})D_m)=1 \tag{460}$$

holds for all $m\in[M]:=\{1,\ldots,M\}$. For $l\in\mathbb{N}$ we set

$$M_{EA}(l,\mathcal{N}):=\max\{M:\exists\ \ \textit{zero-error } (l,M)\ \textit{ea-code for } \mathcal{N}\}. \tag{461}$$

Definition 83. *The entanglement assisted classical zero-error capacity $C_{0EA}(\mathcal{N})$ of $\mathcal{N}\in\mathcal{C}(\mathcal{H},\mathcal{K})$ is given by*

$$C_{0EA}(\mathcal{N}):=\lim_{l\to\infty}\frac{1}{l}\log M_{EA}(l,\mathcal{N}). \tag{462}$$

If we restrict the definition of zero-error ea-code to states $\sigma_{\mathcal{F}\mathcal{F}'}$ with $\dim\mathcal{F}=\dim\mathcal{F}'=1$ we obtain the perfomance parameter $M(l,\mathcal{N})$ as a special case of $M_{\mathrm{EA}}(l,\mathcal{N})$ in (461) and the classical zero-error capacity $C_0(\mathcal{N})$ of a quantum channel $\mathcal{N}$.

Definition 84. *Given a bipartite state $\rho\in\mathcal{S}(\mathcal{H}_A\otimes\mathcal{H}_B)$. An (l,k_l) zero-error entanglement distillation protocol (EDP for short) for ρ consists of an LOCC operation $\mathcal{D}\in\mathcal{C}(\mathcal{H}_A^{\otimes l}\otimes\mathcal{H}_B^{\otimes l},\mathbb{C}^{k_l}\otimes\mathbb{C}^{k_l})$ and a maximally entangled state vector $\varphi_{k_l}=\frac{1}{\sqrt{k_l}}\sum_{i=1}^{k_l}e_i\otimes e_i\in\mathbb{C}^{k_l}\otimes\mathbb{C}^{k_l}$ with an orthonormal basis $\{e_1,\ldots,e_{k_l}\}$ of $\mathbb{C}^{k_l}$ such that*

$$\langle\varphi_{k_l},\mathcal{D}(\rho^{\otimes l})\varphi_{k_l}\rangle=1. \tag{463}$$

Let for $l\in\mathbb{N}$

$$d(l,\rho):=\max\{k_l:\exists(l,k_l)\ \textit{zero-error EDP for } \rho\}, \tag{464}$$

and we define the zero-error distillable entanglement of $\rho\in\mathcal{S}(\mathcal{H}_A\otimes\mathcal{H}_B)$ as

$$D_0(\rho):=\lim_{l\to\infty}\frac{1}{l}\log d(l,\rho). \tag{465}$$

4.1.4 Entanglement generation

Definition 85. *An entanglement-generation (l,k_l)-code for the AVQC $\mathfrak{I}=\{\mathcal{N}_s\}_{s\in\mathbf{S}}$ consists of a pair $(\mathcal{R}^l,\varphi_l)$ where $\mathcal{R}^l\in\mathcal{C}(\mathcal{K}^{\otimes l},\mathcal{F}_l)$ with $k_l=\dim\mathcal{F}_l$ and φ_l is a state on $\mathcal{F}_l\otimes\mathcal{H}^{\otimes l}$.*

Definition 86. *$R\in\mathbb{R}_+$ is called a deterministically achievable entanglement generation rate for $\mathfrak{I}$ if there is a sequence of (l,k_l) entanglement-generating codes with*

1. *$\liminf_{l\to\infty}\frac{1}{l}\log k_l\geq R$, and*
2. *$\lim_{l\to\infty}\inf_{s^l\in\mathbf{S}^l}F(|\psi_l\rangle\langle\psi_l|,(id_{\mathcal{F}_l}\otimes\mathcal{R}^l\circ\mathcal{N}_{s^l})(\varphi_l))=1$ where ψ_l denotes the standard maximally entangled state on $\mathcal{F}_l\otimes\mathcal{F}_l$ and $F(\cdot,\cdot)$ is the fidelity.*

Randomly achievable entanglement generation rates are defined in analogy to Definitions 72 and 77.

Definition 87. *The entanglement-generation capacities of $\mathfrak{I}$ are defined as*

$$\begin{aligned}G_{\mathrm{det}}(\mathfrak{I})&:=\sup\{R\in\mathbb{R}_+:R\ \textit{is deterministically achievable entanglement generation rate for}\ \mathfrak{I}\},\\ G_{\mathrm{random}}(\mathfrak{I})&:=\sup\{R\in\mathbb{R}_+:R\ \textit{is randomly achievable entanglement generation rate for}\ \mathfrak{I}\}.\end{aligned} \tag{466}$$

For these entanglement generation capacities, the following theorem holds.

Theorem 88. *Let $\mathfrak{I}\subset\mathcal{C}(\mathcal{H},\mathcal{K})$ be an AVQC. Then*

1. *$G_{\mathrm{random}}(\mathfrak{I})=\mathcal{A}_{\mathrm{random}}(\mathfrak{I})$*
2. *If $C_{\mathrm{det}}(\mathfrak{I})>0$ then $G_{\mathrm{det}}(\mathfrak{I})=G_{\mathrm{random}}(\mathfrak{I})$.*

4.2 Equivalence of strong subspace and entanglement transmission

We will now use results from section 3.9 to show, that strong subspace- and entanglement transmission are also equivalent criteria w.r.t. AVQCs.

Proof of Theorem 80. First we set, for every $l \in \mathbb{N}$, $\mathcal{I}^l := \mathbf{S}^l$. Assuming that $R > 0$ is an achievable rate for entanglement transmission over a *finite* AVQC $\mathfrak{I}$ (with random codes), we show that it is also an achievable strong subspace transmission rate (with random codes) for $\mathfrak{I}$. The proof does not depend on the form of the sequence of probability distributions assigned to the codes, so it applies to the case of deterministically achievable rates as well.
So, let there be a sequence of (l, k_l) random entanglement transmission codes with

$$\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R, \tag{467}$$

$$\min_{s^l \in \mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) = 1 - f_l, \text{ where } f_l \searrow 0. \tag{468}$$

Thus, there is $l' \in \mathbb{N}$ such that $k_l \geq 2^{l3R/4} + 1$ for all $l \geq l'$. Choose $\varepsilon_l = 2^{-lR/4}$ (this is just one of many possible choices). Obviously, since R is *strictly* greater than zero there is $\hat{l} \in \mathbb{N}$ such that

$$\begin{aligned} |\mathbf{S}|^l \sqrt{2/\pi} e^{-\varepsilon_l^2 (k_l - 1)/128} &\leq |\mathbf{S}|^l \sqrt{2/\pi} e^{-\varepsilon_l^2 2^{l3R/4}/128} && (469)\\ &= |\mathbf{S}|^l \sqrt{2/\pi} e^{-2^{lR/4}/128} && (470)\\ &< 1 && (471) \end{aligned}$$

holds for all $l \geq \hat{l}$. Application of Lemma 66 then yields a sequence of subspaces $\hat{\mathcal{F}}_l$ with dimensions $\hat{k}_l$ such that

$$\liminf_{l\to\infty} \frac{1}{l} \log \hat{k}_l = \liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R, \tag{472}$$

$$\min_{s^l \in \mathbf{S}^l} \min_{\phi \in S(\hat{\mathcal{F}}_l)} \int \langle \phi, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l(|\phi\rangle\langle\phi|)\phi\rangle d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \geq 1 - f_l - \frac{4 \cdot 12}{\sqrt{2(k_l - 1)}} - \frac{1}{l} \quad \forall\, l \geq \max\{l', \hat{l}\}. \tag{473}$$

Since the right hand side of (473) goes to zero for l going to infinity, we have shown that R is an achievable rate for strong subspace transmission (with random codes).
In case that $|\mathfrak{I}| = \infty$ holds we have to take care of some extra issues that arise from approximating $\mathfrak{I}$ by a finite AVQC. Such an approximation is carried out in detail in the proof of Lemma 102 or in that of Lemma 67 .
Now let $R = 0$ be an achievable rate for entanglement transmission with (random) codes. We show that it is achievable for strong subspace transmission by demonstrating that we can *always* achieve a strong subspace transmission rate of zero:
Choose any sequence $(|x_l\rangle\langle x_l|)_{l\in\mathbb{N}}$ of pure states such that $|x_l\rangle\langle x_l| \in \mathcal{S}(\mathcal{H}^{\otimes l})\ \forall l \in \mathbb{N}$. Set $\mathcal{F}_l := \mathbb{C} \cdot x_l$ $(l \in \mathbb{N})$. Define a sequence of recovery operations by $\mathcal{R}^l(a) := \mathrm{tr}(a) \cdot |x_l\rangle\langle x_l|$ $(a \in \mathcal{B}(\mathcal{K}^{\otimes l}),\ l \in \mathbb{N})$. Then $F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l}) = 1$ for all $l \in \mathbb{N},\ s^l \in \mathbf{S}^l$ and $\liminf_{l\to\infty} \frac{1}{l} \log(\dim \mathcal{F}_l) = 0$.

Now let $R \geq 0$ be an achievable rate for strong subspace transmission over some AVQC $\mathfrak{I}$ (with random codes). Thus, there exists a sequence of (random) strong subspace transmission codes with

$$\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R, \tag{474}$$

$$\inf_{s^l\in\mathbf{S}^l}\min_{\phi\in\mathcal{S}(\mathcal{F}_l)}\int\langle\phi,\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l(|\phi\rangle\langle\phi|)\phi\rangle d\mu_l(\mathcal{R}^l,\mathcal{P}^l)=1-f_l\ \forall l\in\mathbb{N},\ \text{where } f_l\searrow 0. \tag{475}$$

Now consider, for every $l\in\mathbb{N}$ and $s^l\in\mathbf{S}^l$, the channels $\int\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l d\mu_l(\mathcal{R}^l,\mathcal{P}^l)$. Then (475) implies that for these channels we have the estimate

$$\inf_{s^l\in\mathbf{S}^l}\min_{\phi\in\mathcal{S}(\mathcal{F}_l)}\langle\phi,\int\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l d\mu_l(\mathcal{R}^l,\mathcal{P}^l)(|\phi\rangle\langle\phi|)\phi\rangle=1-f_l, \tag{476}$$

and by a well-known result ([10], Theorem 2) we get

$$\inf_{s^l\in\mathbf{S}^l}F_e(\pi_{\mathcal{F}_l},\int\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l d\mu_l(\mathcal{R}^l,\mathcal{P}^l))\geq 1-\frac{3}{2}f_l, \tag{477}$$

which by convex-linearity of the entanglement fidelity in the channel implies

$$\inf_{s^l\in\mathbf{S}^l}\int F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l)d\mu_l(\mathcal{R}^l,\mathcal{P}^l)\geq 1-\frac{3}{2}f_l. \tag{478}$$

But $\lim_{l\to\infty}\frac{3}{2}f_l=0$ by assumption, implying that R is an achievable rate for entanglement transmission (with random codes) as well. □

4.3 Proof of the converse part

The basic technical obstacle we are faced with is that the converse part of the coding theorem for an AVQC cannot be reduced immediately to that of the single stationary memoryless quantum channel via Minimax Theorem (cf. [18] and [20]). In order to circumvent this problem we derive a relation between $\mathcal{A}_{\text{random}}(\mathfrak{I})$ and the corresponding random capacity of a suitable compound channel.
To be explicit, let us consider a finite AVQC $\mathfrak{I}=\{\mathcal{N}_s\}_{s\in\mathbf{S}}$ and let $(\mu_l)_{l\in\mathbb{N}}$ be a sequence of random $(l,k_l)-$codes for the AVQC $\mathfrak{I}$ with

$$\lim_{l\to\infty}\inf_{s^l\in\mathbf{S}^l}\int F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l)d\mu_l(\mathcal{P}^l,\mathcal{R}^l)=1. \tag{479}$$

On the other hand, for the infinite channel set $\text{conv}(\mathfrak{I})$, defined in (11), and each $\mathcal{N}_q\in\text{conv}(\mathfrak{I})$ we obtain

$$\begin{aligned}\int F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}_q^{\otimes l}\circ\mathcal{P}^l)d\mu_l(\mathcal{P}^l,\mathcal{R}^l) &= \sum_{s^l\in\mathbf{S}^l}\prod_{i=1}^{l}q(s_i)\int F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l)d\mu_l(\mathcal{P}^l,\mathcal{R}^l) &(480)\\ &\geq \inf_{s^l\in\mathbf{S}^l}\int F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l)d\mu_l(\mathcal{P}^l,\mathcal{R}^l). &(481)\end{aligned}$$

Consequently, (479) and (480) imply

$$\lim_{l\to\infty}\inf_{q\in\mathfrak{P}(\mathbf{S})}\int F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}_q^{\otimes l}\circ\mathcal{P}^l)d\mu_l(\mathcal{P}^l,\mathcal{R}^l)=1. \tag{482}$$

Defining the random entanglement transmission capacity $Q_{\text{comp, random}}(\text{conv}(\mathfrak{I}))$ for the *compound quantum channel* with uninformed users (see the definitions in subsection 3.1.3) built up from $\text{conv}(\mathfrak{I})$ in a similar fashion to $\mathcal{A}_{\text{random}}(\mathfrak{I})$ we can infer from the considerations presented above that

$$\mathcal{A}_{\text{random}}(\mathfrak{I})\leq Q_{\text{comp, random}}(\text{conv}(\mathfrak{I})). \tag{483}$$

Since the inequality $\mathcal{A}_{\text{det}}(\mathfrak{I})\leq\mathcal{A}_{\text{random}}(\mathfrak{I})$ is obvious, we obtain the following basic lemma.

Lemma 89. *Let $\mathfrak{I}=\{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be any finite set of channels and let $\text{conv}(\mathfrak{I})$ be the associated infinite set given in (11). Then*

$$\mathcal{A}_{\text{det}}(\mathfrak{I})\leq\mathcal{A}_{\text{random}}(\mathfrak{I})\leq Q_{\text{comp, random}}(\text{conv}(\mathfrak{I})). \tag{484}$$

Thus, our remaining task is to show that right-most capacity in (484) is upper bounded by the last term in (449). This is done in the following two subsections for finite and infinite AVQCs respectively.

4.3.1 Converse for the finite AVQC

First, we prove the converse to the coding theorem for finite compound quantum channels with random codes.

Theorem 90 (Converse Part: Compound Channel, $|\mathfrak{I}| < \infty$)**.** *Let $\mathfrak{I} = \{\mathcal{N}_1, \ldots, \mathcal{N}_N\} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a finite compound channel. The capacity $Q_{\text{comp, random}}(\mathfrak{I})$ of $\mathfrak{I}$ is bounded from above by*

$$Q_{\text{comp, random}}(\mathfrak{I}) \leq \lim_{l\to\infty} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{\mathcal{N}_i\in\mathfrak{I}} \frac{1}{l} I_c(\rho, \mathcal{N}_i^{\otimes l}). \tag{485}$$

Proof. Let for arbitrary $l \in \mathbb{N}$ a random (l, k_l) code for a compound channel $\mathfrak{I} = \{\mathcal{N}_1, \ldots, \mathcal{N}_N\}$ with the property

$$\min_{1\leq i\leq N} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) \geq 1 - \varepsilon_l \tag{486}$$

be given, where $\varepsilon_l \in [0, 1]$ and $\lim_{l\to\infty} \varepsilon_l = 0$. Obviously, the above code then satisfies

$$\begin{aligned} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \frac{1}{N}\sum_{i=1}^N \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) &= \frac{1}{N}\sum_{i=1}^N \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) & (487)\\ &\geq 1 - \varepsilon_l. & (488) \end{aligned}$$

This implies the existence of at least one pair $(\mathcal{R}^l, \mathcal{P}^l)$ such that

$$F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \frac{1}{N}\sum_{i=1}^N \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) \geq 1 - \varepsilon_l, \tag{489}$$

hence for all $i = 1, \ldots, N$

$$F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) \geq 1 - N\varepsilon_l. \tag{490}$$

The rest of the proof is identical to that of Theorem 52. □

Using the approximation techniques developed in the previous (sub)sections, we will now prove the converse for random codes and general compound channels.

Theorem 91 (Converse Part: Compound Channel)**.** *Let $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be an arbitrary compound quantum channel. The capacity $Q_{\text{comp, random}}(\mathfrak{I})$ of $\mathfrak{I}$ is bounded from above by*

$$Q_{\text{comp, random}}(\mathfrak{I}) \leq \lim_{l\to\infty} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} \frac{1}{l} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{491}$$

Proof of Theorem 91. Let a sequence $(l, k_l)_{l\in\mathbb{N}}$ of random codes for $\mathfrak{I}$ be given such that

- $\liminf_{l\to\infty} \frac{1}{l} \log \dim \mathcal{F}_l = R$
- $\inf_{\mathcal{N}\in\mathfrak{I}} \int F_e(\mathcal{F}_l, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) = 1 - \varepsilon_l,$

where the sequence $(\varepsilon_l)_{l\in\mathbb{N}}$ satisfies $\lim_{l\to\infty} \varepsilon_l = 0$. Let, for some $\tau > 0$, $\bigcup_{\mathcal{N}\in\mathfrak{I}} B_\diamondsuit(\mathcal{N}, \tau)$ be an open cover for $\mathfrak{I}$. Clearly, it also covers the compact set $\bar{\mathfrak{I}}$. Thus, there exist finitely many channels $\mathcal{N}_1, \ldots, \mathcal{N}_{M_\tau}$ such that $\bigcup_{i=1}^{M_\tau} B_\diamondsuit(\mathcal{N}_i, \tau) \supset \bar{\mathfrak{I}}$ and, therefore, $\mathcal{M}_\tau := \{\mathcal{N}_1, \ldots, \mathcal{N}_{M_\tau}\}$ is a τ-net for $\mathfrak{I}$.
By $\mathcal{M}_\tau \subset \mathfrak{I}$ we get, for every $\tau > 0$, the following result:

- $\liminf_{l\to\infty} \frac{1}{l} \log \dim \mathcal{F}_l = R$
- $\min_{\mathcal{N}_i\in\mathcal{M}_\tau} \int F_e(\mathcal{F}_l, \mathcal{R}^l \circ \mathcal{N}_i^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) \geq 1 - \varepsilon_l.$

By Theorem 90, this immediately implies

$$R \leq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{\mathcal{N}_i\in\mathcal{M}_\tau} I_c(\rho, \mathcal{N}_i^{\otimes l}). \tag{492}$$

From Lemma 55 we get, by noting that $D_\diamondsuit(\mathfrak{I}, \mathcal{M}_\tau) \leq \tau$ the estimate

$$R \leq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathfrak{I}} I_c(\rho, \mathcal{N}^{\otimes l}) + \nu(2\tau). \tag{493}$$

Taking the limit $\tau \to 0$ proves the theorem. □

Theorem 92 (Converse: finite AVQC)**.** *Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be a finite AVQC. Then*

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) \leq Q_{\text{comp, random}}(\text{conv}(\mathfrak{I})) = \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{494}$$

Proof. Just combine Lemma 89 and Theorem 91 applied to $\text{conv}(\mathfrak{I})$. □

4.3.2 Case $|\mathfrak{I}| = \infty$

The proof of the converse part of Theorem 76 requires just a bit of additional work. Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an arbitrary AVQC and let $\mathfrak{P}_{\text{fin}}(S)$ denote the set of probability distributions on S with *finite* support. Then

$$\text{conv}(\mathfrak{I}) = \left\{ \mathcal{N}_q \in \mathcal{C}(\mathcal{H}, \mathcal{K}) : \mathcal{N}_q := \sum_{s\in\mathbf{S}} q(s)\mathcal{N}_s, \text{ and } q \in \mathfrak{P}_{\text{fin}}(S) \right\}. \tag{495}$$

The argument that led us to the inequality (480) accompanied by the continuity of the entanglement fidelity with respect to $||\cdot||_\diamondsuit$ and an application of the dominated convergence theorem show that for each $\mathcal{N} \in \overline{\text{conv}(\mathfrak{I})}$

$$\int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) \geq \inf_{s^l\in\mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) \tag{496}$$

holds. Then Lemma 89 holds *mutatis mutandis* with $\text{conv}(\mathfrak{I})$ replaced by $\overline{\text{conv}(\mathfrak{I})}$. Additionally, if we apply Theorem 91 to $\overline{\text{conv}(\mathfrak{I})}$ we are led to the following theorem.

Theorem 93 (Converse: general AVC)**.** *Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an arbitrary AVQC. Then*

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) \leq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{\mathcal{N}\in\overline{\text{conv}(\mathfrak{I})}} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{497}$$

4.4 Achievability of entanglement transmission rate I: Random codes

We show in this section how the achievability results for compound quantum channels obtained in sections 3.3 and 3.5 imply existence of reliable random codes for AVQC via Ahlswede's robustification technique [3].

Let $l \in \mathbb{N}$ and let Perm_l denote the set of permutations acting on $\{1, \ldots, l\}$. Let us further suppose that we are given a finite set $\mathbf{S}$. Then each permutation $\sigma \in \text{Perm}_l$ induces a natural action on $\mathbf{S}^l$ by $\sigma : \mathbf{S}^l \to \mathbf{S}^l$, $\sigma(s^l)_i := s_{\sigma(i)}$. Moreover, let $T(l, \mathbf{S})$ denote the set of types on $\mathbf{S}$ induced by the elements of $\mathbf{S}^l$, i.e. the set of empirical distributions on $\mathbf{S}$ generated by sequences in $\mathbf{S}^l$. Then Ahlswede's robustification can be stated as follows.

Theorem 94 (Robustification technique, cf. Theorem 6 in [3]). *If a function $f : \mathbf{S}^l \to [0,1]$ satisfies*

$$\sum_{s^l \in \mathbf{S}^l} f(s^l) q(s_1) \cdot \ldots \cdot q(s_l) \geq 1 - \gamma \tag{498}$$

for all $q \in T(l, \mathbf{S})$ and some $\gamma \in [0,1]$, then

$$\frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} f(\sigma(s^l)) \geq 1 - (l+1)^{|\mathbf{S}|} \cdot \gamma \qquad \forall s^l \in \mathbf{S}^l. \tag{499}$$

Remark 95. *Ahlswede's original approach in [3] gives*

$$\frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} f(\sigma(s^l)) \geq 1 - 3 \cdot (l+1)^{|\mathbf{S}|} \cdot \sqrt{\gamma} \qquad \forall s^l \in \mathbf{S}^l. \tag{500}$$

The better bound (499) is from [4].

Proof. Because the result of Theorem 94 is a central tool in this work and the proof given in [4] is particularly simple we reproduce it here in full for reader's convenience.
Notice first that (498) is equivalent to

$$\sum_{s^l \in \mathbf{S}^l} (1 - f(s^l)) q(s_1) \cdot \ldots \cdot q(s_l) \leq \gamma \qquad \forall q \in T(l, \mathbf{S}), \tag{501}$$

which in turn is equivalent to

$$\sum_{s^l \in \mathbf{S}^l} (1 - f(\sigma(s^l))) q(s_{\sigma(1)}) \cdot \ldots \cdot q(s_{\sigma(l)}) \leq \gamma \qquad \forall q \in T(l, \mathbf{S}), \tag{502}$$

and $\sigma \in \mathrm{Perm}_l$, since σ is bijective. Clearly, we have

$$q(s_{\sigma(1)}) \cdot \ldots \cdot q(s_{\sigma(l)}) = q(s_1) \cdot \ldots \cdot q(s_l) \qquad \forall \sigma \in \mathrm{Perm}_l, \forall s^l \in \mathbf{S}^l, \tag{503}$$

and therefore, we obtain

$$\sum_{s^l \in \mathbf{S}^l} \left(1 - \frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} f(\sigma(s^l)) \right) q(s_1) \cdot \ldots \cdot q(s_l) \leq \gamma \qquad \forall q \in T(l, \mathbf{S}). \tag{504}$$

Now, for $q \in T(l, \mathbf{S})$ let $T_q^l \subset \mathbf{S}^l$ denote the set of sequences whose empirical distribution is q. Since f takes values in $[0,1]$ we have $1 - \frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} f(\sigma(s^l)) \geq 0$ and thus from (504)

$$\sum_{s^l \in T_q^l} \left(1 - \frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} f(\sigma(s^l)) \right) q(s_1) \cdot \ldots \cdot q(s_l) \leq \gamma \qquad \forall q \in T(l, \mathbf{S}). \tag{505}$$

It is clear from definition that for each $s^l \in T_q^l$ we have $\bigcup_{\sigma \in \mathrm{Perm}_l} \{\sigma(s^l)\} = T_q^l$ and, consequently, $\sum_{\sigma \in \mathrm{Perm}_l} f(\sigma(s^l))$ does not depend on $s^l \in T_q^l$. Therefore, from (505) we obtain

$$\left(1 - \frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} f(\sigma(s^l)) \right) q^{\otimes l}(T_q^l) \leq \gamma \qquad \forall q \in T(l, \mathbf{S}),\ \forall s^l \in T_q^l. \tag{506}$$

On the other hand

$$q^{\otimes l}(T_q^l) \geq \frac{1}{(l+1)^{|\mathbf{S}|}} \qquad \forall q \in T(l, \mathbf{S}) \tag{507}$$

holds (cf. [20] page 30), which, by (506), implies

$$\left(1-\frac{1}{l!}\sum_{\sigma\in\mathrm{Perm}_l} f(\sigma(s^l))\right)\leq (l+1)^{|\mathbf{S}|}\cdot\gamma \qquad \forall q\in T(l,\mathbf{S}),\ \forall s^l\in T_q^l. \tag{508}$$

This is the inequality we aimed to prove since $\mathbf{S}^l=\bigcup_{q\in T(l,\mathbf{S})} T_q^l$. □

The function f appearing in Theorem 94 will be built up from the entanglement fidelities of the channels constituting a finite AVQC that approximates our AVQC $\mathfrak{I}=\{\mathcal{N}_s\}_{s\in\mathbf{S}}$.
As another ingredient for the arguments to follow we need an achievability result for compound channels.

Lemma 96. *Let $k\in\mathbb{N}$ and $\mathfrak{T}\subset\mathcal{C}(\mathcal{H},\mathcal{K})$. For each $\eta>0$ there is a sequence of (l,k_l)-codes $(\mathcal{P}^l,\mathcal{R}^l)_{l\in\mathbb{N}}$ and an $l_0(\eta)\in\mathbb{N})$ such that for all $l\geq l_0(\eta)$*

$$F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}^{\otimes l}\circ\mathcal{P}^l)\geq 1-2^{-lc} \qquad \forall\mathcal{N}\in\mathfrak{T}, \tag{509}$$

and

$$\frac{1}{l}\log\dim\mathcal{F}_l\geq\frac{1}{k}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes k})}\inf_{\mathcal{N}\in\mathfrak{T}} I_c(\rho,\mathcal{N}^{\otimes k})-\eta, \tag{510}$$

hold with a real constant $c=c(k,\dim\mathcal{H},\dim\mathcal{K},\mathfrak{T},\eta)>0$.

Remark 97. *Lemma 96 is a strengthening of Theorem 44 insofar as it explicitly points out the following: For any compound channel, at any rate below its capacity for transmission of entanglement, there exist sequences of codes such that entanglement fidelity goes to one exponentially fast.*
The importance of this result for the investigations at hand can be understood by looking at equation (528) in Theorem 98.

Proof. We will give the details for the case $k=1$ only. The proof for arbitrary k follows by an almost identical argument.
According to the compound BSST Lemma (cf. [15], Lemma 6.1) to any $\eta>0$ we can find $m=m(\mathfrak{T},\eta)\in\mathbb{N}$ and a subspace $\mathcal{G}\subset\mathcal{H}^{\otimes m}$ such that

$$\frac{1}{m}\inf_{\mathcal{N}\in\mathfrak{T}} I_c(\pi_{\mathcal{G}},\mathcal{N}^{\otimes m})\geq\max_{\rho\in\mathcal{S}(\mathcal{H})}\inf_{\mathcal{N}\in\mathfrak{T}} I_c(\rho,\mathcal{N})-\frac{\eta}{3}. \tag{511}$$

Explicitly stated, $\mathcal{G}$ is the eigenspace to eigenvalue one of a frequency typical projection of the maximizer on the r.h.s. of (511) with suitably chosen parameters.
Let us consider the compound quantum channel built up from $\{\mathcal{N}^{\otimes m}:\mathcal{N}\in\mathfrak{T}\}$.

Looking at the last equation in the proof of Lemma 43 we see, that for this channel, for large enough $t\in\mathbb{N}$, there exists a sequence of (t,k_t)-codes $(\bar{\mathcal{P}}^t,\bar{\mathcal{R}}^t)_{t\in\mathbb{N}}$, $\bar{\mathcal{P}}^t\in\mathcal{C}(\mathcal{F}_t,\mathcal{H}^{\otimes mt})$, $\bar{\mathcal{R}}^t\in\mathcal{C}(\mathcal{K}^{\otimes mt},\mathcal{F}_t')$ with

$$\inf_{\mathcal{N}\in\mathfrak{T}} F_e(\pi_{\mathcal{F}_t},\mathcal{R}^t\circ\mathcal{N}^{\otimes mt}\circ\mathcal{P}^t)\geq 1-N_{\tau_t}\varepsilon_t-t\tau_t \tag{512}$$

holds. Here, $\varepsilon_t\leq 2^{-tc_1}$ for some $c_1>0$ (compare equation (216) in the proof of Theorem 38), while $N_{\tau_t}\leq(3/\tau_t)^{2(d\cdot d')^2}$, where $d=\dim\mathcal{H}$ and $d'=\dim\mathcal{K}$. Although the proof of Lemma 9 uses a subexponential growth of N_{τ_t}, this is not at all necessary. Set

$$c_2:=c_1/(2\cdot d\cdot d')^2, \qquad\qquad \tau_t:=2^{-tc_2}\quad(t\in\mathbb{N}). \tag{513}$$

Then

$$\inf_{\mathcal{N}\in\mathfrak{T}} F_e(\pi_{\mathcal{F}_t},\mathcal{R}^t\circ\mathcal{N}^{\otimes mt}\circ\mathcal{P}^t)\geq 1-3^{2(d\cdot d')^2}2t\cdot 2^{-tc_1/2} \tag{514}$$

and thus, defining $c' := c_1/4$, we know that for each $\eta > 0$ there exists $t(\eta) \in \mathbb{N}$ such that

$$\inf_{\mathcal{N}\in\mathfrak{T}} F_e(\pi_{\mathcal{F}_t}, \bar{\mathcal{R}}^t \circ \mathcal{N}^{\otimes mt} \circ \bar{\mathcal{P}}^t) \geq 1 - 2^{-tc'} \qquad \forall t \geq t(\eta), \tag{515}$$

as well as

$$\frac{1}{t}\log k_t = \frac{1}{t}\log\dim\mathcal{F}_t \geq (\inf_{\mathcal{N}\in\mathfrak{T}} I_c(\pi_{\mathcal{G}}, \mathcal{N}^{\otimes m}) - \frac{\eta}{3})\cdot m \qquad \forall t \geq t(\eta) \tag{516}$$

where, clearly, $c' = c'(\dim\mathcal{H}, \dim\mathcal{K}, \mathfrak{T}, \eta)$.
For $t, l \in \mathbb{N}$ let $r \in \{0, 1, \ldots, m-1\}$ be the unique non-negative integer such that $l = mt + r$. Furthermore, let us choose for each $r \in \{0, 1, \ldots, m-1\}$ a state vector $x_r \in \mathcal{H}^{\otimes r}$ and set

$$\mathcal{F}_l := \mathcal{F}_t \otimes \mathbb{C}\cdot\{x_r\}. \tag{517}$$

Then

$$\pi_{\mathcal{F}_l} = \pi_{\mathcal{F}_t} \otimes |x_r\rangle\langle x_r|. \tag{518}$$

Moreover we set

$$\mathcal{P}^l := \bar{\mathcal{P}}^t \otimes id_{\mathcal{B}(\mathcal{H}^{\otimes r})} \quad \text{and} \quad \mathcal{R}^l := \bar{\mathcal{R}}^t \otimes T^r, \tag{519}$$

where $T^r \in \mathcal{C}(\mathcal{K}^{\otimes r}, \mathcal{H}^{\otimes r})$ is given by $T^r(a) := \mathrm{tr}(a)|x_r\rangle\langle x_r|$. Then it is clear that

$$\begin{aligned} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l) &= F_e(\pi_{\mathcal{F}_t}, \bar{\mathcal{R}}^t \circ \mathcal{N}^{\otimes mt} \circ \bar{\mathcal{P}}^t) && (520)\\ &\geq 1 - 2^{-tc'} && (521)\\ &= 1 - 2^{-\frac{l-r}{m}c'} && (522)\\ &\geq 1 - 2^{-lc} \qquad \forall\mathcal{N}\in\mathfrak{T} && (523)\end{aligned}$$

for all $l \geq l_1(\eta)$ with $c := \frac{c'}{2m}$, and where in the second line we have used (515).
On the other hand, from equations (511), (516) and (517) we obtain for $t \geq t(\eta)$

$$\begin{aligned} \frac{1}{l}\log\dim\mathcal{F}_l &= \frac{1}{tm+r}\log\dim\mathcal{F}_t && (524)\\ &\geq \frac{1}{1+\frac{r}{tm}}(\max_{\rho\in\mathcal{S}(\mathcal{H})}\inf_{\mathcal{N}\in\mathfrak{T}} I_c(\rho,\mathcal{N}) - \frac{\eta}{3} - \frac{\eta}{3m}) && (525)\\ &\geq \max_{\rho\in\mathcal{S}(\mathcal{H})}\inf_{\mathcal{N}\in\mathfrak{T}} I_c(\rho,\mathcal{N}) - \eta && (526)\end{aligned}$$

if t and consequently l is sufficiently large. Therefore there is an $l_0(\eta) \in \mathbb{N}$ such that (520) and (524) hold simultaneously for all $l \geq l_0(\eta)$ which concludes the proof in the case $k = 1$. □

In the next step we will combine the robustification technique and Lemma 96 to prove the existence of good random codes for the AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$.
Recall that there is a canonical action of Perm_l on $\mathcal{B}(\mathcal{H})^{\otimes l}$ given by $A_{\sigma,\mathcal{H}}(a_1 \otimes \ldots \otimes a_l) := a_{\sigma^{-1}(1)} \otimes \ldots \otimes a_{\sigma^{-1}(l)}$. It is easy to see that $A_{\sigma,\mathcal{H}}(a) = U_\sigma a U_\sigma^*$, $(a \in \mathcal{B}(\mathcal{H})^{\otimes l})$ with the unitary operator $U_\sigma : \mathcal{H}^{\otimes l} \to \mathcal{H}^{\otimes l}$ defined by $U_\sigma(x_1 \otimes \ldots \otimes x_l) = x_{\sigma^{-1}(1)} \otimes \ldots \otimes x_{\sigma^{-1}(l)}$.

Theorem 98 (Conversion of compound codes)**.** *Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an AVQC. For each $k \in \mathbb{N}$ and any sufficiently small $\eta > 0$ there is a sequence of codes $(\mathcal{P}^l, \mathcal{R}^l)_{l\in\mathbb{N}}$, $\mathcal{P}^l \in \mathcal{C}(\mathcal{F}_l, \mathcal{H}^{\otimes l}), \mathcal{R}^l \in \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}'_l)$, for the compound channel built up from* $\mathrm{conv}(\mathfrak{I})$ *(cf. (11)) satisfying*

$$\frac{1}{l}\log\dim\mathcal{F}_l \geq \frac{1}{k}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes k})}\inf_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho,\mathcal{N}^{\otimes k}) - 2\cdot\nu(8\eta), \tag{527}$$

$$\frac{1}{l!}\sum_{\sigma\in \mathrm{Perm}_l} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ A_{\sigma^{-1},\mathcal{K}} \circ \mathcal{N}_{s^l} \circ A_{\sigma,\mathcal{H}} \circ \mathcal{P}^l) \geq 1-(l+1)^{N_\eta}\cdot 2^{-lc} \qquad \forall s^l \in \mathbf{S}^l \tag{528}$$

for all sufficiently large l with a positive number $c = c(k, \dim\mathcal{H}, \dim\mathcal{K}, \mathrm{conv}(\mathfrak{I}), \eta)$, $\nu : [0,1] \to \mathbb{R}$ defined by $\nu(x) := x + 8x\log(d_\mathcal{K}) + 4\cdot h(x)$ ($h(\cdot)$ being the binary entropy) and an integer N_η which depends on the set $\mathfrak{I}$ as well.

The idea of the proof is the following. We want to approximate the set $\mathrm{conv}(\mathfrak{I})$ from the outside by using a polytope P_η with N_η extreme points. Then, our results for compound codes and an application of the robustification technique yield a sequence of codes which have asymptotically optimal performance for the AVQC P_η. Since $\mathrm{conv}(\mathfrak{I}) \subset P_\eta$, they will also have asymptotically optimal performance for $\mathfrak{I}$.
A problem occurs if $\mathrm{conv}(\mathfrak{I})$ touches the boundary of the set of quantum channels because parts of that boundary are curved and the approximating polytope P_η may contain maps that are not channels. Therefore, an intermediate step consists of slightly moving $\mathrm{conv}(\mathfrak{I})$ away from the boundary. This may be seen as application of a completely positive map and can therefore be absorbed into the recovery operation. During the proof we are going to make use of the following Lemma, that will be proven first:

Lemma 99. *Let A, B be compact convex sets in $\mathbb{C}^n$ with $A \subset B$ and*

$$d(\mathrm{rebd}\, B, A) := \inf\{||b-a|| : b \in \mathrm{rebd}\, B, a \in A\} = t > 0, \tag{529}$$

where $||\cdot||$ denotes any norm.
Let $P \supset A$ be a polytope with $D(A,P) \leq \delta$, where $\delta \in (0,t]$ and D is the Hausdorff distance induced by $||\cdot||$. Then $P' := P \cap \mathrm{aff}\, B$ is a polytope and $P' \subset B$.

Proof of Lemma 99. The assertion that P' is a polytope is clear. Suppose $\exists p \in P'\backslash B$. Then since $D(A,P) \leq \delta$ we have $P \subset (A)_\delta$ (cf. [68], Theorem 2.7.3). But this means, since $P' \subset P$, that to our $p \in P'\backslash B$ we can find $a_\delta \in A$ with

$$||p - a_\delta|| \leq \delta. \tag{530}$$

For $\lambda \in [0,1]$ define

$$x_\lambda := (1-\lambda)a_\delta + \lambda p. \tag{531}$$

Then there is $\lambda^* \in (0,1)$ such that

$$x := x^*_\lambda \quad \in \quad \mathrm{rebd}\, B. \tag{532}$$

This is seen as follows: Since $d(\mathrm{rebd}\, B, A) = t > 0$ we have $A \subset \mathrm{ri}\, B$. Set

$$L := \{\lambda \in (0,1] : (1-\lambda)a_\delta + \lambda p \in B\}. \tag{533}$$

From $a_\delta \in \mathrm{ri}\, B$ it follows that $L \neq \emptyset$ and from the fact that B is compact and convex we then get that $L = (0, \lambda^*]$. Now,

$$\begin{aligned} ||x - a_\delta|| &= ||(1-\lambda^*)a_\delta + \lambda^* p - (1-\lambda^*)a_\delta - \lambda^* a_\delta|| && (534)\\ &= \lambda^* ||p - a_\delta|| && (535)\\ &\leq \lambda^* \cdot \delta && (536)\\ &< t, && (537)\end{aligned}$$

where the last line follows from $\lambda \in (0,1)$. This is a contradiction to $d(\mathrm{rebd}\, B, A) = t$. □

Proof of Theorem 98. We can suppose that

$$\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes k})} \inf_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes k}) > 0, \tag{538}$$

because otherwise our claim is obviously true. We will further assume that $\mathfrak{I}$, and therefore $\mathrm{conv}(\mathfrak{I})$ as well, is compact. Since the Hausdorff-distance of $\mathrm{conv}(\mathfrak{I})$ to its closure (in $||\cdot||_\diamond$) is zero, this does not change the left hand side of equation (538), due to the estimates in Lemma 55. Since $\mathfrak{I}$ is a subset of its norm-closure, good codes for the norm-closure will also work for $\mathfrak{I}$. Thus, our assumption is a pure technicality and, indeed, without loss of generality.
Now let us, for $\varepsilon \leq 1$, by $\mathfrak{D}_\varepsilon$ denote the operation $\mathfrak{D}_\varepsilon(\cdot) := (1-\varepsilon)\mathrm{id}_{\mathcal{B}(\mathcal{K})}(\cdot) + \frac{\varepsilon}{\dim \mathcal{K}} \mathbf{1}_{\mathcal{K}} \mathrm{tr}(\cdot)$. If $\varepsilon \geq 0$, this is nothing but a depolarizing channel.
By Lemma 2.3.3 in [68] and since $\mathfrak{D}_1 \circ \mathcal{N} \notin \mathrm{rebd}\, \mathcal{C}(\mathcal{H}, \mathcal{K})$ for arbitrary $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ and $\eta > 0$, we have

$$\mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I})) \subset \mathrm{ri}\, \mathcal{C}(\mathcal{H}, \mathcal{K}). \tag{539}$$

Since $\mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I}))$ is compact, we know that

$$c' := \min\{||\mathcal{N} - \mathcal{N}'||_\diamond : \mathcal{N} \in \mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I})), \mathcal{N}' \in \mathrm{rebd}\, \mathcal{C}(\mathcal{H}, \mathcal{K})\} \tag{540}$$

satisfies $c' > 0$. Thus, by Lemma 99 and Theorem 3.1.6 in [68] there exists a polytope $P_\eta \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ such that $\mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I})) \subset P_\eta$ and

$$D_\diamond(\mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I})), P_\eta) \leq 2\eta. \tag{541}$$

The set of extremal points of P_η we denote by $\mathrm{ext}(P_\eta) = \{\mathcal{N}_e\}_{e \in E_\eta}$, where E_η is a finite set indexing the extremal points, the number of which we label N_η. Consider the compound quantum channel P_η. It follows from Lemma 96 that there exists a sequence of (l, k_l)-codes $(\mathcal{P}^l, \mathcal{R}^l)_{l \in \mathbb{N}}$ such that for all $l \geq l_0(\eta)$

$$F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}^{\otimes l} \circ \mathcal{P}^l) \geq 1 - 2^{-lc} \qquad \forall \mathcal{N} \in P_\eta, \tag{542}$$

and

$$\frac{1}{l} \log \dim \mathcal{F}_l \geq \frac{1}{k} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes k})} \inf_{\mathcal{N} \in P_\eta} I_c(\rho, \mathcal{N}^{\otimes k}) - \eta, \tag{543}$$

with a positive number $c = c(k, \dim \mathcal{H}, \dim \mathcal{K}, \mathrm{conv}(\mathfrak{I}), \eta)$.
Let us define $f : E_\eta^l \to [0,1]$ by

$$f(e^l) := F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{e^l} \circ \mathcal{P}^l). \tag{544}$$

Then (542) implies that

$$\sum_{e^l \in E_\eta^l} f(e^l) q(e_1) \cdot \ldots \cdot q(e_l) \geq 1 - 2^{-lc} \qquad \forall q \in T(l, E_\eta). \tag{545}$$

But (545) and Theorem 94 yield

$$\frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{\sigma(e^l)} \circ \mathcal{P}^l) \geq 1 - (l+1)^{N_\eta} \cdot 2^{-lc} \qquad \forall e^l \in E_\eta^l. \tag{546}$$

By (543) and (546) we are guaranteed the existence of a good random code for P_η if we can somehow consider permutations as part of the encoding and recovery procedure. More precisely, we will now show that

$$\mathcal{N}_{\sigma(e^l)} = A_{\sigma^{-1}, \mathcal{K}} \circ \mathcal{N}_{e^l} \circ A_{\sigma, \mathcal{H}} \qquad \forall e^l \in E_\eta^l. \tag{547}$$

To this end, let $\psi = \psi_1 \otimes \ldots \otimes \psi_l$, $\varphi = \varphi_1 \otimes \ldots \otimes \varphi_l \in \mathcal{H}^{\otimes l}$. Then

$$\begin{aligned}
A_{\sigma^{-1}, \mathcal{K}} \circ \mathcal{N}_{e^l} \circ A_{\sigma, \mathcal{H}}(|\psi\rangle\langle\varphi|) &= (A_{\sigma^{-1}, \mathcal{K}} \circ \mathcal{N}_{e^l})(|\psi_{\sigma^{-1}(1)}\rangle\langle\varphi_{\sigma^{-1}(1)}| \otimes \ldots \otimes |\psi_{\sigma^{-1}(l)}\rangle\langle\varphi_{\sigma^{-1}(l)}|) && (548)\\
&= A_{\sigma^{-1}, \mathcal{K}}(\otimes_{i=1}^l \mathcal{N}_{s_i}(|\psi_{\sigma^{-1}(i)}\rangle\langle\varphi_{\sigma^{-1}(i)}|)) && (549)\\
&= \otimes_{i=1}^l \mathcal{N}_{s_{\sigma(i)}}(|\psi_i\rangle\langle\varphi_i|) && (550)\\
&= \mathcal{N}_{\sigma(e^l)}(\otimes_{i=1}^l |\psi_i\rangle\langle\varphi_i|) && (551)\\
&= \mathcal{N}_{\sigma(e^l)}(|\psi\rangle\langle\varphi|). && (552)
\end{aligned}$$

Therefore,

$$A_{\sigma^{-1},\mathcal{K}} \circ \mathcal{N}_{e^l} \circ A_{\sigma,\mathcal{H}} = \mathcal{N}_{\sigma(e^l)}. \tag{553}$$

By construction of P_η we know that for every $\mathcal{N}_s \in \mathfrak{I}$ there exists a probability distribution $q(\cdot|s) \in \mathfrak{P}(E_\eta)$ such that

$$\mathfrak{D}_\eta \circ \mathcal{N}_s = \sum_{e \in E_\eta} q(e|s)\mathcal{N}_e \tag{554}$$

holds. We define

$$\tilde{\mathcal{R}}^l_\sigma := \mathcal{R}^l \circ A_{\sigma^{-1},\mathcal{K}} \circ \mathfrak{D}_\eta^{\otimes l}, \qquad \tilde{\mathcal{P}}^l_\sigma := A_{\sigma,\mathcal{H}} \circ \mathcal{P}^l. \tag{555}$$

Combining the equations (546),(547),(554),(555) we get for every $s^l \in \mathbf{S}^l$:

$$\begin{aligned}
\sum_{\sigma \in \mathrm{Perm}_l} F_e(\pi_{\mathcal{F}_l}, \tilde{\mathcal{R}}^l_\sigma \circ \mathcal{N}_{s^l} \circ \tilde{\mathcal{P}}^l_\sigma) &= \sum_{\sigma \in \mathrm{Perm}_l} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ A_{\sigma^{-1},\mathcal{K}} \circ \mathfrak{D}_\eta^{\otimes l} \circ \mathcal{N}_{s^l} \circ A_{\sigma,\mathcal{H}} \circ \mathcal{P}^l) && (556)\\
&= \sum_{\sigma \in \mathrm{Perm}_l} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ A_{\sigma^{-1},\mathcal{K}} \circ \sum_{e^l \in E_\eta^l} \prod_{i=1}^{l} q(e_i|s_i)\mathcal{N}_{e^l} \circ A_{\sigma,\mathcal{H}} \circ \mathcal{P}^l)\\
&= \sum_{e^l \in E_\eta^l} \prod_{i=1}^{l} q(e_i|s_i) \sum_{\sigma \in \mathrm{Perm}_l} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ A_{\sigma^{-1},\mathcal{K}} \circ \mathcal{N}_{e^l} \circ A_{\sigma,\mathcal{H}} \circ \mathcal{P}^l)\\
&= \sum_{e^l \in E_\eta^l} \prod_{i=1}^{l} q(e_i|s_i) \sum_{\sigma \in \mathrm{Perm}_l} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{\sigma(e^l)} \circ \mathcal{P}^l) && (557)\\
&\geq (l!)(1 - (l+1)^{N_\eta} \cdot 2^{-lc}) && (558)
\end{aligned}$$

Now, defining a discretely supported probability measure μ_l, $l \in \mathbb{N}$ by

$$\mu_l := \frac{1}{l!} \sum_{\sigma \in \mathrm{Perm}_l} \delta_{(\tilde{\mathcal{R}}^l_\sigma, \tilde{\mathcal{P}}^l_\sigma)}, \tag{559}$$

where $\delta_{(\tilde{\mathcal{R}}^l_\sigma, \tilde{\mathcal{P}}^l_\sigma)}$ denotes the probability measure that puts measure 1 on the point $(\tilde{\mathcal{R}}^l_\sigma, \mathcal{P}^l_\sigma)$, we obtain for each $k \in \mathbb{N}$ a sequence of (l, k_l)-random codes for $\mathfrak{I}$ achieving

$$\frac{1}{k} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes k})} \inf_{\mathcal{N} \in P_\eta} I_c(\rho, \mathcal{N}^{\otimes k}) - \eta. \tag{560}$$

It remains to show that this last number is close to (527). This in turn is true mostly because, by construction, $D_\diamondsuit(P_\eta, \mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I}))) \leq 2\eta$ holds and, as will be shown, $D_\diamondsuit(\mathrm{conv}(\mathfrak{I}), \mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I}))) \leq 2\eta$ holds.

We start with the upper bound on $D_\diamondsuit(\mathrm{conv}(\mathfrak{I}), \mathfrak{D}_\eta(\mathrm{conv}(\mathfrak{I})))$, which will be derived in a slightly more general way. For arbitrary $s \leq 1$ and a compact $A \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$

$$D_\diamondsuit(\mathfrak{D}_s(A), A) \leq |s| \cdot \max_{x \in A} ||x - \mathfrak{D}_1 \circ x|| \leq 2|s| \tag{561}$$

holds, where the second inequality follows from $A \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ in an obvious way and we only prove the first one:

$$\begin{aligned}
\max_{x \in \mathfrak{D}_s(A)} \min_{y \in A} ||x - y||_\diamondsuit &= \max_{x \in A} \min_{y \in A} ||\mathfrak{D}_s(x) - y||_\diamondsuit && (562)\\
&= \max_{x \in A} \min_{y \in A} ||(1-s)x + s\mathfrak{D}_1 \circ x - (1-s)y - sy||_\diamondsuit && (563)\\
&\leq \max_{x \in A} \min_{y \in A} (||(1-s)x - (1-s)y||_\diamondsuit + ||sy - s\mathfrak{D}_1 \circ x||_\diamondsuit) && (564)\\
&\leq \max_{x \in A} |s| \cdot ||x - \mathfrak{D}_1 \circ x||_\diamondsuit. && (565)
\end{aligned}$$

A similar calculation leads to

$$\max_{x\in A}\min_{y\in\mathfrak{D}_s(A)}||x-y||_\diamond \leq |s|\cdot\max_{x\in A}\cdot||x-\mathfrak{D}_1\circ x||_\diamond. \tag{566}$$

Application of the triangle inequality for $D_\diamond$ gives us the estimate

$$D_\diamond(P_\eta,\mathrm{conv}(\mathfrak{I}))\leq 4\eta. \tag{567}$$

Lemma c-lemma:estimate-for-coherent-information (originating back to [51]), finally makes the connection between our set-theoretic approximations and the capacity formula:

$$|\frac{1}{k}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes k})}\inf_{\mathcal{N}\in P_\eta} I_c(\rho,\mathcal{N}^{\otimes k})-\frac{1}{k}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes k})}\inf_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho,\mathcal{N}^{\otimes k})|\leq\nu(8\eta) \tag{568}$$

with $\nu(x)=x+8x\log(d_\mathcal{K})+4h(x)$. It is obvious that $-\eta\geq-\nu(8\eta)$ holds, therefore for l large enough

$$\frac{1}{l}\log\dim\mathcal{F}_l\geq\frac{1}{k}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes k})}\inf_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho,\mathcal{N}^{\otimes k})-2\nu(8\eta). \tag{569}$$

□

This leads to the following corollary to Theorem 98.

Corollary 100. *For any AVQC $\mathfrak{I}=\{\mathcal{N}_s\}_{s\in\mathbf{S}}$ we have*

$$\mathcal{A}_{\mathrm{random}}(\mathfrak{I})\geq\lim_{l\to\infty}\frac{1}{l}\max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})}\inf_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho,\mathcal{N}^{\otimes l}). \tag{570}$$

Together with Theorem 91 this proves the first part of Theorem 76.

4.5 Achievability of entanglement transmission rate II: Derandomization

In this section we will prove the second claim made in Theorem 76 by following Ahlswede's elimination technique. The main result of this section is the following Theorem.

Theorem 101. *Let $\mathfrak{I}=\{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an AVQC. Then $C_{\mathrm{det}}(\mathfrak{I})>0$ implies $\mathcal{A}_{\mathrm{det}}(\mathfrak{I})=\mathcal{A}_{\mathrm{random}}(\mathfrak{I})$.*

The proof of Theorem 101 is based mainly on the following lemma, which shows that not much of common randomness is needed to achieve $\mathcal{A}_{\mathrm{random}}(\mathfrak{I})$.

Lemma 102 (Random Code Reduction)**.** *Let $\mathfrak{I}=\{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be an AVQC, $l\in\mathbb{N}$, and μ_l an (l,k_l)-random code for the AVQC $\mathfrak{I}$ with*

$$e(\mu_l,\mathfrak{I}):=\inf_{s^l\in\mathbf{S}^l}\int F_e(\pi_{\mathcal{F}_l},\mathcal{R}^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}^l)d\mu_l(\mathcal{P}^l,\mathcal{R}^l)\geq 1-\varepsilon_l \tag{571}$$

for sequence $(\varepsilon_l)_{l\in\mathbb{N}}$ such that $\varepsilon_l\searrow 0$.
Let $\varepsilon\in(0,1)$. Then for all sufficiently large $l\in\mathbb{N}$ there exist l^2 codes $\{(\mathcal{P}_i^l,\mathcal{R}_i^l):i=1,\ldots,l^2\}\subset\mathcal{C}(\mathcal{F}_l,\mathcal{H}^{\otimes l})\times\mathcal{C}(\mathcal{K}^{\otimes l},\mathcal{F}_l')$ such that

$$\frac{1}{l^2}\sum_{i=1}^{l^2}F_e(\pi_{\mathcal{F}_l},\mathcal{R}_i^l\circ\mathcal{N}_{s^l}\circ\mathcal{P}_i^l)>1-\varepsilon\qquad\forall s^l\in\mathbf{S}^l. \tag{572}$$

Proof. Before we get into the details, we should note that the whole proof can be read much more easily if one restricts to the case $|\mathfrak{I}| < \infty$ and sets each of the approximating sets occurring in the sequel equal to $\mathfrak{I}$.
Let (Λ_i, Ω_i), $i = 1, \dots, K$, be independent random variables with values in $\mathcal{C}(\mathcal{F}_l, \mathcal{H}^{\otimes l}) \times \mathcal{C}(\mathcal{K}^{\otimes l}, \mathcal{F}_l')$ which are distributed according to $\mu_l^{\otimes K}$. Let $(P_l)_{l\in\mathbb{N}}$ be a sequence of polytopes with, for all $l \in \mathbb{N}$, the properties

1. $P_l \subset \mathrm{conv}(\mathfrak{I})$
2. $D_\diamondsuit(P_l, \mathrm{conv}(\mathfrak{I})) \leq 1/l^2$.

Denote by $ext(P_l)$ the extremal points of P_l. Consider an indexing such that we can write $ext(P_l) = \{\mathcal{N}_e\}_{e\in E_l}$ and note that the polytope P_l can be chosen in such a way that $N_l := |E_l|$ satisfies $N_l \leq (6l)^{4\dim(\mathcal{H})^2\dim(\mathcal{K})^2}$ (see, for example, Lemma 5.2 in [15]).
For every $e^l \in E_l^l$ and corresponding channel $\mathcal{N}_{e^l}$, an application of Markov's inequality yields for any $\varepsilon \in (0,1)$ and any $\gamma > 0$ the following:

$$\mathbb{P}\left(1 - \frac{1}{K}\sum_{i=1}^{K} F_e(\pi_{\mathcal{F}_l}, \Lambda_i \circ \mathcal{N}_{e^l} \circ \Omega_i) \geq \varepsilon/2\right) = \mathbb{P}\left(2^{K\gamma - \gamma\sum_{i=1}^K F_e(\pi_{\mathcal{F}_l}, \Lambda_i\circ\mathcal{N}_{e^l}\circ\Omega_i)} \geq 2^{K\gamma(\varepsilon/2)}\right) \quad (573)$$

$$\leq 2^{-K\gamma(\varepsilon/2)} \cdot \mathbb{E}\left(2^{\gamma(K - \sum_{i=1}^K F_e(\pi_{\mathcal{F}_l}, \Lambda_i\circ\mathcal{N}_{e^l}\circ\Omega_i))}\right). \quad (574)$$

We will derive an upper bound on the expectation in the preceding line:

$$\mathbb{E}\left(2^{\gamma(K - \sum_{i=1}^K F_e(\pi_{\mathcal{F}_l}, \Lambda_i\circ\mathcal{N}_{e^l}\circ\Omega_i))}\right) = \mathbb{E}\left(2^{\gamma(\sum_{i=1}^K (1 - F_e(\pi_{\mathcal{F}_l}, \Lambda_i\circ\mathcal{N}_{e^l}\circ\Omega_i)))}\right) \quad (575)$$

$$\overset{(a)}{=} \left[\mathbb{E}\left(2^{\gamma(1 - F_e(\pi_{\mathcal{F}_l}, \Lambda_1\circ\mathcal{N}_{e^l}\circ\Omega_1))}\right)\right]^K \quad (576)$$

$$\overset{(b)}{\leq} [\mathbb{E}(1 + 2^\gamma(1 - F_e(\pi_{\mathcal{F}_l}, \Lambda_1 \circ \mathcal{N}_{e^l} \circ \Omega_1)))]^K \quad (577)$$

$$\overset{(c)}{\leq} [1 + 2^\gamma \varepsilon_l]^K. \quad (578)$$

We used (a) independence of the (Λ_i, Ω_i), (b) the inequality $2^{\gamma t} \leq (1-t)2^{\gamma\cdot 0} + t2^\gamma \leq 1 + t2^\gamma,\ t \in [0,1]$, where the first inequality is simply the convexity of $[0,1] \ni t \mapsto 2^{\gamma t}$, (c) holds by (571) and by $P_l \subset \mathrm{conv}(\mathfrak{I})$. Now, for $K = l^2$, $\gamma = 2$ there is an $l_0(\varepsilon) \in \mathbb{N}$ such that for all $l \geq l_0(\varepsilon)$ we have

$$(1 + 2^2\varepsilon_l)^{l^2} \leq 2^{l^2(\varepsilon/2)}. \quad (579)$$

Therefore, we obtain from (573), (575), and (579) that for all sufficiently large $l \in \mathbb{N}$

$$\mathbb{P}\left(1 - \frac{1}{l^2}\sum_{i=1}^{l^2} F_e(\pi_{\mathcal{F}_l}, \Lambda_i \circ \mathcal{N}_{e^l} \circ \Omega_i) \geq (\varepsilon/2)\right) \leq 2^{-l^2(\varepsilon/2)} \quad (580)$$

uniformly in $e^l \in E_l^l$. It follows from (580) that

$$\mathbb{P}\left(\frac{1}{l^2}\sum_{i=1}^{l^2} F_e(\pi_{\mathcal{F}_l}, \Lambda_i \circ \mathcal{N}_{e^l} \circ \Omega_i) > 1 - \varepsilon/2\ \forall e^l \in E_l^l\right) \geq 1 - N_l^l \cdot 2^{-l^2(\varepsilon/2)} \quad (581)$$

implying the existence of a realization $(\mathcal{P}_i^l, \mathcal{R}_i^l)_{i=1}^{l^2}$ with

$$\frac{1}{l^2}\sum_{i=1}^{l^2} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_i^l \circ \mathcal{N}_{e^l} \circ \mathcal{P}_i^l) > 1 - \varepsilon/2 \qquad \forall e^l \in \mathbf{E}_l^l \quad (582)$$

whenever $N_l^l \cdot 2^{-l^2\varepsilon} < 1$, which is clearly fulfilled for all sufficiently large $l \in \mathbb{N}$.
Finally, we note that for every $l \in \mathbb{N}$ and $\mathcal{N}_s \in \mathfrak{I}$ there is $\mathcal{N}_e \in E_l$ such that $||\mathcal{N}_s - \mathcal{N}_e||_\diamond \leq \frac{1}{l^2}$ and, therefore, to every $\mathcal{N}_{s^l}$ there exists $\mathcal{N}_{e^l}$ (with each $\mathcal{N}_{e_i} \in E_l$) such that (see the proof of Lemma 41 for details)

$$||\mathcal{N}_{s^l} - \mathcal{N}_{e^l}||_\diamond \leq \sum_{i=1}^{l} ||\mathcal{N}_{s_i} - \mathcal{N}_{e_i}||_\diamond \leq \frac{1}{l}, \tag{583}$$

and therefore for every $s^l \in \mathbf{S}^l$ we have, for a maybe even larger l as before (satisfying $1/l < \varepsilon/2$, additionally),

$$\frac{1}{l^2} \sum_{i=1}^{l^2} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_i^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}_i^l) > 1 - \varepsilon \qquad \forall s^l \in \mathbf{S}^l. \tag{584}$$

□

We proceed with the proof of Theorem 101. Since $C_{\det}(\mathfrak{I}) > 0$ according to the assumption of the theorem, there is an (m_l, l^2)-deterministic code $\mathfrak{C}_{m_l} = (\rho_i, D_i)_{i=1}^{l^2}$ with $\rho_1, \ldots, \rho_{l^2} \in \mathcal{S}(\mathcal{H}^{\otimes m_l})$, $D_1, \ldots, D_{l^2} \in \mathcal{B}(\mathcal{K}^{\otimes m_l})$ with $m_l = o(l)$ and

$$\bar{P}_{e,m_l} = \sup_{s^{m_l} \in \mathbf{S}^{m_l}} P_e(\mathfrak{C}_{m_l}, s^{m_l}) \leq \varepsilon. \tag{585}$$

On the other hand, let us consider an (l, k_l)-random code as in Lemma 102, i.e. with

$$\frac{1}{l^2} \sum_{i=1}^{l^2} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_i^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}_i^l) > 1 - \varepsilon \qquad \forall s^l \in \mathbf{S}^l. \tag{586}$$

Define CPTP maps $\mathcal{P}^{l+m_l} \in \mathcal{C}(\mathcal{F}_l, \mathcal{H}^{\otimes l+m_l})$, $\mathcal{R}^{l+m_l} \in \mathcal{C}(\mathcal{K}^{\otimes l+m_l}, \mathcal{F}_l')$ by

$$\mathcal{P}^{l+m_l}(a) := \frac{1}{l^2} \sum_{i=1}^{l^2} \mathcal{P}_i^l(a) \otimes \rho_i \quad \text{and} \quad \mathcal{R}^{l+m_l}(b \otimes d) := \sum_{i=1}^{l^2} \operatorname{tr}(D_i d) \mathcal{R}_i^l(b). \tag{587}$$

Then for each $s^{l+m_l} = (v^l, u^{m_l}) \in \mathbf{S}^{l+m_l}$

$$\begin{aligned} F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^{l+m_l} \circ (\mathcal{N}_{v^l} \otimes \mathcal{N}_{u^{m_l}}) \circ \mathcal{P}^{l+m_l}) &= \frac{1}{l^2} \sum_{i,j=1}^{l^2} \operatorname{tr}(D_j \mathcal{N}_{u^{m_l}}(\rho_i)) F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_j^l \circ \mathcal{N}_{v^l} \circ \mathcal{P}_i^l) \quad (588) \\ &\geq \frac{1}{l^2} \sum_{i=1}^{l^2} \operatorname{tr}(D_i \mathcal{N}_{u^{m_l}}(\rho_i)) F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_i^l \circ \mathcal{N}_{v^l} \circ \mathcal{P}_i^l), \quad (589) \end{aligned}$$

where in the last line we have used that all involved terms are non-negative. In order to show that the fidelity on the left-hand side of (588) is at least $1 - 2\varepsilon$ we need the following lemma from [2].

Lemma 103. *Let $K \in \mathbb{N}$ and real numbers $a_1, \ldots, a_K, b_1, \ldots, b_K \in [0, 1]$ be given. Assume that*

$$\frac{1}{K} \sum_{i=1}^{K} a_i \geq 1 - \varepsilon \qquad \text{and} \qquad \frac{1}{K} \sum_{i=1}^{K} b_i \geq 1 - \varepsilon, \tag{590}$$

hold. Then

$$\frac{1}{K} \sum_{i=1}^{K} a_i b_i \geq 1 - 2\varepsilon. \tag{591}$$

Applying this lemma with $K = l^2$,

$$a_i = \mathrm{tr}(D_i \mathcal{N}_{u^{m_l}}(\rho_i)), \quad \text{and} \quad b_i = F_e(\pi_{\mathcal{F}_l}, \mathcal{R}_i^l \circ \mathcal{N}_{v^l} \circ \mathcal{P}_i^l) \tag{592}$$

along with (585), (586), and (588) shows that

$$F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^{l+m_l} \circ (\mathcal{N}_{v^l} \otimes \mathcal{N}_{u^{m_l}}) \circ \mathcal{P}^{l+m_l}) \geq 1 - 2\varepsilon. \tag{593}$$

On the other hand we know from Theorem 98 that for each sufficiently small $\eta > 0$ there is a random code μ_l for the AVQC $\mathfrak{I}$ with

$$\frac{1}{l} \log \dim \mathcal{F}_l \geq \frac{1}{k} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes k})} \inf_{\mathcal{N} \in \mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes k}) - \eta, \tag{594}$$

and

$$e(\mu_l, \mathfrak{I}) = \inf_{s^l \in \mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{P}^l, \mathcal{R}^l) \geq 1 - (l+1)^{N_\eta} 2^{-lc} \tag{595}$$

for all sufficiently large l with $c = c(k, \dim \mathcal{H}, \dim \mathcal{K}, \mathrm{conv}(\mathfrak{I}), \eta)$ and $N_\eta \in \mathbb{N}$. Thus the arguments that led us to (593) show that for all sufficiently large l there is a deterministic $(l + m_l, k_l)$-code for the AVQC $\mathfrak{I}$ with

$$F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^{l+m_l} \circ (\mathcal{N}_{v^l} \otimes \mathcal{N}_{u^{m_l}}) \circ \mathcal{P}^{l+m_l}) \geq 1 - 2\varepsilon, \tag{596}$$

and

$$\frac{1}{l+m_l} \log \dim \mathcal{F}_l \geq \frac{1}{k} \max_{\rho \in \mathcal{S}(\mathcal{H}^{\otimes k})} \inf_{\mathcal{N} \in \mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes k}) - 2\eta \tag{597}$$

by (594) and since $m_l = o(l)$. This shows that $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) \geq \mathcal{A}_{\mathrm{random}}(\mathfrak{I})$. Since the reverse inequality is trivially true we are done.

4.6 Zero-capacity-conditions: Symmetrizability

The most basic quality feature of an information processing system is whether it can be used for communication at a positive rate or not. This applies especially to such rather complex systems as AVCs or AVQCs. The notion of symmetrizability stems from the theory of classical AVCs and it addresses exactly that question. A classical AVC has deterministic capacity for message transmission equal to zero if and only if it is symmetrizable (with the definition of symmetrizability adjusted to the two different scenarios 'average error criterion' and 'maximal error criterion') [29] and [21], [45].
Of course, a similar statement for $\mathcal{A}_{\mathrm{det}}$ would be of great interest.

In this section we give three different conditions for three different capacities of an AVQC to be equal to zero. We restrict ourselves to the case $|\mathfrak{I}| < \infty$. Starting with the statement that has the weakest information theoretic consequences, we proceed to stronger statements. The case $|\mathfrak{I}| = \infty$ requires some involved continuity issues which shall be carried out elsewhere.
All three conditions have in common that they enable the adversary to simulate, on average over some probability distribution, a different output at the receiver side than the one that was originally put into the channel by the sender. The first two conditions, dealing with message transmission, exhibit a possibly nonlinear dependence between message set and probability distribution. They are direct (but not single-letter) analogs of their classical counterparts.
The third one is a sufficient condition for $\mathcal{A}_{\mathrm{random}}$ to be equal to zero. It employs a linear dependence between input state and probability distribution. If this condition is valid, the adversary is not only able to simulate a wrong output, he can also simulate an entanglement breaking channel between sender and receiver. In contrast to the first two criteria, this third one is a single-letter criterion.
There is a fourth and, at first sight, trivial condition, given by the following: An AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s \in \mathbf{S}}$ has

(deterministic *and* random) capacity for transmission of entanglement equal to zero if there is an $s \in \mathbf{S}$ such that $\mathcal{N}_s$ has zero capacity for transmission of entanglement.
We note that this fourth condition is nontrivial only because of the following reason: there is, until now, no way of telling exactly when a given (memoryless) quantum channel has a capacity greater than zero (except for calculating (449) for a single channel, an awkward task in general). This is in sharp contrast to the classical case, where the question can be trivially answered: A classical memoryless channel has a nonzero capacity if and only if there are at least two input states that lead to different output states.
Since our results do not answer the question whether it can happen that $C_{\mathrm{det}}(\mathfrak{I}) = 0$, $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = 0$ and $\mathcal{A}_{\mathrm{random}}(\mathfrak{I}) > 0$ hold simultaneously for a given AVQC $\mathfrak{I}$, we are left with two interesting and intimately related questions:
First, there is the zero-capacity question for single memoryless channels. Second, we need to find a criterion telling us exactly when $\mathcal{A}_{\mathrm{det}}$ is equal to zero.

4.6.1 Classical capacity with deterministic codes and average error

We now introduce a notion of symmetrizability which is a sufficient and necessary condition for $C_{\mathrm{det}}(\mathfrak{I}) = 0$. Our approach is motivated by the corresponding concept for arbitrarily varying channels with classical input and quantum output (cq-AVC) given in [6].
A nontrivial example of a non-symmetrizable AVQC can be found in subsection 4.8.1, see step **D** in the proof of Lemma 113.

Definition 104. *Let* $\mathbf{S}$ *be a finite set and* $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ *an AVQC.*

1. *$\mathfrak{I}$ is called l-symmetrizable, $l \in \mathbb{N}$, if for each finite set $\{\rho_1, \ldots, \rho_K\} \subset \mathcal{S}(\mathcal{H}^{\otimes l})$, $K \in \mathbb{N}$, there is a map $p : \{\rho_1, \ldots, \rho_K\} \to \mathfrak{P}(\mathbf{S}^l)$ such that for all $i, j \in \{1, \ldots, K\}$*

$$\sum_{s^l\in\mathbf{S}^l} p(\rho_i)(s^l)\mathcal{N}_{s^l}(\rho_j) = \sum_{s^l\in\mathbf{S}^l} p(\rho_j)(s^l)\mathcal{N}_{s^l}(\rho_i) \tag{598}$$

   holds.

2. *We call $\mathfrak{I}$ symmetrizable if it is l-symmetrizable for all $l \in \mathbb{N}$.*

We now state the main statement of this subsection.

Theorem 105. *Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$, $|\mathbf{S}| < \infty$, be an AVQC. Then $\mathfrak{I}$ is symmetrizable if and only if $C_{\mathrm{det}}(\mathfrak{I}) = 0$.*

Proof. 1. "Symmetrizability implies $C_{\mathrm{det}}(\mathfrak{I}) = 0$".
The proof follows closely the corresponding arguments given in [29], [21], and [6]. We give the full proof for reader's convenience. Suppose that $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ is symmetrizable and let $(\rho_i, D_i)_{i=1}^M$, $M \geq 2$, be a code for transmission of messages over $\mathfrak{I}$ with $\{\rho_1, \ldots, \rho_M\} \subset \mathcal{S}(\mathcal{H}^{\otimes l})$ and POVM $\{D_i\}_{i=1}^M$ on $\mathcal{H}^{\otimes l}$. Since $\mathfrak{I}$ is symmetrizable there is a map $p : \{\rho_1, \ldots, \rho_M\} \to \mathfrak{P}(\mathbf{S}^l)$ such that for all $i, j \in \{1, \ldots, M\}$

$$\sum_{s^l\in\mathbf{S}^l} p(\rho_i)(s^l)\mathcal{N}_{s^l}(\rho_j) = \sum_{s^l\in\mathbf{S}^l} p(\rho_j)(s^l)\mathcal{N}_{s^l}(\rho_i) \tag{599}$$

For $s^l \in \mathbf{S}^l$ and $i \in \{1, \ldots, M\}$ we set

$$e(i, s^l) := 1 - \mathrm{tr}(\mathcal{N}_{s^l}(\rho_i)D_i) = \sum_{\substack{j=1\\ j\neq i}}^{M} \mathrm{tr}(\mathcal{N}_{s^l}(\rho_i)D_j). \tag{600}$$

For $k \in \{1, \ldots, M\}$ let S_k^l be a random variable taking values in $\mathbf{S}^l$ and which is distributed according to $(p(\rho_k)(s^l))_{s^l \in \mathbf{S}^l}$. Then using relation (600) we can write

$$\begin{aligned}
\mathbb{E}(e(i, S_k^l)) &= \sum_{s^l \in \mathbf{S}^l} \sum_{\substack{j=1 \\ j \neq i}}^{M} p(\rho_k)(s^l) \mathrm{tr}(\mathcal{N}_{s^l}(\rho_i) D_j) && (601) \\
&= \sum_{\substack{j=1 \\ j \neq i}}^{M} \mathrm{tr}\{ \sum_{s^l \in \mathbf{S}^l} p(\rho_k)(s^l) \mathcal{N}_{s^l}(\rho_i) D_j \} && (602) \\
&= \sum_{\substack{j=1 \\ j \neq i}}^{M} \mathrm{tr}(\sum_{s^l \in \mathbf{S}^l} p(\rho_i)(s^l) \mathcal{N}_{s^l}(\rho_k) D_j) && (603) \\
&= \sum_{\substack{j=1 \\ j \neq i}}^{M} \sum_{s^l \in \mathbf{S}^l} p(\rho_i)(s^l) \mathrm{tr}(\mathcal{N}_{s^l}(\rho_k) D_j), && (604)
\end{aligned}$$

where the third line is by (599). On the other hand we have

$$\mathbb{E}(e(k, S_i^l)) = \sum_{s^l \in \mathbf{S}^l} \sum_{\substack{j=1 \\ j \neq k}}^{M} p(\rho_i)(s^l) \mathrm{tr}(\mathcal{N}_{s^l}(\rho_k) D_j). \tag{605}$$

Since $\{D_i\}_{i=1}^M$ is a POVM (601) and (605) imply that for $i \neq k$

$$\mathbb{E}(e(i, S_k^l)) + \mathbb{E}(e(k, S_i^l)) \geq 1 \tag{606}$$

holds. Let us abbreviate $\mathfrak{C} := (\rho_i, D_i)_{i=1}^M$, then with

$$\bar{P}_e(\mathfrak{C}, s^l) = \frac{1}{M} \sum_{k=1}^{M} (1 - \mathrm{tr}(\mathcal{N}_{s^l}(\rho_k) D_k)) \tag{607}$$

for $s^l \in \mathbf{S}^l$ we obtain

$$\begin{aligned}
\mathbb{E}(\bar{P}_e(\mathfrak{C}, S_j^l)) &= \sum_{s^l \in \mathbf{S}^l} p(\rho_j)(s^l) \frac{1}{M} \sum_{k=1}^{M} (1 - \mathrm{tr}(\mathcal{N}_{s^l}(\rho_k) D_k)) && (608) \\
&= \frac{1}{M} \sum_{k=1}^{M} \mathbb{E}(e(k, S_j^l)). && (609)
\end{aligned}$$

(606) and (608) yield

$$\begin{aligned}
\frac{1}{M} \sum_{j=1}^{M} \mathbb{E}(\bar{P}_e(\mathfrak{C}, S_j^l)) &= \frac{1}{M^2} \sum_{i,j=1}^{M} \mathbb{E}(e(k, S_j^l)) && (610) \\
&\geq \frac{1}{M^2} \binom{M}{2} && (611) \\
&= \frac{M-1}{2M} \geq \frac{1}{4} && (612)
\end{aligned}$$

for $M \geq 2$. Thus it follows that there is at least one $j \in \{1, \ldots, M\}$ with

$$\mathbb{E}(\bar{P}_e(\mathfrak{C}, S_j^l)) \geq \frac{1}{4} \tag{613}$$

and consequently there is at least one $s^l \in \mathbf{S}^l$ with

$$\bar{P}_e(\mathfrak{C}, s^l) \geq \frac{1}{4} \tag{614}$$

implying that $C_{\det}(\mathfrak{I}) = 0$.
2. "$C_{\det}(\mathfrak{I}) = 0$ implies symmetrizability".
Suppose that $\mathfrak{I}$ is non-symmetrizable. Then there is an $\hat{l} \in \mathbb{N}$ and a finite set $\{\rho_x\}_{x\in\mathcal{X}} \subset \mathcal{S}(\mathcal{H}^{\otimes\hat{l}})$ such that for no map $p : \{\rho_x\}_{x\in\mathcal{X}} \to \mathfrak{P}(\mathbf{S}^{\hat{l}})$ the relation (598) holds. Let us define for each $s^{\hat{l}} \in \mathbf{S}^{\hat{l}}$ a cq-channel $\mathcal{X} \ni x \mapsto W_{s^{\hat{l}}}(x) := \mathcal{N}_{s^{\hat{l}}}(\rho_x) \in \mathcal{S}(\mathcal{K}^{\otimes\hat{l}})$, and consider the cq-AVC generated by the set $\mathfrak{I}_{cq} := \{W_{s^{\hat{l}}}\}_{s^{\hat{l}}\in\mathbf{S}^{\hat{l}}}$. Then, due to the assumed non-symmetrizability of $\mathfrak{I}$, our new cq-AVC $\mathfrak{I}_{cq}$ is non-symmetrizable in the sense of [6].
Since $\mathfrak{I}_{cq}$ is non-symmetrizable the reduction argument from [6] to the results of [21] show that the cq-AVC $\mathfrak{I}_{cq}$ has positive capacity. This implies the existence of a sequence $(K_m, f_m, D_m, \varepsilon_m)_{m\in\mathbb{N}}$, where $K_m \in \mathbb{N}$, $f_m : \{1, \ldots, K_m\} \to \mathcal{X}^m$, $D_m \in \mathcal{B}_+(\mathcal{H}^{\otimes l\cdot m})$, $\lim_{m\to\infty} \varepsilon_m \searrow 0$, $\liminf_{m\to\infty} \frac{1}{m} \log K_m = c > 0$ and $\frac{1}{K_m} \sum_{i=1}^{K_m} (1 - \mathrm{tr}(D_i W_{s^{\hat{l}}}^m(f(i)))) = \varepsilon_m$.
We may use this sequence to construct another sequence $(\rho_i, D_i)_{i=1}^{M_l}$ of deterministic codes for message transmission over $\mathfrak{I}$, thereby achieving a capacity of $\frac{1}{\hat{l}} c > 0$. A similar construction is carried out explicitly at the end of the proof of the following Theorem 107. □

Corollary 106. *If the AVQC $\mathfrak{I} = \{\mathcal{N}\}_{s\in\mathbf{S}}$ is symmetrizable then $\mathcal{A}_{\det}(\mathfrak{I}) = 0$.*

Proof. Note that $\mathcal{A}_{\det}(\mathfrak{I}) \leq C_{\det}(\mathfrak{I})$ and apply Theorem 105. □

4.6.2 Classical capacity with deterministic codes and maximal error

We will now investigate, when exactly it is possible to send classical messages at positive rate over a finite AVQC, with the error criterion being that of maximal rather than average error.

Theorem 107. *Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}} \subset \mathcal{C}(\mathcal{H}, \mathcal{K})$ be a finite AVQC. The classical deterministic maximal error capacity $C_{\det,\max}(\mathfrak{I})$ of $\mathfrak{I}$ is equal to zero if and only if for every $l \in \mathbb{N}$ and every set $\{\rho_1, \rho_2\} \subset \mathcal{S}(\mathcal{H}^{\otimes l})$ we have*

$$\mathrm{conv}(\{\mathcal{N}_{s^l}(\rho_1)\}_{s^l\in\mathbf{S}^l}) \cap \mathrm{conv}(\{\mathcal{N}_{s^l}(\rho_2)\}_{s^l\in\mathbf{S}^l}) \neq \emptyset. \tag{615}$$

Proof. We closely follow the line of proof given in [45]. Let us begin with the 'if' part. Let $K, l \in \mathbb{N}$, $\{\rho_1, \ldots, \rho_K\} \subset \mathcal{S}(\mathcal{H}^{\otimes l})$ and $D_1, \ldots, D_K \in \mathcal{B}_+(\mathcal{K}^{\otimes l})$ with $\sum_{i=1}^K D_i = \mathbf{1}_{\mathcal{K}^{\otimes l}}$ be a code for transmission of classical messages over $\mathfrak{I}$.
We show that the maximal error probability of this code is bounded away from zero for large enough l. Let, without loss of generality, l be such that

$$\mathrm{tr}(D_1 \mathcal{N}_{s^l}(\rho_1)) > 1/2 \qquad \forall\ s^l \in \mathbf{S}^l \tag{616}$$

$$\mathrm{tr}(D_2 \mathcal{N}_{s^l}(\rho_2)) > 1/2 \qquad \forall\ s^l \in \mathbf{S}^l. \tag{617}$$

We show that there is a contradiction between (616) and (617). By assumption, there exist probability distributions $p_1, p_2 \in \mathfrak{P}(\mathbf{S}^l)$ such that

$$\sum_{s^l\in\mathbf{S}^l} p_1(s^l) \mathcal{N}_{s^l}(\rho_1) = \sum_{s^l\in\mathbf{S}^l} p_2(s^l) \mathcal{N}_{s^l}(\rho_2). \tag{618}$$

Of course, (616) implies

$$\sum_{s^l\in\mathbf{S}^l} p_1(s^l)\mathrm{tr}(D_1\mathcal{N}_{s^l}(\rho_1)) > 1/2. \tag{619}$$

Together with (618) this leads to

$$\begin{aligned}
1/2 &< \sum_{s^l\in\mathbf{S}^l} p_1(s^l)\mathrm{tr}(D_1\mathcal{N}_{s^l}(\rho_1)) && (620)\\
&= \sum_{s^l\in\mathbf{S}^l} p_2(s^l)\mathrm{tr}(D_1\mathcal{N}_{s^l}(\rho_2)) && (621)\\
&\leq \sum_{s^l\in\mathbf{S}^l} p_2(s^l)\mathrm{tr}((D_1+\sum_{i=3}^{K} D_i)\mathcal{N}_{s^l}(\rho_2)) && (622)\\
&= \sum_{s^l\in\mathbf{S}^l} p_2(s^l)\mathrm{tr}((\mathbf{1}-D_2)\mathcal{N}_{s^l}(\rho_2)) && (623)\\
&= 1-\sum_{s^l\in\mathbf{S}^l} p_2(s^l)\mathrm{tr}(D_2\mathcal{N}_{s^l}(\rho_2)) && (624)\\
&< 1-1/2, && (625)
\end{aligned}$$

a clear contradiction. Thus, for every code the maximal error probability is bounded from below by $1/2$.
Let us turn to the 'only if' part.
Assume there is an $\hat{l}\in\mathbb{N}$ and a set $\{\rho_1,\rho_2\}\subset\mathcal{S}(\mathcal{H}^{\otimes\hat{l}})$ such that

$$\mathrm{conv}(\{\mathcal{N}_{s^{\hat{l}}}(\rho_1)\}_{s^{\hat{l}}\in\mathbf{S}^{\hat{l}}})\cap\mathrm{conv}(\{\mathcal{N}_{s^{\hat{l}}}(\rho_2)\}_{s^{\hat{l}}\in\mathbf{S}^{\hat{l}}})=\emptyset. \tag{626}$$

Thus, there exists a self adjoint operator $A\in\mathcal{B}(\mathcal{K}^{\otimes\hat{l}})$ such that

$$\mathrm{tr}(A\rho)<0\ \ \forall\ \rho\in\mathrm{conv}(\{\mathcal{N}_{s^{\hat{l}}}(\rho_1)\}_{s^{\hat{l}}\in\mathbf{S}^{\hat{l}}}),\qquad \mathrm{tr}(A\rho)>0\ \ \forall\ \rho\in\mathrm{conv}(\{\mathcal{N}_{s^{\hat{l}}}(\rho_2)\}_{s^{\hat{l}}\in\mathbf{S}^{\hat{l}}}). \tag{627}$$

Let A have a decomposition $A=\sum_{x=1}^{d}a_xA_x$, where a_x are real numbers (including the possibility of $a_x=0$ for some x) and A_x are one dimensional projections fulfilling $\sum_{x=1}^{d}A_x=\mathbf{1}_{\mathcal{K}^{\otimes\hat{l}}}$. For every $m\in\mathbb{N}$, define

$$P_1^m:=\sum_{x^m:\frac{1}{m}\sum_{i=1}^{m}a_{x_i}<0} A_{x_1}\otimes\ldots\otimes A_{x_m},\qquad P_2^m:=\sum_{x^m:\frac{1}{m}\sum_{i=1}^{m}a_{x_i}\geq 0} A_{x_1}\otimes\ldots\otimes A_{x_m}. \tag{628}$$

Then $P_1^m+P_2^m=\mathbf{1}_{\mathcal{K}^{\otimes\hat{l}\cdot m}}$. Let us denote elements of $\mathbf{S}^{\hat{l}m}$ by $s^{\hat{l}m}=(s_1^{\hat{l}},\ldots,s_m^{\hat{l}})$, where each $s_i^{\hat{l}}\in\mathbf{S}^{\hat{l}}$.
To every $s^{\hat{l}}\in\mathbf{S}^{\hat{l}}$, define probability distributions $p_{s^{\hat{l}}},q_{s^{\hat{l}}}\in\mathcal{S}(\{1,\ldots,d\})$ according to

$$p_{s^{\hat{l}}}(x):=\mathrm{tr}(A_x\mathcal{N}_{s^{\hat{l}}}(\rho_1)),\qquad q_{s^{\hat{l}}}(x):=\mathrm{tr}(A_x\mathcal{N}_{s^{\hat{l}}}(\rho_2)),\ \ \forall\ x\in\{1,\ldots,d\} \tag{629}$$

and to every $s^{\hat{l}m}\in\mathbf{S}^{\hat{l}m}$ we associate two real numbers $\bar{A}_{s^{\hat{l}m}}(\rho_1),\bar{A}_{s^{\hat{l}m}}(\rho_2)$ by

$$\bar{A}_{s^{\hat{l}m}}(\rho_1):=\sum_{s^{\hat{l}}\in\mathbf{S}^{\hat{l}}}\frac{1}{m}N(s^{\hat{l}}|s^{\hat{l}m})\mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(\rho_1)),\qquad \bar{A}_{s^{\hat{l}m}}(\rho_2):=\sum_{s^{\hat{l}}\in\mathbf{S}^{\hat{l}}}\frac{1}{m}N(s^{\hat{l}}|s^{\hat{l}m})\mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(\rho_2)), \tag{630}$$

with natural numbers $N(s^{\hat{l}}|s^{\hat{l}m}):=|\{i:s_i^{\hat{l}}=s^{\hat{l}},\ i\in\{1,\ldots,m\}\}|$ for every $s^{\hat{l}m}\in\mathbf{S}^{\hat{l}m}$ and $s^{\hat{l}}\in\mathbf{S}^{\hat{l}}$.
Obviously, $\bar{A}_{s^{\hat{l}m}}(\rho_1)<0$ and $\bar{A}_{s^{\hat{l}m}}(\rho_2)>0$. Setting

$$C:=\max_{(s^{\hat{l}},X)\in\mathbf{S}^{\hat{l}}\times\{\rho_1,\rho_2\}}(\mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(X))/2)^{-2}(\mathrm{tr}(A^2\mathcal{N}_{s^{\hat{l}}}(X))-\mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(X))^2) \tag{631}$$

we arrive, by application of Chebyshev's inequality and for every $s^{\hat{l}m} = (s_1^{\hat{l}}, \ldots, s_m^{\hat{l}}) \in \mathbf{S}^{\hat{l}m}$ at

$$\begin{aligned}
\mathrm{tr}(P_1^m \mathcal{N}_{s^{\hat{l}m}}(\rho_1^{\otimes m})) &= \sum_{x^m : \frac{1}{m}\sum_{i=1}^m a_{x_i} < 0} \mathrm{tr}(A_{x_1} \otimes \ldots \otimes A_{x_m} \mathcal{N}_{s_1^{\hat{l}}}(\rho_1) \otimes \ldots \otimes \mathcal{N}_{s_m^{\hat{l}}}(\rho_1)) && (632)\\
&= \sum_{x^m : \frac{1}{m}\sum_{i=1}^m a_{x_i} < 0} p_{s_1^{\hat{l}}}(x_1) \cdot \ldots \cdot p_{s_m^{\hat{l}}}(x_m) && (633)\\
&\geq \sum_{x^m : |\frac{1}{m}\sum_{i=1}^m a_{x_i} - \bar{A}_{s^{\hat{l}m}}(\rho_1)| \leq |\bar{A}_{s^{\hat{l}m}}(\rho_1)/2|} p_{s_1^{\hat{l}}}(x_1) \cdot \ldots \cdot p_{s_m^{\hat{l}}}(x_m) && (634)\\
&\geq 1 - \frac{1}{m}(\bar{A}_{s^{\hat{l}m}}(\rho_1)/2)^{-2} \sum_{s^{\hat{l}} \in \mathbf{S}^{\hat{l}}} \frac{1}{m} N(s^{\hat{l}}|s^{\hat{l}m})(\mathrm{tr}(A^2 \mathcal{N}_{s^{\hat{l}}}(\rho_1)) - \mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(\rho_1))^2) && (635)\\
&\geq 1 - \frac{1}{m} \max_{s^{\hat{l}}}(\mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(\rho_1))/2)^{-2}(\mathrm{tr}(A^2 \mathcal{N}_{s^{\hat{l}}}(\rho_1)) - \mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(\rho_1))^2) && (636)\\
&\geq 1 - \frac{1}{m} \cdot C. && (637)
\end{aligned}$$

In the very same way, we can prove that

$$\begin{aligned}
\mathrm{tr}(P_2^m \mathcal{N}_{s^{\hat{l}m}}(\rho_2^{\otimes m})) &= \sum_{x^m : \frac{1}{m}\sum_{i=1}^m a_{x_i} \geq 0} \mathrm{tr}(A_{x_1} \otimes \ldots \otimes A_{x_m} \mathcal{N}_{s_1^{\hat{l}}}(\rho_2) \otimes \ldots \otimes \mathcal{N}_{s_m^{\hat{l}}}(\rho_2)) && (638)\\
&= \sum_{x^m : \frac{1}{m}\sum_{i=1}^m a_{x_i} \geq 0} q_{s_1^{\hat{l}}}(x_1) \cdot \ldots \cdot q_{s_m^{\hat{l}}}(x_m) && (639)\\
&\geq \sum_{x^m : |\frac{1}{m}\sum_{i=1}^m a_{x_i} - \bar{A}_{s^{\hat{l}m}}(\rho_2)| \leq |\bar{A}_{s^{\hat{l}m}}(\rho_2)|/2} q_{s_1^{\hat{l}}}(x_1) \cdot \ldots \cdot q_{s_m^{\hat{l}}}(x_m) && (640)\\
&\geq 1 - \frac{1}{m} \max_{s^{\hat{l}}}(\mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(\rho_2))/2)^{-2}(\mathrm{tr}(A^2 \mathcal{N}_{s^{\hat{l}}}(\rho_2)) - \mathrm{tr}(A\mathcal{N}_{s^{\hat{l}}}(\rho_2))^2) && (641)\\
&\geq 1 - \frac{1}{m} \cdot C. && (642)
\end{aligned}$$

Take any $0 < \varepsilon < 1/4$. Let $m' = \min\{m \in \mathbb{N} : \frac{1}{m} \cdot C < \varepsilon\}$. Then

$$\mathrm{tr}(P_1^{m'} \mathcal{N}_{s^{\hat{l}m'}}(\rho_1^{\otimes m'})) \geq 1 - \varepsilon \qquad \mathrm{tr}(P_2^{m'} \mathcal{N}_{s^{\hat{l}m'}}(\rho_2^{\otimes m'})) \geq 1 - \varepsilon \qquad (643)$$

hold. Consider the classical AVC given by the family $J := \{c_{\nu,\delta}\}_{\delta,\nu \in [3/4,1]}$ of classical channels $c_{\nu,\delta} : \{0,1\} \to \{0,1\}$ with stochastic matrices defined via $c_{\nu,\delta}(1|1) := 1 - \nu,\ c_{\nu,\delta}(2|2) := 1 - \delta$. Clearly, J is a convex set and, for every $c_{\nu,\delta} \in J$ we have that

$$\begin{aligned}
\max_{p \in \mathfrak{P}(\{0,1\})} I(p, c_{\nu,\delta}) &\geq 1 - \frac{1}{2}(h(\nu) + h(\delta)) && (644)\\
&\geq 1 - h(3/4) && (645)\\
&> 0, && (646)
\end{aligned}$$

where $I(p, c_{\nu,\delta})$ is the mutual information of the probability distribution q on $\{1,2\} \times \{1,2\}$ which is generated by p and $c_{\nu,\delta}$ through $q(i,j) := p(i)c_{\nu,\delta}(j|i)$ $((i,j) \in \{1,2\} \times \{1,2\})$. The lower bound given here is calculated using an equidistributed input. Note further that for this special AVC, with notation taken from [7], $\bar{\bar{J}} = \mathrm{conv}(J) = J$.

At this point in their proof of the classical zero-capacity-condition for AVCs [45], Kiefer and Wolfowitz made reference to a result by Gilbert [33], who proved existence of codes that achieve a positive rate.

Kiefer and Wolfowitz used these codes for message transmission over an AVC with binary input and output alphabet. Our strategy of proof is to use the existence of codes for AVCs with binary input and output that is guaranteed by Theorem 1 of [7] instead. Together with (646) this theorem gives us the existence of a number $C' > 0$, a function $\kappa : \mathbb{N} \to \mathbb{R}$ with $\lim_{r\to\infty} \kappa(r) = 0$ and a sequence $(M^r, f^r, \varepsilon_r, (D_1^r, \ldots, D_{\lceil |M^r|}^r))_{r\in\mathbb{N}}$ where for each $r \in \mathbb{N}$:

1. $M^r = \{1, \ldots, N\}$ is a finite set of cardinality $N = |M^r| = 2^{r(C'-\kappa(r))}$,
2. $f^r : M^r \to \{1, 2\}^r$,
3. $\varepsilon_r \geq 0$ and $\lim_{r\to\infty} \varepsilon_r = 0$,
4. $D_1^r, \ldots, D_{\lceil |M^r|}^r \subset \{1, 2\}^n$ are pairwise disjoint and
5. for every sequence $x^r \in ([3/4, 1] \times [3/4, 1])^r$ and every $i \in M^r$ we have that

$$\sum_{y^n \in D_i^r} \prod_{j=1}^{r} c_{x_j}(y_j | f^r(i)_j) \geq 1 - \varepsilon_r. \tag{647}$$

For $n \in \mathbb{N}$, take the unique numbers $r \in \mathbb{N}$, $t \in \{0, \ldots, m' - 1\}$ such that $n = m'r + t$ holds. The code for $\mathfrak{I}$ is then defined as follows:

$$M_n := M^r, \tag{648}$$

$$f_n(i) := (\rho_{f^r(i)_1})^{\otimes m'} \otimes \ldots \otimes (\rho_{f^r(i)_r})^{\otimes m'} \otimes \sigma^{\otimes t}, \tag{649}$$

$$P_i^n := \sum_{y^r \in D_i^r} P_{y_1}^{m'} \otimes \ldots \otimes P_{y_r}^{m'} \otimes \mathbf{1}_{\mathcal{K}}^{\otimes t}. \tag{650}$$

Let, for every $s^{m'} \in \mathbf{S}^{m'}$, $x = (\nu, \delta) \in [3/4, 1]^2$ be such that

$$c_{\nu,\delta}(0|0) := \operatorname{tr}(P_1^{m'} \mathcal{N}_{s^{nm'}}(\rho_1^{\otimes m'})) = 1 - \nu \qquad c_{\nu,\delta}(1|1) := \operatorname{tr}(P_1^{m'} \mathcal{N}_{s^{nm'}}(\rho_2^{\otimes m'})) = 1 - \delta. \tag{651}$$

Then for every $s^n \in \mathbf{S}^n$ we use the decomposition $s^n = (s_1^{m'}, \ldots, s_r^{m'}, s^t)$ and get, using equation (647) and the definition (651), for every $i \in M_n$,

$$\operatorname{tr}\{P_i^n f_n(i)\} = \operatorname{tr}\{[\sum_{y^r \in D_i^r} P_{y_1}^{m'} \otimes \ldots \otimes P_{y_r}^{m'} \otimes \mathbf{1}_{\mathcal{K}}^{\otimes t}](\rho_{f^r(i)_1})^{\otimes m'} \otimes \ldots \otimes (\rho_{f^r(i)_r})^{\otimes m'} \otimes \sigma^{\otimes t}\} \tag{652}$$

$$= \sum_{y^r \in D_i^r} \prod_{j=1}^{r} \operatorname{tr}\{P_{y_j}^{m'} (\rho_{f^r(i)_j})^{\otimes m'}\} \tag{653}$$

$$= \sum_{y^r \in D_i^r} \prod_{j=1}^{r} c_{\nu,\delta}(y_j | f^r(i)_j) \tag{654}$$

$$\geq 1 - \varepsilon_r. \tag{655}$$

Obviously, this implies

$$\lim_{n\to\infty} \min_{s^n \in \mathbf{S}^n} \max_{i \in M_n} \operatorname{tr}\{P_i^n f_n(i)\} = 0. \tag{656}$$

Together with

$$\lim_{n\to\infty} \frac{1}{n} \log |M_n| = \frac{1}{m'} C' > 0 \tag{657}$$

we have shown that $C_{det,max}(\mathfrak{I}) > 0$ holds. □

Notice that the statements made in (618) and (615) are equivalent and a glance at Definition 104 reveals that the assertion of (618) is nothing else than the symmetrizability restricted to sets of states consisting of two elements.

4.6.3 Entanglement transmission capacity with random codes

The final issue in this section is a sufficient condition for $\mathcal{A}_{\text{random}}(\mathfrak{I}) = 0$ which is based on the notion of qc-symmetrizability.
Let $\mathfrak{F}_{\mathbb{C}}(\mathbf{S})$ stand for the set of $\mathbb{C}$-valued functions defined on $\mathbf{S}$ in what follows and we consider the set of channels with quantum input and classical output (qc-channels)[3]

$$\mathrm{QC}(\mathcal{H}, \mathbf{S}) := \{T : \mathcal{B}(\mathcal{H}) \to \mathfrak{F}_{\mathbb{C}}(\mathbf{S}) : T \text{ is linear, positive, and trace preserving}\}. \tag{658}$$

The condition that $T \in \mathrm{QC}(\mathcal{H}, \mathbf{S})$ is trace preserving means that

$$\sum_{s \in \mathbf{S}} [T(b)](s) = \mathrm{tr}(b) \tag{659}$$

holds for all $b \in \mathcal{B}(\mathcal{H})$. By Riesz' representation theorem there is a one-to-one correspondence between elements $T \in \mathrm{QC}(\mathcal{H}, \mathbf{S})$ and (discrete) positive operator-valued measures (POVM) $\{E_s\}_{s \in \mathbf{S}}$.
For a given finite set of quantum channels $\mathfrak{I} = \{\mathcal{N}_s\}_{s \in \mathbf{S}}$ and $T \in \mathrm{QC}(\mathcal{H}, \mathbf{S})$ we define a CPTP map $\mathcal{M}_{T,\mathbf{S}} : \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ by

$$\begin{aligned} \mathcal{M}_{T,\mathbf{S}}(a \otimes b) &:= \sum_{s \in \mathbf{S}} [T(a)](s) \mathcal{N}_s(b) && (660) \\ &= \sum_{s \in \mathbf{S}} \mathrm{tr}(E_s a) \mathcal{N}_s(b), && (661) \end{aligned}$$

where $\{E_s\}_{s \in \mathbf{S}}$ is the unique POVM associated with T.

Definition 108. *An arbitrarily varying quantum channel, generated by a finite set $\mathfrak{I} = \{\mathcal{N}_s\}_{s \in \mathbf{S}}$, is called qc-symmetrizable if there is $T \in QC(\mathcal{H}, \mathbf{S})$ such that for all $a, b \in \mathcal{B}(\mathcal{H})$*

$$\mathcal{M}_{T,\mathbf{S}}(a \otimes b) = \mathcal{M}_{T,\mathbf{S}}(b \otimes a) \tag{662}$$

holds, where $\mathcal{M}_{T,\mathbf{S}} : \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ is the CPTP map defined in (660).

The best illustration of the definition of qc-symmetrizability is given in the proof of our next theorem.

Theorem 109. *If an arbitrarily varying quantum channel generated by a finite set $\mathfrak{I} = \{\mathcal{N}_s\}_{s \in \mathbf{S}}$ is qc-symmetrizable, then for any sequence of (l, k_l)-random codes $(\mu_l)_{l \in \mathbb{N}}$ with $k_l = \dim \mathcal{F}_l \geq 2$ for all $l \in \mathbb{N}$ we have*

$$\inf_{s^l \in \mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \leq \frac{1}{2}, \tag{663}$$

for all $l \in \mathbb{N}$. Thus

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) = 0, \tag{664}$$

and consequently

$$\mathcal{A}_{\text{det}}(\mathfrak{I}) = 0. \tag{665}$$

Proof. We have to show that for the codes $(\mathcal{P}^l, \mathcal{R}^l)$ with the properties as stated in the lemma the inequality

$$\inf_{s^l \in \mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \leq \frac{1}{2} \tag{666}$$

[3]Mere positivity is sufficient here because $\mathfrak{F}_{\mathbb{C}}(\mathbf{S})$ is commutative, cf. [58].

holds for all $l \in \mathbb{N}$.
Let $\psi_l \in \mathcal{S}(\mathcal{F}_l \otimes \mathcal{F}_l)$ be a purification of $\pi_{\mathcal{F}_l}$ which is, clearly, maximally entangled. Inequality (666) can then be equivalently reformulated as

$$\inf_{s^l \in \mathbf{S}^l} \int \langle \psi_l, (id_{\mathcal{F}_l} \otimes (\mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l))(|\psi_l\rangle\langle\psi_l|)\psi_l \rangle d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \leq \frac{1}{2}. \tag{667}$$

We fix $\sigma \in \mathcal{S}(\mathcal{H})$ and define CPTP maps $E_1, E_2 : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ by

$$E_1(a) := \mathcal{M}_{T,\mathbf{S}}(\sigma \otimes a) = \sum_{s \in \mathbf{S}} \mathrm{tr}(E_s \sigma)\mathcal{N}_s(a) \tag{668}$$

and

$$E_2(a) := \mathcal{M}_{T,\mathbf{S}}(a \otimes \sigma) = \sum_{s \in \mathbf{S}} \mathrm{tr}(E_s a)\mathcal{N}_s(\sigma). \tag{669}$$

Then

$$\begin{aligned} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ E_1^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) &= \sum_{s^l \in \mathbf{S}^l} \mathrm{tr}(E_{s^l}\sigma^{\otimes l}) \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) & (670) \\ &\geq \inf_{s^l \in \mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l), & (671) \end{aligned}$$

where $E_{s^l} := E_{s_1} \otimes \ldots \otimes E_{s_l}$. Therefore, we are done if we can show that

$$\int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ E_1^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \leq \frac{1}{2} \tag{672}$$

for all $l \in \mathbb{N}$.
On the other hand, choosing bases $\{e_{i,j}\}_{i,j=1}^{k_l}$ and $\{f_{k,m}\}_{k,m=1}^{d^l}$ of $\mathcal{B}(\mathcal{F}_l)$ and $\mathcal{B}(\mathcal{H})^{\otimes l}$ respectively, we can write

$$id_{\mathcal{F}_l} \otimes \mathcal{P}^l(|\psi_l\rangle\langle\psi_l|) =: \rho_l = \sum_{i,j,k,m} \rho_{i,j,k,m} e_{i,j} \otimes f_{k,m}, \tag{673}$$

and obtain

$$\begin{aligned} id_{\mathcal{F}_l} \otimes (\mathcal{R}^l \circ E_1^{\otimes l})(\rho_l) &= \sum_{i,j,k,m} \rho_{i,j,k,m} e_{i,j} \otimes \mathcal{R}^l(\mathcal{M}_{T,\mathbf{S}}^{\otimes l}(\sigma^{\otimes l} \otimes f_{k,m})) & (674) \\ &= \sum_{i,j,k,m} \rho_{i,j,k,m} e_{i,j} \otimes \mathcal{R}^l(\mathcal{M}_{T,\mathbf{S}}^{\otimes l}(f_{k,m} \otimes \sigma^{\otimes l})) & (675) \\ &= \sum_{i,j,k,m} \rho_{i,j,k,m} e_{i,j} \otimes \mathcal{R}^l(E_2^{\otimes l}(f_{k,m})) & (676) \\ &= id_{\mathcal{F}_l} \otimes (\mathcal{R}^l \circ E_2^{\otimes l})(\rho_l), & (677) \end{aligned}$$

where the second equality follows from the assumed qc-symmetrizability. Thus, we end up with

$$F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ E_1^{\otimes l} \circ \mathcal{P}^l) = F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ E_2^{\otimes l} \circ \mathcal{P}^l), \tag{678}$$

for any encoding operation $\mathcal{P}^l$ and any recovery operation $\mathcal{R}^l$. Consequently, by (678) and (670) we have to show that for all $l \in \mathbb{N}$

$$\int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ E_2^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \leq \frac{1}{2} \tag{679}$$

holds. But the channel

$$E_2(a) = \sum_{s\in S} \mathrm{tr}(E_s a)\mathcal{N}_s(\sigma) \qquad (a \in \mathcal{B}(\mathcal{H})) \tag{680}$$

is entanglement breaking implying that the state

$$(id_{\mathcal{F}_l} \otimes \mathcal{R}^l \circ E_2^{\otimes l} \circ \mathcal{P}^l)(|\psi_l\rangle\langle\psi_l|) \tag{681}$$

is separable. A standard result from entanglement theory implies that

$$\langle\psi_l, (id_{\mathcal{F}_l} \otimes \mathcal{R}^l \circ E_2^{\otimes l} \circ \mathcal{P}^l)(|\psi_l\rangle\langle\psi_l|)\psi_l\rangle \leq \frac{1}{k_l} \tag{682}$$

holds for any $\mathcal{R}^l$ and $\mathcal{P}^l$ since ψ_l is maximally entangled with Schmidt rank k_l. Now, our assumption that for each $l \in \mathbb{N}$ the relation $k_l \geq 2$ holds implies along with (682) that for all $l \in \mathbb{N}$

$$\int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ E_2^{\otimes l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) = \int \langle\psi_l, (id_{\mathcal{F}_l} \otimes \mathcal{R}^l \circ E_2^{\otimes l} \circ \mathcal{P}^l)(|\psi_l\rangle\langle\psi_l|)\psi_l\rangle d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \leq \frac{1}{2}, \tag{683}$$

and by (678) and (670) we obtain

$$\inf_{s^l\in\mathbf{S}^l} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu_l(\mathcal{R}^l, \mathcal{P}^l) \leq \frac{1}{2} \tag{684}$$

which concludes the proof. □

Our Definition 108 addresses the notion of qc-symmetrizability for block length $l = 1$. Thus the question arises whether a less restrictive requirement, as stated in the following definition, gives a better sufficient condition for an arbitrarily varying quantum channel to have capacity 0.

Definition 110. *An arbitrarily varying quantum channel, generated by a finite set $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$, is called l-qc-symmetrizable, $l \in \mathbb{N}$, if there is $T \in QC(\mathcal{H}^{\otimes l}, \mathbf{S}^l)$ such that for all $a, b \in \mathcal{B}(\mathcal{H})^{\otimes l}$*

$$\mathcal{M}^l_{T,\mathbf{S}}(a \otimes b) = \mathcal{M}^l_{T,\mathbf{S}}(b \otimes a) \tag{685}$$

holds, where $\mathcal{M}^l_{T,\mathbf{S}} : \mathcal{B}(\mathcal{H})^{\otimes l} \otimes \mathcal{B}(\mathcal{H})^{\otimes l} \to \mathcal{B}(\mathcal{K})^{\otimes l}$ is the CPTP map defined by

$$\mathcal{M}^l_{T,\mathbf{S}}(a \otimes b) := \sum_{s^l\in\mathbf{S}^l} tr(E_{s^l} a)\mathcal{N}_{s^l}(b), \tag{686}$$

and $\{E_{s^l}\}_{s^l\in\mathbf{S}^l}$ is the unique POVM corresponding to $T \in QC(\mathcal{H}^{\otimes l}, \mathbf{S}^l)$.

Obviously, qc-symmetrizability implies l-qc-symmetrizability for all $l \in \mathbb{N}$. The next lemma states that the reverse implication is true too.

Lemma 111. *For any finitely generated AVQC given by $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ l-qc-symmetrizability implies qc-symmetrizability for any $l \in \mathbb{N}$.*

Proof. For a given finite set of quantum channels $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ and $l \in \mathbb{N}$ let $T \in \mathrm{QC}(\mathcal{H}^{\otimes l}, \mathbf{S}^l)$ be such that for all $a, b \in \mathcal{B}(\mathcal{H})^{\otimes l}$

$$\mathcal{M}^l_{T,\mathbf{S}}(a \otimes b) = \mathcal{M}^l_{T,\mathbf{S}}(b \otimes a), \tag{687}$$

where $\mathcal{M}^l_{T,\mathbf{S}}$ is defined in (686).
Let $b \in \mathcal{B}(\mathcal{H})$ and for each $s \in \mathbf{S}$ define a linear functional

$$\phi_s(b) := \mathrm{tr}\left(\left(b \otimes \left(\frac{1}{\dim\mathcal{H}}\mathbf{1}_{\mathcal{H}}\right)^{\otimes l-1}\right) \sum_{s_2^l\in\mathbf{S}^{l-1}} E_{ss_2^l}\right), \tag{688}$$

where $ss_2^l := (s, s_2, \ldots, s_l) \in \mathbf{S}^l$. Clearly, ϕ_s is positive. Consequently, Riesz' representation theorem shows that there is a unique positive $\tilde{E}_s \in \mathcal{B}(\mathcal{H})$ with

$$\phi_s(b) = \mathrm{tr}(\tilde{E}_s b) \qquad (b \in \mathcal{B}(\mathcal{H})). \tag{689}$$

Obviously, $\{\tilde{E}_s\}_{s\in\mathbf{S}}$ is a POVM and let $\tilde{T} \in \mathrm{QC}(\mathcal{H}, \mathbf{S})$ denote the associated qc-channel.
Some simple algebra shows that for each $a, b \in \mathcal{B}(\mathcal{H})$

$$\mathcal{M}_{\tilde{T},\mathbf{S}}(a \otimes b) = \mathrm{tr}_{\mathcal{K}^{\otimes l-1}} \mathcal{M}^l_{T,\mathbf{S}} \left(\left(a \otimes \left(\frac{1}{\dim \mathcal{H}} \mathbf{1}_{\mathcal{H}} \right)^{\otimes l-1} \right) \otimes \left(b \otimes \left(\frac{1}{\dim \mathcal{H}} \mathbf{1}_{\mathcal{H}} \right)^{\otimes l-1} \right) \right) \tag{690}$$

where $\mathrm{tr}_{\mathcal{H}^{\otimes l-1}}$ denotes the partial trace over the last $l-1$ tensor factors. The relation (690) immediately implies that for all $a, b \in \mathcal{B}(\mathcal{H})$

$$\mathcal{M}_{\tilde{T},\mathbf{S}}(a \otimes b) = \mathcal{M}_{\tilde{T},\mathbf{S}}(b \otimes a), \tag{691}$$

and

$$\begin{aligned}
\mathcal{M}_{\tilde{T},\mathbf{S}}(a \otimes b) &= \sum_{s^l \in \mathbf{S}^l} \mathrm{tr} \left(\left(b \otimes \left(\frac{1}{\dim \mathcal{H}} \mathbf{1}_{\mathcal{H}} \right)^{\otimes l-1} \right) E_{s^l} \right) \mathcal{N}_{s_1}(a) && (692)\\
&= \sum_{s_1 \in \mathbf{S}} \mathrm{tr} \left(\left(b \otimes \left(\frac{1}{\dim \mathcal{H}} \mathbf{1}_{\mathcal{H}} \right)^{\otimes l-1} \right) \sum_{s_2^l \in \mathbf{S}^{l-1}} E_{s^l} \right) \mathcal{N}_{s_1}(a) && (693)\\
&= \sum_{s_1 \in \mathbf{S}} \phi_{s_1}(b) \mathcal{N}_{s_1}(a) && (694)\\
&= \sum_{s_1 \in \mathbf{S}} \mathrm{tr}(b \tilde{E}_{s_1}) \mathcal{N}_{s_1}(a). && (695)
\end{aligned}$$

Equations (691) and (692) show that $\mathfrak{I}$ is qc-symmetrizable. □

4.7 Conditions for single-letter-capacities

In this section we give two conditions on the structure of a finite AVQC which guarantee that their quantum capacity is given by a single-letter formula. The first one is empty in the case of a single channel, while the second one generalizes the degradability condition from [25] that we repeat here for readers convenience:
A channel $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ is called degradable if for any Hilbert space $\mathcal{K}_E$ and any partial isometry $V : \mathcal{H} \to \mathcal{K} \otimes \mathcal{K}_E$ such that $\mathcal{N}(\cdot) = \mathrm{tr}_{\mathcal{K}_E}(V \cdot V^*)$ there is $\mathcal{N}_V \in \mathcal{C}(\mathcal{K}, \mathcal{K}_E)$ such that $\mathcal{N}_V \circ \mathcal{N} = \mathrm{tr}_{\mathcal{K}}(V \cdot V^*)$. The above definition is in fact independent from the choice of $\mathcal{K}_E$ and V - if it holds for only one such choice, then it holds for all possible choices.

Lemma 112. *Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ satisfy $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{random}}(\mathfrak{I})$ (for example, $\mathfrak{I}$ might be non-symmetrizable). We have a single-letter formula for $\mathcal{A}_{\mathrm{det}}(\mathfrak{I})$ in any of the following two cases:*

1. *There is $\mathcal{N}_* \in \mathrm{conv}(\mathfrak{I})$ such that for any $\mathcal{N} \in \mathrm{conv}(\mathfrak{I})$ there is $\mathcal{D}_{\mathcal{N}} \in \mathcal{C}(\mathcal{K}, \mathcal{K})$ with the property $\mathcal{N}_* = \mathcal{D}_{\mathcal{N}} \circ \mathcal{N}$ and, additionally, $Q(\mathcal{N}_*) = \max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{N}_*)$ holds for the entanglement transmission capacity $Q(\mathcal{N}_*)$ of the memoryless channel $\mathcal{N}_*$.*
2. *Each $\mathcal{N} \in \mathrm{conv}(\mathfrak{I})$ is degradable.*

Proof. 1. It is clear that

$$\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) \leq Q(\mathcal{N}_*) = \max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{N}_*). \tag{696}$$

By assumption, we have

$$\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{random}}(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{697}$$

On the other hand by application of the data-processing inequality [60] we have, for all $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$, $\mathcal{N} \in \mathrm{conv}(\mathfrak{I})$ and $l \in \mathbb{N}$,

$$\begin{aligned} I_c(\rho, \mathcal{N}^{\otimes l}) &\geq I_c(\rho, \mathcal{D}_{\mathcal{N}}^{\otimes l} \circ \mathcal{N}^{\otimes l}) && (698)\\ &= I_c(\rho, \mathcal{N}_*^{\otimes l}). && (699) \end{aligned}$$

It follows that

$$\frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}) \geq \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}_*^{\otimes l}) \tag{700}$$

and by (697):

$$\begin{aligned} \mathcal{A}_{\mathrm{det}}(\mathfrak{I}) &\geq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}_*^{\otimes l}) && (701)\\ &= Q(\mathcal{N}_*) && (702)\\ &= \max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{N}_*). && (703) \end{aligned}$$

Equations (696) and (703) give us the desired result:

$$\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = \max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{N}_*). \tag{704}$$

2. It is well known that the following three properties are valid:

- P1 If a $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ is degradable, then the map $\rho \mapsto I_c(\rho, \mathcal{N})$ is concave ([74], Lemma 5).
- P2 For every fixed $\rho \in \mathcal{S}(\mathcal{H})$, $\mathcal{N} \mapsto I_c(\rho, \mathcal{N})$ is convex (see [52], Theorem 1).
- P3 Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ be degradable. For an arbitrary $l \in \mathbb{N}$, write $\mathcal{H}^{\otimes l} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_l$ with $\mathcal{H}_i := \mathcal{H}$ for every $i \in \mathbb{N}$. Let $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$ with marginal states $\rho_i := \mathrm{tr}_{\mathcal{H}_1\otimes\ldots\otimes\mathcal{H}_{i-1}\otimes\mathcal{H}_{i+1}\otimes\ldots\otimes\mathcal{H}_l}(\rho)$. Then the inequality $I_c(\rho, \mathcal{N}^{\otimes l}) \leq \sum_{i=1}^n I_c(\rho_i, \mathcal{N})$ holds [25].
- P4 The coherent information is continuous in both of its entries.

By the minimax-theorem [67, 44], properties P1, P2 and P4 imply that

$$\max_{\rho\in\mathcal{S}(\mathcal{H})} \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}) = \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} \max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{N}). \tag{705}$$

Suppose now, that each $\mathcal{N} \in \mathrm{conv}(\mathfrak{I})$ is degradable. It then holds, for every $l \in \mathbb{N}$,

$$\begin{aligned} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}) &\leq \frac{1}{l} \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} I_c(\rho, \mathcal{N}^{\otimes l}) && (706)\\ &\leq \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} \max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{N}) && (707)\\ &= \max_{\rho\in\mathcal{S}(\mathcal{H})} \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}), && (708) \end{aligned}$$

where the second inequality follows from P3 and the equality from P1, P2 via the minimax-theorem. It follows that

$$\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) \leq \mathcal{A}_{\mathrm{random}}(\mathfrak{I}) \leq \max_{\rho\in\mathcal{S}(\mathcal{H})} \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}). \tag{709}$$

By assumption, we also have $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{random}}(\mathfrak{I})$. The obvious relation $\mathcal{A}_{\mathrm{random}}(\mathfrak{I}) \geq \max_{\rho\in\mathcal{S}(\mathcal{H})} \min_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N})$ then implies the reverse inequality. □

4.8 An example and an application to zero-error capacities

4.8.1 Erasure-AVQC

As an application and illustration of most of the results obtained so far we calculate the quantum capacity of finite AVQC $\mathfrak{I}$ consisting of erasure quantum channels. As expected, we obtain that $\mathcal{A}_{\mathrm{det}}(\mathfrak{I})$ equals the capacity of the worst erasure channel in the set $\mathfrak{I}$.

Lemma 113. *Let $d \in \mathbb{N}$, $d \geq 2$ and denote by $\{e_1, \dots e_d\}$, $\{e_1, \dots, e_{d+1}\}$, the standard basis of $\mathbb{C}^d, \mathbb{C}^{d+1}$. Set $\mathcal{H} = \mathbb{C}^d$, $\mathcal{K} = \mathbb{C}^{d+1}$. Define, for $p \in [0,1]$, the erasure channel $\mathcal{E}_p \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ by*

$$\mathcal{E}_p(x) := (1-p)x + p \cdot \mathrm{tr}(x)|e_{d+1}\rangle\langle e_{d+1}| \qquad \forall x \in \mathcal{B}(\mathcal{H}). \tag{710}$$

Let, for a finite collection $\{p_s\}_{s\in\mathbf{S}} \subset [0,1]$, an AVQC be given by $\mathfrak{I} = \{\mathcal{E}_{p_s}\}_{s\in\mathbf{S}}$.
The following are true.

1. *If $p_s \geq 1/2$ for some $s \in \mathbf{S}$, then $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{random}}(\mathfrak{I}) = 0$.*
2. *If $p_s < 1/2$ for every $s \in \mathbf{S}$, then $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{random}}(\mathfrak{I}) = \min_{s\in\mathbf{S}}(1-2p_s)\log(d)$.*

Proof. We start with *2.* by showing the validity of the following properties.

A. For $q \in \mathfrak{P}(\mathbf{S})$, we have $\sum_{s\in\mathbf{S}} q(s)\mathcal{E}_{p_s} = \mathcal{E}_{q(p)}$, where $q(p) := \sum_{s\in\mathbf{S}} q(s)p_s$.

B. There is a set $\{\widehat{\mathcal{E}_{p_s}}\}_{s\in\mathbf{S}}$ of complementary maps given by $\widehat{\mathcal{E}_{p_s}} = \mathcal{E}_{1-p_s}$, $s \in \mathbf{S}$.

C. $\mathcal{E}_p$ is degradable for $p \in [0, 1/2)$.

D. $\{\mathcal{E}_{p_s}\}_{s\in\mathbf{S}}$ is non-symmetrizable if $p_s \in [0,1)$ for all $s \in \mathbf{S}$.

A.: For every $x \in \mathcal{B}(\mathbb{C}^d)$,

$$\sum_{s\in\mathbf{S}} q(s)\mathcal{E}_{p_s}(x) = \sum_{s\in\mathbf{S}} q(s)[(1-p_s)x + p_s \cdot \mathrm{tr}(x)|e_{d+1}\rangle\langle e_{d+1}|] \tag{711}$$

$$= (1 - \sum_{s\in\mathbf{S}} q(s)p_s)x + \sum_{s\in\mathbf{S}} q(s)p_s \cdot \mathrm{tr}(x)|e_{d+1}\rangle\langle e_{d+1}|. \tag{712}$$

B.: Consider an environment defined by $\mathcal{K}_{env} := \mathcal{K}$. For every $p \in [0,1]$ we can give a Stinespring isometry $V_p : \mathcal{H} \to \mathcal{K} \otimes \mathcal{K}_{env}$ of $\mathcal{E}_p$ by

$$V_p u := \sqrt{1-p} \cdot u \otimes e_{d+1} + \sqrt{p} \cdot e_{d+1} \otimes u, \qquad u \in \mathcal{B}(\mathcal{H}). \tag{713}$$

The claim becomes clear by tracing out the first or second subsystem, depending on whether one wants to calculate $\hat{\mathcal{E}}_p$ or $\mathcal{E}_p$.
C.: Set $\mu := \frac{1-2p}{1-p}$ and define $E_\mu \in \mathcal{C}(\mathcal{K}, \mathcal{K})$ by

$$E_\mu(x) := (1-\mu) \cdot x + \mu \cdot \mathrm{tr}(x) \cdot |e_{d+1}\rangle\langle e_{d+1}|, \qquad x \in \mathcal{B}(\mathcal{K}). \tag{714}$$

Then by $p \in [0, 1/2)$ we have $\mu \in (0,1]$. We show that $\mathcal{E}_{1-p} = E_\mu \circ \mathcal{E}_p$ holds. Let $x \in \mathcal{B}(\mathcal{H})$, then

$$E_\mu \circ \mathcal{E}_p(x) = (1-p)E_\mu(x) + pE_\mu(|e_{d+1}\rangle\langle e_{d+1}|) \tag{715}$$

$$= (1-p)(1-\mu) \cdot x + \mu(1-p) \cdot |e_{d+1}\rangle\langle e_{d+1}| + p|e_{d+1}\rangle\langle e_{d+1}| \tag{716}$$

$$= (1-p-1+2p) \cdot x + (1-2p)|e_{d+1}\rangle\langle e_{d+1}| + p|e_{d+1}\rangle\langle e_{d+1}| \tag{717}$$

$$= p \cdot x + (1-p) \cdot |e_{d+1}\rangle\langle e_{d+1}| \tag{718}$$

$$= \mathcal{E}_{1-p}(x). \tag{719}$$

D.: Let $\rho_1 := |e_1\rangle\langle e_1|$, $\rho_2 := |e_2\rangle\langle e_2| \in \mathcal{S}(\mathcal{H})$. We show by contradiction that there are no two probability distributions $r_1, r_2 \in \mathfrak{P}(\mathbf{S})$ such that

$$\sum_{s\in\mathbf{S}} r_1(s)\mathcal{E}_{p_s}(\rho_1) = \sum_{s\in\mathbf{S}} r_2(s)\mathcal{E}_{p_s}(\rho_2). \tag{720}$$

Assume there are $r_1, r_2 \in \mathfrak{P}(\mathbf{S})$ such that (720) is true. This is equivalent to

$$\sum_{s\in\mathbf{S}}(1-p_s)[r_1(s)|e_1\rangle\langle e_1| - r_2(s)|e_2\rangle\langle e_2|] = 0, \qquad \sum_{s\in\mathbf{S}} p_s[r_1(s) - r_2(s)] \cdot |e_{d+1}\rangle\langle e_{d+1}| = 0. \tag{721}$$

By linear independence of $|e_1\rangle\langle e_1|, |e_2\rangle\langle e_2|$ and since $p_s \in [0,1)$ for every $s \in \mathbf{S}$ the first equality implies $r_1(s) = r_2(s) = 0\ \forall s \in \mathbf{S}$, in clear contradiction to the assumption $r_1, r_2 \in \mathfrak{P}(\mathbf{S})$.
Thus, $\{\mathcal{E}_{p_s}\}_{s\in\mathbf{S}}$ with all $p_s \in [0,1)$ is non-symmetrizable.

Using **A** and the fact that $\{p_s\}_{s\in\mathbf{S}} \subset [0,1/2)$ we see that for an arbitrary $q \in \mathfrak{P}(\mathbf{S})$ we have $\sum_{s\in\mathbf{S}} q(s)\mathcal{E}_{p_s} = \mathcal{E}_{q(p)}$ with $q(p) \in [0,1/2)$.
Now **B** implies that for every $q \in \mathfrak{P}(\mathbf{S})$ the channel $\sum_{s\in\mathbf{S}} q(s)\mathcal{E}_{p_s}$ is degradable.
Thus by Lemma 112, *2.*, the regularization in the identity

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) = \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}) \tag{722}$$

is not necessary, so

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) = \max_{\rho\in\mathcal{S}(\mathcal{H})} \inf_{\mathcal{N}\in\text{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}). \tag{723}$$

Further, for a fixed degradable channel, the coherent information is concave in the input state [74] and thus by the minimax theorem for concave-convex functions [44, 67] we can interchange min and max in (723). Now, to any given $\rho \in \mathcal{S}(\mathcal{H})$, we may write $\rho = \sum_{i=1}^d \lambda_i |v_i\rangle\langle v_i|$ for some set $\{v_1, \ldots, v_d\}$ of orthonormal vectors that satisfy, by standard identification of $\mathbb{C}^d$ and $\mathbb{C}^{d+1}$, $v_i \perp e_{d+1}$ $(1 \le i \le d)$ and write a purification of ρ as $|\psi_\rho\rangle\langle\psi_\rho| = \sum_{i,j=1}^d \lambda_i\lambda_j |v_i\rangle\langle v_j| \otimes |v_i\rangle\langle v_j|$. Then for every $\mathcal{E}_p \in \text{conv}(\mathfrak{I})$ we have

$$\begin{aligned}
\max_{\rho\in\mathcal{S}(\mathcal{H})} I_c(\rho, \mathcal{E}_p) &= \max_{\rho\in\mathcal{S}(\mathcal{H})} (S(\mathcal{E}_p(\rho)) - S(Id_{\mathcal{H}} \otimes \mathcal{E}_p(|\psi_\rho\rangle\langle\psi_\rho|)) && (724)\\
&= \max_{\rho\in\mathcal{S}(\mathcal{H})} (S((1-p)\rho + p|e_{d+1}\rangle\langle e_{d+1}|) - S((1-p)|\psi_\rho\rangle\langle\psi_\rho| + p\rho\otimes|e_{d+1}\rangle\langle e_{d+1}|)) && (725)\\
&= \max_{\rho\in\mathcal{S}(\mathcal{H})} ((1-p)S(\rho) + pS(|e_{d+1}\rangle\langle e_{d+1}|) + H(p) && (726)\\
&\quad - (1-p)S(|\psi_\rho\rangle\langle\psi_\rho|) - pS(\rho\otimes|e_{d+1}\rangle\langle e_{d+1}|) - H(p)) && (727)\\
&= \max_{\rho\in\mathcal{S}(\mathcal{H})} ((1-p)S(\rho) - pS(\rho\otimes|e_{d+1}\rangle\langle e_{d+1}|)) && (728)\\
&= \max_{\rho\in\mathcal{S}(\mathcal{H})} (1-2p)S(\rho) && (729)\\
&= (1-2p)\log(d). && (730)
\end{aligned}$$

This leads to

$$\mathcal{A}_{\text{random}}(\mathfrak{I}) = \min_{s\in\mathbf{S}}(1-2p_s)\log(d), \tag{731}$$

a formula that was first discovered for the case of a single memoryless channel and $d = 2$ by [13].
From **D** it follows that $\mathcal{A}_{\text{det}}(\mathfrak{I}) = \mathcal{A}_{\text{random}}(\mathfrak{I})$.
We can now prove *1.*: Set $p_{\max} := \max_{s\in\mathbf{S}} p_s$. It holds $p_{\max} \ge 1/2$, therefore the channel $\mathcal{E}_{\max} := \mathcal{E}_{p_{\max}}$ satisfies $\mathcal{A}_{\text{det}}(\{\mathcal{E}_{\max}\}) = \mathcal{A}_{\text{random}}(\{\mathcal{E}_{\max}\}) = 0$, since by **B** and **C** $\widehat{\mathcal{E}_{\max}}$ is degradable, hence $\mathcal{E}_{\max}$ is anti-degradable. Thus, for every $l \in \mathbb{N}$, the adversary can always choose $\mathcal{E}_{\max}^{\otimes l}$ to ensure that transmission of entanglement will fail. □

4.8.2 Qualitative behavior of zero-error capacities

Let us, first, embark on the connection between AVQCs and zero-error capacities. Classical information theory exhibits an interesting connection between the zero-error capacity of stationary memoryless channels and the deterministic capacity with asymptotically vanishing maximal error probability criterion of certain arbitrarily varying channels. This connection was first described by Ahlswede in [1].
Ahlsedes result can be formulated using the following notation. For two finite sets $\mathbf{A}, \mathbf{B}$, $C(\mathbf{A}, \mathbf{B})$ stands for the set of channels from $\mathbf{A}$ to $\mathbf{B}$, i.e. each element of $W \in C(\mathbf{A}, \mathbf{B})$ defines a set of output probability distributions $\{W(\cdot|a)\}_{a\in\mathbf{A}}$. With slight abuse of notation, for each $D \subset \mathbf{B}$ and $a \in \mathbf{A}$, $W(D|a) := \sum_{b\in D} W(b|a)$. The (finite) set of extremal points of the (convex) set $C(\mathbf{A}, \mathbf{B})$ will be written $E(\mathbf{A}, \mathbf{B})$.
For two channels $W_1, W_2 \in C(\mathbf{A}, \mathbf{B})$, their product $W_1 \otimes W_2 \in C(\mathbf{A}^2, \mathbf{B}^2)$ is defined through $(W_1 \otimes W_2)(b^2|a^2) := W_1(b_1|a_1)W_2(b_2|a_2)$. An arbitrarily varying channel (AVC) is, in this setting, defined through a set $\mathbb{W} = \{W_s\}_{s\in\mathbf{S}} \subset C(\mathbf{A}, \mathbf{B})$ (we assume $\mathbf{S}$ and, hence, $|\mathbb{W}|$, to be finite). The different realizations of the channel are written

$$W_{s^l} := W_{s_1} \otimes \ldots \otimes W_{s_l} \qquad (s^l \in \mathbf{S}^l) \tag{732}$$

and, formally, the AVC $\mathbb{W}$ consists of the set $\{W_{s^l}\}_{s^l\in\mathbf{S}^l,\ l\in\mathbb{N}}$.
An (l, M_l)-code for the AVC $\mathbb{W}$ is given by a set $\{a_i^l\}_{i=1}^{M_l} \subset \mathbf{A}^l$ called the 'codewords' and a set $\{D_i^l\}_{i=1}^{M_l}$ of subsets of $\mathbf{B}^l$ called the 'decoding sets', that satiesfies $D_i^l \cap D_j^l = \emptyset,\ i \neq j$.
A nonnegative number $R \in \mathbb{R}$ is called an achievable maximal-error rate for the AVC $\mathbb{W}$, if there exists a sequence of (l, M_l) codes for $\mathbb{W}$ such that both

$$\liminf_{l\to\infty} \frac{1}{l} \log M_l \geq R \qquad \text{and} \qquad \lim_{l\to\infty} \min_{s^l\in\mathbf{S}^l} \min_{1\leq i\leq M_l} W_{s^l}(D_i^l|x_i^l) = 1. \tag{733}$$

The (deterministic) maximal error capacity $C_{\max}(\mathbb{W})$ of the AVC $\mathbb{W}$ is, as usually, defined as the supremum over all achievable maximal-error rates for $\mathbb{W}$.
Much stronger requirements concerning the quality of codes can be made. An (l, M_l)-code is said to have zero error for the AVC $\mathbb{W}$, if for all $1 \leq i \leq M_l$ and $s^l \in \mathbf{S}^l$ the equality $W_{s^l}(D_i^l|x_i^l) = 1$ holds.
The zero error capacity $C_0(\mathbb{W})$ of the AVC $\mathbb{W}$ is defined as

$$C_0(\mathbb{W}) := \lim_{l\to\infty} \max\{\frac{1}{l} \log M_l : \exists\ (l, M_l)-\text{code with zero error for } \mathbb{W}\}. \tag{734}$$

The above definitions carry over to single channels $W \in C(\mathbf{A}, \mathbf{B})$ by identifying W with the set $\{W\}$.
In short form, the connection [1, Theorem 3] between the capacity of certain arbitrarily varying channels and the zero-error capacity of stationary memoryless channels can now be reformulated as follows:

Theorem 114. *Let $W \in C(\mathbf{A}, \mathbf{B})$ have a decomposition $W = \sum_{s\in\mathbf{S}} q(s)W_s$, where $\{W_s\}_{s\in\mathbf{S}} \subset E(\mathbf{A}, \mathbf{B})$ and $q(s) > 0\ \forall s \in \mathbf{S}$. Then for the AVC $\mathbb{W} := \{W_s\}_{s\in\mathbf{S}}$:*

$$C_0(W) = C_{\max}(\mathbb{W}). \tag{735}$$

Conversely, for every AVC $\mathbb{W} = \{W_s\}_{s\in\mathbf{S}} \subset E(\mathbf{A}, \mathbf{B})$ and every $q \in \mathfrak{P}(\mathbf{S})$ with $q(s) > 0\ \forall s \in \mathbf{S}$, equation (735) holds for the channel $W := \sum_{s\in\mathbf{S}} q(s)W_s$.

Remark 115. *Let us note at this point, that the original formulation of the theorem did not make reference to extremal points of the set of channels, but rather used the equivalent notion "channels of $0-1$-type".*

Remark 116. *By choosing $W \in E(\mathbf{A}, \mathbf{B})$, one gets the equality $C_0(W) = C_{max}(W)$. The quantity $C_{max}(W)$ being well-known and easily computable, it may seem that Theorem 114 solves Shannons's zero-error problem. This is not the case, as one can verify by looking at the famous pentagon channel that was introduced in [63, Figure 2.]. The pentagon channel is far from being extremal. That its zero-error capacity is positive [63] is due to the fact that it is not a member of the relative interior $riE(\mathbf{A}, \mathbf{B})$.*

We can now formulate a straightforward analogy of Theorem 114 for quantum channels:

Conjecture 117. *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ have a decomposition $\mathcal{N} = \sum_{s\in\mathbf{S}} q(s)\mathcal{N}_s$, where each $\mathcal{N}_s$ is extremal in $\mathcal{C}(\mathcal{H}, \mathcal{K})$ and $q(s) > 0\ \forall s \in \mathbf{S}$. Then for the AVQC $\mathfrak{I} := \{\mathcal{N}_s\}_{s\in\mathbf{S}}$:*

$$Q_0(\mathcal{N}) = \mathcal{A}_{\mathrm{s,det}}(\mathfrak{I}). \tag{736}$$

Conversely, for every AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ with $\mathcal{N}_s$ being extremal for every $s \in \mathbf{S}$ and every $q \in \mathfrak{P}(\mathbf{S})$ with $q(s) > 0\ \forall s \in \mathbf{S}$, equation (736) holds for the channel $\mathcal{N} := \sum_{s\in\mathbf{S}} q(s)\mathcal{N}_s$.

Remark 118. *One could formulate weaker conjectures than the one above. A crucial property of extremal classical channels that was used in the proof of Theorem 114 was that $W_{s^l}(\cdot|x_i^l)$ is a point-measure for every codeword x_i^l, if only $\{W_{s^l}\}_{s\in\mathbf{S}} \subset E(\mathbf{A}, \mathbf{B})$.*
This property gets lost for the extremal points of $\mathcal{C}(\mathcal{H}, \mathcal{K})$ (see the channels that are used in the proof of Theorem 119), but could be regained by restriction to channels consisting of only one single Kraus operator.

This conjecture leads us to the following theorem:

Theorem 119. *Conjecture 117 is wrong.*

Remark 120. *As indicated in Remark 118, there could still be interesting connections between (for example) the deterministic strong subspace transmission capacity of AVQCs and the zero-error entanglement transmission of stationary memoryless quantum channels.*

Proof. Let $\mathcal{H} = \mathcal{K} = \mathbb{C}^2$. Let $\{e_0, e_1\}$ be the standard basis of $\mathbb{C}^2$. Consider, for a fixed but arbitrary $x \in [0, 1]$ the channel $\mathcal{N}_x \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ defined by Kraus operators $A_1 := \sqrt{1 - x^2}|e_0\rangle\langle e_1|$ and $A_2 := |e_0\rangle\langle e_0| + x|e_1\rangle\langle e_1|$. As was shown in [71], this channel is extremal in $\mathcal{C}(\mathcal{H}, \mathcal{K})$. It is also readily seen from the definition of Kraus operators, that it approximates the identity channel $id_{\mathbb{C}^2} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$:

$$\lim_{x\to 1} \|\mathcal{N}_x - id_{\mathbb{C}^2}\|_\diamond = 0. \tag{737}$$

Now, on the one hand, $\mathcal{N}_x$ being extremal implies $\mathrm{span}(\{A_i^* A_j\}_{i,j=1}^2) = M(\mathbb{C}^2)$ for all $x \in [0, 1)$ (where $M(\mathbb{C}^2)$ denotes the set of complex 2×2 matrices) by [19, Theorem 5]. This carries over to the channels $\mathcal{N}_x^{\otimes l}$ for every $l \in \mathbb{N}$: Let the Kraus operators of $\mathcal{N}_x^{\otimes l}$ be denoted $\{A_{i^l}\}_{i^l\in\{1,2\}^l}$, then

$$\mathrm{span}(\{A_{i^l}^* A_{j^l}\}_{i^l,j^l\in\{1,2\}^l}) = \{M : M \text{ is complex } 2^l \times 2^l\text{−matrix}\}. \tag{738}$$

On the other hand, it was observed e.g. in [27], that for two pure states $|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi| \in \mathcal{S}((\mathbb{C}^2)^{\otimes l})$, the subspace spanned by them can be transmitted with zero error only if

$$|\psi\rangle\langle\phi| \perp \mathrm{span}(\{A_{i^l}^* A_{j^l}\}_{i^l,j^l\in\{1,2\}^l}). \tag{739}$$

This is in obvious contradiction to equation (738), therefore $Q_0(\mathcal{N}_x) = 0\ \forall x \in [0, 1)$.
On the other hand, from equation (737) and continuity of coherent information in the channel [51] we see that there is an $X \in [0, 1)$ such that for all $x \geq X$ we have $Q(\mathcal{N}_x) > 0$. Letting $x = X$ we obtain $Q_0(\mathcal{N}_X) = 0$ and $Q(\mathcal{N}_X) > 0$, so $Q_0(\mathcal{N}_X) \neq Q(\mathcal{N}_X)$ in contradiction to the statement of the conjecture. □

We now show (following closely the lines of [1]) what remains of Theorem 114 in the quantum case: Let $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ be a finite AVQC. Consider

$$\mathcal{N}_\mathfrak{I} := \frac{1}{|\mathbf{S}|} \sum_{s\in\mathbf{S}} \mathcal{N}_s. \tag{740}$$

By definition of zero-error capacity, to any $\delta > 0$ there exists an $l \in \mathbb{N}$, a maximally mixed state $\pi_{\mathcal{F}_l}$ with $\frac{1}{l} \log \dim \mathcal{F}_l \geq Q_0(\mathcal{N}_\mathfrak{I}) - \delta$ and a pair $(\mathcal{R}^l, \mathcal{P}^l)$ of recovery and encoding map such that

$$\min_{x \in \mathcal{F}_l, ||x||=1} \langle x, \mathcal{R}^l \circ \mathcal{N}_\mathfrak{I}^{\otimes l} \circ \mathcal{P}(|x\rangle\langle x|)x\rangle = 1 \tag{741}$$

holds. But this directly implies

$$\min_{x \in \mathcal{F}_l, ||x||=1} \langle x, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}(|x\rangle\langle x|)x\rangle = 1 \qquad \forall s^l \in \mathbf{S}^l, \tag{742}$$

so $(\pi_{\mathcal{F}_l}, \mathcal{R}^l, \mathcal{P}^l)$ gives a zero-error code for the AVQC $\mathfrak{I}$ as well and therefore

$$\mathcal{A}_{\det}(\mathfrak{I}) \geq Q_0(\mathcal{N}_\mathfrak{I}) - \delta \qquad \forall \delta > 0. \tag{743}$$

This in turn is equivalent to

$$\mathcal{A}_{\det}(\mathfrak{I}) \geq Q_0(\mathcal{N}_\mathfrak{I}). \tag{744}$$

One may now ask when exactly this is a meaningful (nonzero) lower bound. The answer is given by the proof of Lemma 124: On any face of $\mathcal{C}(\mathcal{H}, \mathcal{K})$ the zero-error capacities are constant and the encoding and recovery maps are *universal*. Thus, if $\mathfrak{I}$ is a subset of a face and $\mathfrak{I} \subset \mathrm{ri}(\mathcal{C}(\mathcal{H}, \mathcal{K}))^{\complement}$ then there is good hope to get a nonzero lower bound by means of inequality (743). So far for the connection between AVQCs and zero-error capacities.

Motivated by the above observation, a closer study of zero-error capacities reveals some additional facts that are interesting in their own right.

To be more precise, we investigate continuity of zero-error capacities. This property is a highly desirable property both from the practical and the theoretical point of view. It is of particular importance in situations where full knowledge of the communication system cannot be achieved but only a narrow confidence set containing the unknown channel is given. In [51] it has been shown that the ordinary capacities of stationary memoryless quantum channels are continuous in the finite-dimensional setting and it was demonstrated by examples that these functions become discontinuous in infinite dimensional situations.

In this section we show that quantum, entanglement-assisted, and classical zero-error capacities of quantum channels are discontinuous at every positivity point. Our approach is based on two simple observations. The first one is that the zero-error capacities mentioned above of each quantum channel belonging to the relative interior of the set of quantum channels are equal to 0. The second one is the well known fact that the relative interior of any convex set is open and dense in that set, i.e. generic. Hence any channel can be approximated by a sequence belonging to the relative interior implying the discontinuity result.

Similar arguments can be applied to the recently introduced Lovász $\tilde{\theta}$ function and zero-error distillable entanglement as well, leading to analogous conclusions as shall be shown in the last part of this section.

We now show that all the zero-error capacities defined in subsection 4.1.3 are generically equal to 0 and are discontinuous at any positivity point. Then we demonstrate that the zero-error capacities of quantum channels can be thought of as step functions subordinate to the partition built from the relative interiors of the faces of $\mathcal{C}(\mathcal{H}, \mathcal{K})$.

Discontinuity of zero-error capacities

Theorem 121. *Let $\mathcal{N} \in \mathrm{ri}\,\mathcal{C}(\mathcal{H}, \mathcal{K})$. Then $k(l, \mathcal{N}) = M(l, \mathcal{N}) = M_{\mathrm{EA}}(l, \mathcal{N}) = 1$ for every $l \in \mathbb{N}$. Consequently, $Q_0(\mathcal{N}) = C_0(\mathcal{N}) = C_{0\mathrm{EA}}(\mathcal{N}) = 0$.*

In the proof of Theorem 121 we shall make use of the following elementary fact:

Lemma 122. *Let F be a non-empty convex set and $\mathcal{N}_0, \mathcal{N} \in \mathrm{ri}\, F$ with $\mathcal{N}_0 \neq \mathcal{N}$. Then there exists $\mathcal{N}_1 \in F$ and $\lambda_0, \lambda_1 \in (0, 1)$, $\lambda_0 + \lambda_1 = 1$ with $\mathcal{N} = \lambda_0 \mathcal{N}_0 + \lambda_1 \mathcal{N}_1$.*

Proof of Lemma 122. Since $\mathcal{N} \in \operatorname{ri} F$ there is $\mu' > 1$ such that

$$\mathcal{N}_1 := (1-\mu')\mathcal{N}_0 + \mu'\mathcal{N} \in F. \tag{745}$$

We define now

$$\lambda_1 := \frac{1}{\mu'} \in (0,1), \quad \lambda_0 := 1 - \lambda_1, \tag{746}$$

and obtain using $\mathcal{N}_1$ given in (745) the desired convex decomposition

$$\mathcal{N} = \lambda_0\mathcal{N}_0 + \lambda_1\mathcal{N}_1. \tag{747}$$

□

Proof of Theorem 121. Let $\mathcal{N} \in \operatorname{ri}\mathcal{C}(\mathcal{H},\mathcal{K})$. Observing that the fully depolarizing channel $\mathcal{N}_0(a) = \frac{\operatorname{tr}(a)}{d_\mathcal{K}}\mathbf{1}_\mathcal{K}$, $a \in \mathcal{B}(\mathcal{H})$, belongs to $\operatorname{ri}\mathcal{C}(\mathcal{H},\mathcal{K})$ we obtain from Lemma 122 a convex decomposition of $\mathcal{N}$ as

$$\mathcal{N} = \lambda_0\mathcal{N}_0 + \lambda_1\mathcal{N}_1, \tag{748}$$

where $\lambda_0, \lambda_1 \in (0,1)$, $\lambda_0 + \lambda_1 = 1$.
Clearly, this decomposition implies that

$$\mathcal{N}^{\otimes l} = \sum_{s^l \in \{0,1\}^l} \lambda_{s^l}\mathcal{N}_{s^l}, \tag{749}$$

with $\lambda_{s^l} := \lambda_{s_1} \cdot \ldots \cdot \lambda_{s_l} > 0$ and $\mathcal{N}_{s^l} := \mathcal{N}_{s_1} \otimes \ldots \otimes \mathcal{N}_{s_l}$ for all $s^l \in \{0,1\}^l$. Then for any zero-error (l,M) ea-code $(\sigma_{\mathcal{F}\mathcal{F}'}, \{\mathcal{P}_m, D_m\}_{m=1}^M)$ for $\mathcal{N}$ we get for each $m \in [M]$

$$\begin{aligned} 1 &= \operatorname{tr}((\mathcal{N}^{\otimes l} \circ \mathcal{P}_m \otimes \mathrm{id}_{\mathcal{F}'})(\sigma_{\mathcal{F}\mathcal{F}'})D_m) && (750)\\ &= \sum_{s^l \in \{0,1\}^l} \lambda_{s^l} \operatorname{tr}((\mathcal{N}_{s^l} \circ \mathcal{P}_m \otimes \mathrm{id}_{\mathcal{F}'})(\sigma_{\mathcal{F}\mathcal{F}'})D_m) && (751)\end{aligned}$$

and, consequently, since $\lambda_{s^l} > 0$ for all $s^l \in \{0,1\}^l$

$$\operatorname{tr}((\mathcal{N}_{s^l} \circ \mathcal{P}_m \otimes \mathrm{id}_{\mathcal{F}'})(\sigma_{\mathcal{F}\mathcal{F}'})D_m) = 1 \qquad \forall s^l \in \{0,1\}^l, \tag{752}$$

for all $m \in [M]$. Choosing $\bar{s}^l = (0,\ldots,0)$ we obtain from Eqn. (752) that $(\sigma_{\mathcal{F}\mathcal{F}'}, \{\mathcal{P}_m, D_m\}_{m=1}^M)$ is a zero-error ea-code for $\mathcal{N}_0$. Since $M_{\mathrm{EA}}(l,\mathcal{N}_0) = 1$ for all $l \in \mathbb{N}$ we can conclude that $M_{\mathrm{EA}}(l,\mathcal{N}) \leq 1$ and thus $M_{\mathrm{EA}}(l,\mathcal{N}) = 1$ holds. Consequently $C_{0\mathrm{EA}}(\mathcal{N}) = 0$. The other assertions follow from the observation that $1 \leq k(l,\mathcal{N}) \leq M(l,\mathcal{N}) \leq M_{\mathrm{EA}}(l,\mathcal{N})$. □

Corollary 123. *The function $Q_0 : \mathcal{C}(\mathcal{H},\mathcal{K}) \to \mathbb{R}_+$ that assigns the zero-error quantum capacity to each quantum channel is discontinuous at any $\mathcal{N} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ with $Q_0(\mathcal{N}) > 0$. The same conclusion holds true for C_0 and $C_{0\mathrm{EA}}$.*

Proof. If $Q_0(\mathcal{N}) > 0$ holds then necessarily $\mathcal{N} \in \operatorname{rebd}\mathcal{C}(\mathcal{H},\mathcal{K})$ by Theorem 121. On the other hand $\operatorname{ri}\mathcal{C}(\mathcal{H},\mathcal{K})$ is dense in $\mathcal{C}(\mathcal{H},\mathcal{K})$ (cf. Theorem 2.3.8 in [68]). So there is a sequence of channels $(\mathcal{N}_i)_{i\in\mathbb{N}} \subset \operatorname{ri}\mathcal{C}(\mathcal{H},\mathcal{K})$ with $\lim_{i\to\infty} ||\mathcal{N}_i - \mathcal{N}||_\diamond = 0$ and by Theorem 121 we have $Q_0(\mathcal{N}_i) = 0$ for all $i \in \mathbb{N}$. The arguments for C_0 and $C_{0\mathrm{EA}}$ follow the same line of reasoning. □

Relation to the facial structure of the set of quantum channels Here we shall show that the considered zero-error capacities are basically step functions, the underlying partition consisting of the relative interiors of the faces of $\mathcal{C}(\mathcal{H},\mathcal{K})$.

Lemma 124. *Let $F \subset \mathcal{C}(\mathcal{H},\mathcal{K})$ be convex and let $\tilde{\mathcal{N}} \in \operatorname{ri} F$. Then for any $\mathcal{N} \in \operatorname{ri} F$, $Q_0(\mathcal{N}) = Q_0(\tilde{\mathcal{N}})$, $C_{0\mathrm{EA}}(\mathcal{N}) = C_{0\mathrm{EA}}(\tilde{\mathcal{N}})$, and $C_0(\mathcal{N}) = C_0(\tilde{\mathcal{N}})$ hold.*

Proof. We assume w.l.o.g. that $\mathcal{N} \neq \tilde{\mathcal{N}}$ to avoid trivialities. Then setting $\mathcal{N}_0 := \mathcal{N}$ we can find $\mathcal{N}_1 \in F$ and $\lambda_0, \lambda_1 \in (0,1), \lambda_0 + \lambda_1 = 1$, with

$$\tilde{\mathcal{N}} = \lambda_0 \mathcal{N}_0 + \lambda_1 \mathcal{N}_1 \tag{753}$$

just by applying Lemma 122 to $\mathcal{N}_0, \tilde{\mathcal{N}} \in \operatorname{ri} F$.
Let $(\mathcal{F}_l, \mathcal{P}, \mathcal{R})$ be an (l, k_l) zero-error quantum code for $\tilde{\mathcal{N}}$. Then using the representation (753) we obtain for any $x \in \mathcal{F}_l, ||x|| = 1$

$$\begin{aligned} 1 &= \langle x, \mathcal{R} \circ \tilde{\mathcal{N}}^{\otimes l} \circ \mathcal{P}(|x\rangle\langle x|)x\rangle & (754)\\ &= \sum_{s^l \in \{0,1\}^l} \lambda_{s^l} \langle x, \mathcal{R} \circ \mathcal{N}_{s^l} \circ \mathcal{P}(|x\rangle\langle x|)x\rangle & (755) \end{aligned}$$

and consequently, since $\lambda_{s^l} > 0$ for all $s^l \in \{0,1\}^l$, we are led to

$$\langle x, \mathcal{R} \circ \mathcal{N}_{s^l} \circ \mathcal{P}(|x\rangle\langle x|)x\rangle = 1 \tag{756}$$

for all $s^l \in \{0,1\}^l$ and all $x \in \mathcal{F}_l, ||x|| = 1$. Choosing the sequence $s^l = (0,\ldots,0)$ and recalling that $\mathcal{N}_0 = \mathcal{N}$ we arrive at

$$Q_0(\mathcal{N}) \geq Q_0(\tilde{\mathcal{N}}). \tag{757}$$

The reverse inequality is derived by interchanging the roles of $\mathcal{N}$ and $\tilde{\mathcal{N}}$. The remaining assertions are shown in the same vein. □

We shall now pass to the set of faces $\mathfrak{F} := \{F : \text{face of } \mathcal{C}(\mathcal{H},\mathcal{K})\}$ of $\mathcal{C}(\mathcal{H},\mathcal{K})$.

Theorem 125. *To each $\mathcal{N} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ there is a unique $F \in \mathfrak{F}$ with $\mathcal{N} \in \operatorname{ri} F$. Moreover, each of the capacity functions Q_0, $C_{0\mathrm{EA}}$, and C_0 is constant on $\operatorname{ri} F$.*

Proof. According to Theorem 2.6.10 in [68] the family of sets $\{\operatorname{ri} F : F \in \mathfrak{F}\}$ forms a partition of $\mathcal{C}(\mathcal{H},\mathcal{K})$. This shows the first assertion of the theorem. The second follows from Lemma 124. □

Remark 126. *Notice that the results obtained so far show that the optimal (i.e. capacity achieving) code for any channel $\mathcal{N}$ in the relative interior of any face F of $\mathcal{C}(\mathcal{H},\mathcal{K})$ is also optimal for any other channel in ri F.*

4.8.3 Discontinuity of quantum Lovász $\tilde{\theta}$ function & zero-error distillable entanglement

In this final subsection we show that our methods are not only bound to the zero-error capacities of quantum channels. They apply to the quantum Lovász $\tilde{\theta}$ function from [27] and also to zero-error distillable entanglement.

Discontinuity of quantum Lovász $\tilde{\theta}$ function Preliminarily, following [27], for a given channel $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ with a corresponding set of Kraus operators $\{E_j\}_{j\in[K]}$ we define the non-commutative confusability graph following [27] by

$$\begin{aligned} S(\mathcal{N}) &:= \operatorname{span}\{E_j^* E_i : i, j \in [K]\} && (758)\\ &= \hat{\mathcal{N}}_*(\mathcal{B}(\mathcal{E})), && (759)\end{aligned}$$

where $\hat{\mathcal{N}}_*$ is the adjoint of the complementary channel $\hat{\mathcal{N}} \in \mathcal{C}(\mathcal{H}, \mathcal{E})$ defined via the Stinespring isometry $V : \mathcal{H} \to \mathcal{K} \otimes \mathcal{E}$

$$Vx := \sum_{j=1}^{K} E_j x \otimes f_j \tag{760}$$

with an ONB $\{f_1, \ldots, f_K\}$ in $\mathcal{E}$.
Also, let us recite their definition of the quantum Lovász $\tilde{\theta}$ function and its most fundamental property:

Definition 127 (Quantum Lovász $\tilde{\theta}$ function)**.** *The quantum Lovász $\tilde{\theta}$ function is, for a given confusability graph S defined by*

$$\tilde{\theta}(S) := \sup_{n\in\mathbb{N}} \max\{\|\mathbf{1}_{\mathcal{H}\otimes\mathbb{C}^n} + T\| : T \in S^\perp \otimes \mathcal{B}(\mathbb{C}^n),\ \mathbf{1} + T \geq 0,\ T = T^*\}, \tag{761}$$

where $S^\perp := \{a \in \mathcal{B}(\mathcal{H}) : \operatorname{tr}(ab) = 0\ \forall b \in S\}$.

This function gives an upper bound on the entanglement-assisted capacity for transmission of classical messages with zero error: For every channel $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$:

$$C_{0EA}(\mathcal{N}) \leq \log \tilde{\theta}(S(\mathcal{N})) \tag{762}$$

(Lemma 7 and Corollary 10 in [27]). It can be characterised as a semidefinite program ([27], Theorem 8). We are going to employ the dual formulation:

Theorem 128 (Theorem 9 in [27])**.** *For any $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ we have*

$$\tilde{\theta}(S(\mathcal{N})) = \min\left\{||\operatorname{tr}_{\mathcal{H}} Y|| : Y \in S(\mathcal{N}) \otimes \mathcal{B}(\mathcal{H}'), Y \geq |\Phi\rangle\langle\Phi|\right\}, \tag{763}$$

where $\mathcal{H}'$ is just a copy of $\mathcal{H}$ and $\Phi = \sum_{i=1}^{\dim \mathcal{H}} e_i \otimes e_i'$ with ONBs $\{e_1, \ldots, e_{\dim \mathcal{H}}\}$ and $\{e_1', \ldots, e_{\dim \mathcal{H}}'\}$ of $\mathcal{H}$ and $\mathcal{H}'$.

In the following we shall also need the next simple lemma.

Lemma 129. *Let $\mathcal{N} \in \operatorname{ri}\mathcal{C}(\mathcal{H}, \mathcal{K})$. Then $S(\mathcal{N}) = \mathcal{B}(\mathcal{H})$.*

Proof. Again we can represent $\mathcal{N}$ as

$$\mathcal{N} = \lambda_0 \mathcal{N}_0 + \lambda_1 \mathcal{N}_1 \tag{764}$$

with $\lambda_0, \lambda_1 \in (0, 1)$, $\lambda_0 + \lambda_1 = 1$, $\mathcal{N}_0$ being the fully depolarizing channel, and $\mathcal{N}_1 \in \mathcal{C}(\mathcal{H}, \mathcal{K})$. The proof is concluded by the following simple observation: Given any two channels $\mathcal{N}_0, \mathcal{N}_1$ and $\lambda_0, \lambda_1 \in (0, 1)$ with $\lambda_0 + \lambda_1 = 1$. Then for the channel $\mathcal{N} := \lambda_0 \mathcal{N}_0 + \lambda_1 \mathcal{N}_1$ it holds that

$$S(\mathcal{N}) \supseteq S(\mathcal{N}_0), S(\mathcal{N}_1). \tag{765}$$

Since in our case $S(\mathcal{N}_0) = \mathcal{B}(\mathcal{H})$ we are done. □

fWith Theorem 128 and Lemma 129 at our disposal we can deduce the following discontinuity result for $\tilde{\theta}$:

Theorem 130. *The function $\tilde{\theta} : \mathcal{C}(\mathcal{H}, \mathcal{H}) \to \mathbb{R}_+$ assigning the number $\tilde{\theta}(S(\mathcal{N}))$ to each quantum channel $\mathcal{N}$ is discontinuous at any $\mathcal{N}$ with $C_{0\mathrm{EA}}(\mathcal{N}) > 0$.*

Proof. Note that for $\mathcal{N} \in \mathrm{ri}\,\mathcal{C}(\mathcal{H}, \mathcal{K})$ $S(\mathcal{N}) = \mathcal{B}(\mathcal{H})$ by Lemma 129. Hence $|\Phi\rangle\langle\Phi| \in S(\mathcal{N}) \otimes \mathcal{B}(\mathcal{H}') = \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H})$ and $||\mathrm{tr}_{\mathcal{H}}|\Phi\rangle\langle\Phi|| = 1 = \tilde{\theta}(S(\mathcal{N}))$. On the other hand, (762) implies that for any $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$, $\tilde{\theta}(S(\mathcal{N})) > 1$ if $C_{0\mathrm{EA}}(\mathcal{N}) > 0$.
Since $\mathrm{ri}\,\mathcal{C}(\mathcal{H}, \mathcal{K})$ is dense in $\mathcal{C}(\mathcal{H}, \mathcal{K})$ and since $\tilde{\theta}(S(\mathcal{N})) = 1$ for each $\mathcal{N} \in \mathrm{ri}\,\mathcal{C}(\mathcal{H}, \mathcal{K})$ we are done. □

Notice that the arguments given for the Lovász $\tilde{\theta}$ function apply to any other upper bound to the entanglement-assisted zero-error capacity vanishing in the relative interior of $\mathcal{C}(\mathcal{H}, \mathcal{K})$.

Zero-error distillation of entanglement The simple methods employed so far can also be applied to the problem of zero-error distillation of entanglement as we shall briefly indicate below. Assuming that $\rho \in \mathrm{ri}\,\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ we can find $\lambda_0, \lambda_1 \in (0, 1)$, $\lambda_0 + \lambda_1 = 1$ such that

$$\rho = \lambda_0 \rho_0 + \lambda_1 \rho_1, \tag{766}$$

with $\rho_0 = \frac{1}{d_A}\mathbf{1}_{\mathcal{H}_A} \otimes \frac{1}{d_B}\mathbf{1}_{\mathcal{H}_B} \in \mathrm{ri}\,\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $d_A = \dim \mathcal{H}_A, d_B = \dim \mathcal{H}_B$, and $\rho_1 \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then

$$\rho^{\otimes l} = \sum_{s^l \in \{0,1\}^l} \lambda_{s^l} \rho_{s^l}, \tag{767}$$

and for any (l, k_l) zero-error EDP $(\mathcal{D}, \varphi_{k_l})$ for ρ we obtain

$$1 = \sum_{s^l \in \{0,1\}^l} \lambda_{s^l} \langle \varphi_{k_l}, \mathcal{D}(\rho_{s^l})\varphi_{k_l}\rangle, \tag{768}$$

leading to

$$1 = \langle \varphi_{k_l}, \mathcal{D}(\rho_{s^l})\varphi_{k_l}\rangle \tag{769}$$

for all $s^l \in \{0, 1\}^l$. Choosing $s^l = (0, \ldots, 0)$ and noting that due to the fact that $\mathcal{D}$ is a LOCC operation the state $\mathcal{D}(\rho_0^{\otimes l})$ is separable, we obtain from [41]

$$1 = \langle \varphi_{k_l}, \mathcal{D}(\rho_0^{\otimes l})\varphi_{k_l}\rangle \leq \frac{1}{k_l}. \tag{770}$$

Thus $k_l = 1$ and $d(l, \rho) = 1$ for all $l \in \mathbb{N}$. We collect these observations in the following corollary.

Corollary 131. *Let $\rho \in \mathrm{ri}\,\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then $d(l, \rho) = 1$ for all $l \in \mathbb{N}$ and $D_0(\rho) = 0$. Moreover, the function D_0 is discontinuous at any $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with $D_0(\rho) > 0$.*

4.9 Entanglement generation

Let $(\varpi_l)_{l\in\mathbb{N}}$ be a sequence of (l, k_l)-entanglement generation codes for the AVQC $\mathfrak{I} = \{\mathcal{N}_s\}_{s\in\mathbf{S}}$ and $(\varepsilon_l)_{l\in\mathbb{N}}$ a sequence with $\varepsilon_l \searrow 0$ such that

$$\int F(|\psi_l\rangle\langle\psi_l|, \mathcal{R}^l \circ \mathcal{N}_{s^l}(\phi_l))d\varpi(\mathcal{R}^l, \phi_l) \geq 1 - \varepsilon_l \qquad \forall\ l \in \mathbb{N},\ s^l \in \mathbf{S}^l. \tag{771}$$

Clearly then, as in section 4.3

$$\int F(|\psi_l\rangle\langle\psi_l|, \mathcal{R}^l \circ \mathcal{N}_q^{\otimes l}(\phi_l))d\varpi(\mathcal{R}^l, \phi_l) \geq 1 - \varepsilon_l \qquad \forall\ \mathcal{N}_q \in \mathrm{conv}(\mathfrak{I}),\ l \in \mathbb{N}. \tag{772}$$

Thus, we now have a common-randomness-assisted entanglement generation code for the compound channel $\mathrm{conv}(\mathfrak{I})$. Taking a look at the proof of Theorem 91 will convince the reader that, with the obvious definition of the common-randomness-assisted entanglement generating capacity $E_{\mathrm{compound,random}}$ of a compound quantum channel, the following theorem holds:

Theorem 132 (Entanglement generation converse for an AVQC). *Let $\mathfrak{I}$ be an AVQC. Then*

$$G_{\mathrm{random}}(\mathfrak{I}) \leq E_{\mathrm{compound,random}}(\mathfrak{I}) \leq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}^{\otimes l})} \inf_{\mathcal{N}\in\mathrm{conv}(\mathfrak{I})} I_c(\rho, \mathcal{N}^{\otimes l}). \tag{773}$$

Also, the following theorem holds:

Theorem 133 (Entanglement generation direct part for an AVQC). *Let $\mathfrak{I}$ be an AVQC. Then*

$$G_{\mathrm{random}}(\mathfrak{I}) \geq \mathcal{A}_{\mathrm{random}}(\mathfrak{I}). \tag{774}$$

Proof. Let $\varepsilon > 0$ and $(\mu_l)_{l\in\mathbb{N}}$ be a sequence of (l, k_l) entanglement transmission codes for $\mathfrak{I}$ with $\liminf_{l\to\infty} \frac{1}{l}\log k_l = \mathcal{A}_{\mathrm{random}}(\mathfrak{I}) - \varepsilon$ and $\lim_{l\to\infty} \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu(\mathcal{R}^l, \mathcal{P}^l) = 1$.
Then, defining the measure ϖ_l by

$$\varpi_l(A) := \mu_l(\{(\mathcal{R}^l, \mathcal{P}^l) : (\mathcal{R}^l, id_{\mathcal{F}_l} \otimes \mathcal{P}^l(|\psi_l\rangle\langle\psi_l|)) \in A\}) \tag{775}$$

we get for every $l \in \mathbb{N}$

$$\int F(|\psi_l\rangle\langle\psi_l|, id_{\mathcal{F}_l} \otimes \mathcal{R}^l \circ \mathcal{N}_{s^l}(\phi_l)) d\varpi_l(\mathcal{R}^l, \phi_l) = \int F_e(\pi_{\mathcal{F}_l}, \mathcal{R}^l \circ \mathcal{N}_{s^l} \circ \mathcal{P}^l) d\mu(\mathcal{R}^l, \mathcal{P}^l) \tag{776}$$

and it follows that, for every $\varepsilon > 0$, $G_{\mathrm{random}}(\mathfrak{I}) \geq \mathcal{A}_{\mathrm{random}}(\mathfrak{I}) - \varepsilon$, implying Theorem 133. □

Especially, the above proof is entirely independent from the structure of the sequence $(\mu_l)_{l\in\mathbb{N}}$, it might well consist of pointmeasures. Thus,

$$\mathcal{A}_{\mathrm{random}}(\mathfrak{I}) = G_{\mathrm{random}}(\mathfrak{I}) \geq G_{\mathrm{det}}(\mathfrak{I}) \geq \mathcal{A}_{\mathrm{det}}(\mathfrak{I}). \tag{777}$$

On the other hand, if $C_{\mathrm{det}}(\mathfrak{I}) > 0$, then $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = \mathcal{A}_{\mathrm{random}}(\mathfrak{I})$ and it follows $G_{\mathrm{det}}(\mathfrak{I}) = G_{\mathrm{random}}(\mathfrak{I})$.
This finally proves Theorem 88.

5 Conclusions and open problems

5.1 Conclusion for the compound quantum channel

The results presented are analogous to those well known related results from the classical information theory obtained by Wolfowitz [72], [73], and Blackwell, Breiman and Thomasian [17]. In contrast to the classical results on compound channels there is, in general, no single-letter description of the quantum capacities for entanglement transmission and generation over compound quantum channels. Notice, however, that for compound channels with classical input and quantum output (cq-channels) a single-letter characterization of the capacity is always possible according to the results of [14].
A little surprise is contained in section 3.10, where it becomes clear that a symmetrizability condition analoguous to those given in section 4.6 for arbitrarily varying quantum channels leads to a necessary and sufficient condition for the capacity for transmission of classical messages using the average error criterion over a compound quantum channel to be equal to zero.
This is in contrast to the results for compound channels in classical informaiton theory and will be further disucssed in section 5.3.

5.2 Conclusion for the arbitrarily varying quantum channel

We have been able to derive a multi-letter analog of Ahlswede's dichotomy for quantum capacities of arbitrarily varying quantum channels: Either the classical, deterministic capacity of such a channel with average error criterion is zero, or else its deterministic and common-randomness-assisted entanglement transmission capacities are equal. Moreover, we have shown that the entanglement and strong subspace transmission capacities for this channel model are equal. It should be noted, however, that our proof of this does not rely on a strategy of "hiding" randomness in the encoding operation. In fact, by using a probabilistic variant of Dvoretzky's theorem we achieve this equality of capacities just by restricting to an appropriate code subspace of comparable dimension on the exponential scale.
Simple conditions that guarantee single-letter capacity formulas have been provided. They are generalizations of those for memoryless and stationary quantum channels.
The results for entanglement transmission have been extended to the task of entanglement generation.

5.3 Open problems

We will now enlist the open problems together with some explanatory remarks.

A) Single letter capacity formulae for compound quantum channels Natural candidates of compound quantum channels that might admit a single-letter capacity formula are given by sets of quantum channels consisting entirely of degradable channels. While it is quite easy to see from the results in [22] that the degradable compound quantum channels with *informed encoder* have a single-letter capacity formula for entanglement transmission and generation, the corresponding statement in the uninformed case seems to be less obvious.

B) Classical message transmission over compound quantum channels Apart from the work [14], there is little we know about the classical message transmission capacity of compound quantum channels. An investigation of the model with arbitrary states as inputs would be of great interest, especially in the light of symmetrizability conditions. This leads us to the next point:

C) Symmetrizability for compound quantum channels It is clear from Theorem 70 and the following remark that Definition 69 does not give a tool to answer the question exactly when a given compound quantum channel has a capacity for transmission of entanglement that is strictly greater than

zero.
So, although we could hope that the answer to such a simple question would be given by a rather simple formula, apart from the notoriously intransparent capacity formulae stated in Theorem 27, there is not much that can be said.
But, as it turned out in section 3.10 by exemplary application of one of them, the symmetrizability conditions from section 4.6 *can* be used to gain at least *some* information about compound quantum channels.
We did not at all perform a full investigation of the topic in section 3.10. It is clear, that the obvious variant of Definition 69 (see Theorem 107 for more information) can be used to cope with classical message transmission and maximal error.
It also seems highly likely, that one can write down variants of Definition 108 and Theorem 109 that directly give a sufficient condition for $Q(\mathfrak{I})$ to be equal to zero. A look at equation (670) should warn the reader that such a generalization could possibly be a non-single letter condition.
At this point, it is also worth taking a look at the classical case. Here, the situation is the following. For a set $\mathrm{W} := \{\mathrm{W}_i\}_{i=1}^N$ of channels with corresponding conditional distributions $\{w(\cdot|a|i)\}_{i=1,\ a\in\mathbf{A}}^N \subset \mathfrak{P}(\mathbf{B})$ for finite alphabets $\mathbf{A}$ and $\mathbf{B}$, it holds

C1 The compound channel W has zero capacity for message transmission (with average or maximal error probability criterion) if and only if there is an $i \in \{1, \ldots, N\}$ such that $w(\cdot|a|i) = w(\cdot|a'|i)\ \forall a, a' \in \mathbf{A}$ (see [73]).

C2 It immediately follows that if the compound channel W has zero deterministic capacity for message transmission, then the same holds for the AVC W and, again, this is regardless of which one of the above two success criteria one uses.

C3 For $\mathbf{A} = \mathbf{B} = \{0, 1\}$ the set $\mathrm{V} := \{\mathrm{V}_1, V_2\}$ with $v(0|0, 1) = v(1|1, 1) = 1$ and $v(1|0, 2) = v(0|1, 2) = 1$ has the following properties: The compound channel V has a capacity of one, while the AVC W has a capacity of zero ($\frac{1}{2}(V_1 + V_2)$ is the completely useless channel, but one channel use is sufficient to distinguish V_1 from V_2!).

We conclude that for classical compound channels, (single-letter) symmetrizability is a useless criterion, while in the quantum case, symmetrizability turns out to be a useful criterion for compound quantum channels as well.
This is, at least at first sight, due to the non-single letter character of our symmetrizability criteria for quantum channels.
Under these circumstances, it would be very interesting to find a handy criterion telling us exactly when $Q(\mathfrak{I}) = 0$ occurs. It is possible that this question can only be answered after one finds such a criterion for stationary memoryless quantum channels (compound quantum channels with $|\mathfrak{I}| = 1$).
Therefore, we are led to the following two open problems:

D) Find a handy criterion telling us exactly when the capacity for transmission of entanglement for a stationary memoryless channel vanishes And, since it might be easier to attack a special, but nontrivial case first:

E) Solve problem D) for the case of the depolarizing qubit channel

F) Isometric encodings for the AVQC We have left open the question whether the entanglement transmission capacity of arbitrarily varying quantum channels can be achieved with isometric encoding operations.

G) Common randomness and the entanglement transmission capacity of an AVQC The major unresolved problem that we are left with seems to be the question whether there are AVQCs $\mathfrak{I}$ for which $C_{\mathrm{det}}(\mathfrak{I}) = 0$ and $\mathcal{A}_{\mathrm{random}}(\mathfrak{I}) > 0$ can occur.
Or to put the question into different words: Does common randomness really help to transmit entanglement through arbitrarily varying quantum channels?

H) Entanglement generation over AVQCs Another open question is, whether $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) = G_{\mathrm{det}}(\mathfrak{I})$ holds for arbitrary AVQCs $\mathfrak{I}$. If this was the case, then the quantum Ahlswede dichotomy would hold with $\mathcal{A}$ replaced by G in every statement. However, it is unclear whether the case $\mathcal{A}_{\mathrm{det}}(\mathfrak{I}) < G_{\mathrm{det}}(\mathfrak{I})$ can occur - a question that will have to be addressed in future work as well.

I) Capacity formula for the "complete" AVQC Going back to the model that we started with in the introduction, we can formulate an even more ambitious goal: prove a capacity formula for the transmission of entanglement over a channel defined by the sets given in equation (17) with criterion of success given in equation (16) (Equivalence of strong subspace- and entanglement transmission within this model should hold by arguments analogous to those used in the proof of Theorem 80).
Taking a look at the estimates on the output states and numbers of Kraus operators of our channels in subsections 3.3.1 and 3.3.2 shows that solving this question could become difficult, since we make explicit use of the product structure of our channels.

6 Appendix

Let $\mathcal{E}$ and $\mathcal{G}$ be subspaces of $\mathcal{H}$ with $\mathcal{E} \subset \mathcal{G} \subset \mathcal{H}$ where $k := \dim \mathcal{E}$, $d_{\mathcal{G}} := \dim \mathcal{G}$. p and $p_{\mathcal{G}}$ will denote the orthogonal projections onto $\mathcal{E}$ and $\mathcal{G}$. For a Haar distributed random variable U with values in $\mathfrak{U}(\mathcal{G})$ and $x, y \in \mathcal{B}(\mathcal{H})$ we define a random sesquilinear form

$$b_{UpU^*}(x,y) := \mathrm{tr}(UpU^*x^*UpU^*y) - \frac{1}{k}\mathrm{tr}(UpU^*x^*)\mathrm{tr}(UpU^*y). \tag{778}$$

In this appendix we will give an elementary derivation of the formula

$$\mathbb{E}\{b_{UpU^*}(x,y)\} = \frac{k^2-1}{d_{\mathcal{G}}^2-1}\mathrm{tr}(p_{\mathcal{G}}x^*p_{\mathcal{G}}y) + \frac{1-k^2}{d_{\mathcal{G}}(d_{\mathcal{G}}^2-1)}\mathrm{tr}(p_{\mathcal{G}}x^*)\mathrm{tr}(p_{\mathcal{G}}y) \tag{779}$$

for all $x, y \in \mathcal{B}(\mathcal{H})$ and where the expectation is taken with respect to the random variable U.
Let us set

$$p_U := UpU^*. \tag{780}$$

Since $\mathrm{tr}(p_Ux^*p_Uy)$ and $\mathrm{tr}(p_Ux^*)\mathrm{tr}(p_Uy)$ depend sesquilinearly on $(x,y) \in \mathcal{B}(\mathcal{H}) \times \mathcal{B}(\mathcal{H})$ it suffices to consider operators of the form

$$x = |f_1\rangle\langle g_1| \quad \text{and} \quad y = |f_2\rangle\langle g_2| \tag{781}$$

with suitable $f_1, f_2, g_1, g_2 \in \mathcal{H}$. With x, y as in (781) we obtain

$$\mathrm{tr}(p_Ux^*p_Uy) = \langle f_1, p_Uf_2\rangle\langle g_2, p_Ug_1\rangle \tag{782}$$
$$= \langle f_1 \otimes g_2, (U\otimes U)(p\otimes p)(U^*\otimes U^*)f_2\otimes g_1\rangle, \tag{783}$$

and

$$\mathrm{tr}(p_Ux^*)\mathrm{tr}(p_Uy) = \mathrm{tr}((p_U\otimes p_U)(|g_1\rangle\langle f_1|\otimes|f_2\rangle\langle g_2|)) \tag{784}$$
$$= \langle f_1\otimes g_2, (U\otimes U)(p\otimes p)(U^*\otimes U^*)g_1\otimes f_2\rangle. \tag{785}$$

Since the range of the random projection $(U\otimes U)(p\otimes p)(U^*\otimes U^*)$ is contained in $\mathcal{G}\otimes\mathcal{G}$ we see from (782) and (784) that we may (and will) w.l.o.g. assume that $f_1, f_2, g_1, g_2 \in \mathcal{G}$. Moreover, (782) and (784) show, due to the linearity of expectation, that the whole task of computing the average in (779) is boiled down to the determination of

$$A(p) := \mathbb{E}((U\otimes U)(p\otimes p)(U^*\otimes U^*)) \tag{786}$$
$$= \int_{\mathfrak{U}(\mathcal{G})}(u\otimes u)(p\otimes p)(u^*\otimes u^*)du. \tag{787}$$

Obviously, $A(p)$ is $u\otimes u$-invariant, i.e. $A(p)(u\otimes u) = (u\otimes u)A(p)$ for all $u \in \mathfrak{U}(\mathcal{G})$. It is fairly standard (and proven by elementary means in [69]) that then

$$A(p) = \alpha\Pi_s + \beta\Pi_a, \tag{788}$$

where Π_s and Π_a denote the projections onto the symmetric and antisymmetric subspaces of $\mathcal{G}\otimes\mathcal{G}$. More specifically

$$\Pi_s := \frac{1}{2}(\mathrm{id}+\mathbb{F}) \qquad \Pi_a = \frac{1}{2}(\mathrm{id}-\mathbb{F}), \tag{789}$$

with $\mathrm{id}(f\otimes g) = f\otimes g$ and $\mathbb{F}(f\otimes g) = g\otimes f$, for all $f, g \in \mathcal{G}$.
Since Π_s and Π_a are obviously $u\otimes u$-invariant, and $\Pi_s\Pi_a = \Pi_a\Pi_s = 0$ holds, the coefficients α and β in (788) are given by

$$\alpha = \frac{1}{\mathrm{tr}(\Pi_s)}\mathrm{tr}((p\otimes p)\Pi_s) = \frac{2}{d_{\mathcal{G}}(d_{\mathcal{G}}+1)}\mathrm{tr}((p\otimes p)\Pi_s), \tag{790}$$

and

$$\beta = \frac{1}{\mathrm{tr}(\Pi_a)}\mathrm{tr}((p\otimes p)\Pi_a) = \frac{2}{d_{\mathcal{G}}(d_{\mathcal{G}}-1)}\mathrm{tr}((p\otimes p)\Pi_a), \tag{791}$$

where $d_{\mathcal{G}} = \dim \mathcal{G}$ and we have used the facts that

$$\mathrm{tr}(\Pi_s) = \dim \mathrm{ran}(\Pi_s) = \frac{d_{\mathcal{G}}(d_{\mathcal{G}}+1)}{2} \tag{792}$$

and

$$\mathrm{tr}(\Pi_a) = \dim \mathrm{ran}(\Pi_a) = \frac{d_{\mathcal{G}}(d_{\mathcal{G}}-1)}{2}. \tag{793}$$

It is easily seen by an explicit computation with a suitable basis that

$$\mathrm{tr}((p\otimes p)\Pi_s) = \frac{1}{2}(k^2+k) \quad \text{and} \quad \mathrm{tr}((p\otimes p)\Pi_a) = \frac{1}{2}(k^2-k). \tag{794}$$

For example choosing any orthonormal basis $\{e_1,\ldots,e_{d_{\mathcal{G}}}\}$ of $\mathcal{G}$ with $e_1,\ldots,e_k \in \mathrm{ran}(p)$ we obtain

$$\begin{aligned}
\mathrm{tr}((p\otimes p)\Pi_s) &= \sum_{i,j=1}^{d_{\mathcal{G}}} \langle e_i\otimes e_j, (p\otimes p)\Pi_s e_i\otimes e_j\rangle && (795)\\
&= \sum_{i,j=1}^{k} \langle e_i\otimes e_j, (p\otimes p)\Pi_s e_i\otimes e_j\rangle && (796)\\
&= \frac{1}{2}\Big(\sum_{i,j=1}^{k} \langle e_i,e_i\rangle\langle e_j,e_j\rangle + \langle e_i,e_j\rangle\langle e_j,e_i\rangle\Big) && (797)\\
&= \frac{1}{2}(k^2+k), && (798)
\end{aligned}$$

with a similar calculation for $\mathrm{tr}((p\otimes p)\Pi_a)$. Utilizing (790), (791), (794), and (788) we end up with

$$A(p) = \frac{k^2+k}{d_{\mathcal{G}}(d_{\mathcal{G}}+1)}\Pi_s + \frac{k^2-k}{d_{\mathcal{G}}(d_{\mathcal{G}}-1)}\Pi_a. \tag{799}$$

Now, (799), (786), (784), (782), and some simple algebra show that

$$\begin{aligned}
\mathbb{E}\{\mathrm{tr}(UpU^*x^*UpU^*y) - \frac{1}{k}\mathrm{tr}(UpU^*x^*)\mathrm{tr}(UpU^*y)\} &= \frac{k^2-1}{d_{\mathcal{G}}^2-1}\mathrm{tr}(x^*y) && (800)\\
&\quad + \frac{1-k^2}{d_{\mathcal{G}}(d_{\mathcal{G}}^2-1)}\mathrm{tr}(x^*)\mathrm{tr}(y). && (801)
\end{aligned}$$

7 References

[1] R. Ahlswede, "A Note on the Existence of the Weak Capacity for Channels with Arbitrarily Varying Channel Probability Functions and Its Relation to Shannon's Zero Error Capacity" *The Annals of Mathematical Statistics*, Vol. 41, No. 3. (1970)

[2] R. Ahlswede, "Elimination of Correlation in Random Codes for Arbitrarily Varying Channels", *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 44, 159-175 (1978)

[3] R. Ahlswede, "Coloring Hypergraphs: A New Approach to Multi-user Source Coding-II", *Journal of Combinatorics, Information & System Sciences* Vol. 5, No. 3, 220-268 (1980)

[4] R. Ahlswede, "Arbitrarily Varying Channels with States Sequence Known to the Sender", *IEEE Trans. Inf. Th.* Vol. 32, 621-629, (1986)

[5] R. Ahlswede, I. Bjelakovic, H. Boche, J. Nötzel "Quantum capacity under adversarial noise: arbitrarily varying quantum channels", accepted for publication in *Comm. Math. Phys.*, arXiv 1010.0418

[6] R. Ahlswede, V. Blinovsky, "Classical Capacity of Classical-Quantum Arbitrarily Varying Channels", *IEEE Trans. Inf. Th.* Vol. 53, No. 2, 526-533 (2007)

[7] R. Ahlswede, J. Wolfowitz, "The Capacity of a Channel with Arbitrarily Varying Channel Probability Functions and Binary Output Alphabet" *Z. Wahrscheinlichkeitstheorie verw. Geb.* 15, 186-194 (1970)

[8] R. Alicki, M. Fannes, "Continuity of quantum conditional information", *J. Phys. A* 37, L55, (2004)

[9] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. "An Introduction to Random Matrices", Cambridge University Press, (2009) Available at: http://www.wisdom.weizmann.ac.il/zeitouni/cupbook.pdf.

[10] H. Barnum, E. Knill and M.A. Nielsen, "On Quantum Fidelities and Channel Capacities", *IEEE Trans. Inf. Theory*, VOL. 46, NO. 4, (2000)

[11] H. Barnum, M.A. Nielsen, B. Schumacher, "Information transmission through a noisy quantum channel", *Phys. Rev. A* Vol. 57, No. 6, 4153 (1998)

[12] C.H. Bennett, P.W. Shor, J.A. Smolin, and A.V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem", *IEEE Trans. Inf. Th.* 48, 2637-2655 (2002)

[13] C.H. Bennett, D.P. DiVincenzo, and J.A. Smolin, "Capacities of Quantum Erasure Channels", *Phys. Rev. Lett.* 78, 32173220 (1997)

[14] I. Bjelaković, H. Boche, "Classical Capacities of Averaged and Compound Quantum Channels", *IEEE Trans. Inf. Th.* Vol. 55, 7, 3360 - 3374 (2009)

[15] I. Bjelaković, H. Boche, J. Nötzel, "Quantum capacity of a class of compound channels", *Phys. Rev. A* 78, 042331, (2008)

[16] I. Bjelaković, H. Boche, J. Nötzel, "Entanglement transmission and generation under channel uncertainty: Universal quantum channel coding", *Commun. Math. Phys.* 292, 55-97 (2009) - Available at: http://arxiv.org/abs/0811.4588

[17] D. Blackwell, L. Breiman, A.J. Thomasian, "The capacity of a class of channels", *Ann. Math. Stat.* Vol. 30, No. 4, 1229-1241 (1959)

[18] D. Blackwell, L. Breiman, A.J. Thomasian, "The capacities of certain channel classes under random coding", *Ann. Math. Stat.* 31, 558-567 (1960)

[19] M.-D. Choi, "Completely Positive Linear Maps on Complex Matrices", *Linear Algebra and Its Applications* 10, 285-290 (1975)

[20] I. Csiszar, J. Körner, *Information Theory; Coding Theorems for Discrete Memoryless Systems*, Akadémiai Kiadó, Budapest/Academic Press Inc., New York 1981

[21] I. Csiszar, P. Narayan, "The Capacity of the Arbitrarily Varying Channel Revisited: Positivity, Constraints", *IEEE Trans. Inf. Th.* Vol. 34, No. 2, 181-193 (1989)

[22] T. Cubitt, M. Ruskai and G. Smith, "The structure of degradable quantum channels", *Jour. Math. Physics* Vol. 49, No. 10, 102104 (2008)

[23] N. Datta, T.C. Dorlas, "The coding theorem for a class of quantum channels with long-term memory" *J. Phys. A: Math. Theor.* 40, 8147-8164 (2007)

[24] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel", *IEEE Trans. Inf. Th.* 51, No.1, 44-55 (2005)

[25] I. Devetak and P.W. Shor, "The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information", *Commun. Math. Phys.* Vol. 256, Nr. 2 (2005)

[26] I. Devetak, A. Winter, "Distillation of secret key and entanglement from quantum states", *Proc. R. Soc. A* 461, 207-235 (2005)

[27] R. Duan, S. Severini, A. Winter, "Zero-error communication via quantum channels, non-commutative graphs and a quantum Lovász θ function", arXiv:1002.2514v2

[28] F. Dupuis, "The decoupling approach to quantum information theory", arXiv:1004.1641 or https://papyrus.bib.umontreal.ca/jspui/handle/1866/3363 (2010)

[29] T. Ericson, "Exponential Error Bounds for Random Codes in the Arbitrarily Varying Channel", *IEEE Trans. Inf. Th.* Vol. 31, No. 1, 42-48 (1985)

[30] M. Fannes, "A continuity property of the entropy density for spin lattice systems", *Commun. Math. Phys.* 31, No. 4, 291-294 (1973)

[31] M. Fekete, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten", *Mathematische Zeitschrift* 17, 228 (1923).

[32] C. A. Fuchs and J. van de Graaf. "Cryptographic distinguishability measures for quantum mechanical states", *IEEE Trans. Inf. Theory* 45, 1216 (2007)

[33] E.N. Gilbert, "A comparison of signaling alphabets", *Bell System Tech. J.* 31, 504-522. (1952)

[34] M. Hayashi, "Universal coding for classical-quantum channel", *Commun. Math. Phys.* 289, No. 3, 1087-1098 (2009)

[35] P. Hayden, M. Horodecki, A. Winter, J. Yard, "A decoupling approach to the quantum capacity", *Open. Syst. Inf. Dyn.* 15, 7-19 (2008)

[36] P. Hayden, P.W. Shor, A. Winter, "Random Quantum Codes from Gaussian Ensembles and an Uncertainty Relation", *Open. Syst. Inf. Dyn.* 15, 71-89, (2008)

[37] A.S. Holevo: "Bounds for the quantity of information transmitted by a quantum channel", *Probl. Inform. Transm.* Vol. 9, No. 3, 177-183 (1973)

[38] A.S. Holevo, "The Capacity of the Quantum Channel with General Signal States", *IEEE Trans. Inf. Th.* Vol. 44, No. 1, 269-273, (1998)

[39] A.S. Holevo, "On entanglement-assisted classical capacity" *Jour. Math. Physics* Vol. 43, No. 9, 4326-4333 (2002)

[40] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press (1999)

[41] M. Horodecki, P. Horodecki, "Reduction criterion of separability and limits for a class of distillation protocols", *Phys. Rev. A* Vol. 59, No. 6, 4206 (1999)

[42] M. Horodecki, P. Horodecki, R. Horodecki, "General teleportation channel, singlet fraction, and quasidistillation ", *Phys. Rev. A* 60, 18881898 (1999)

[43] R. Jozsa, M. Horodecki, P. Horodecki, R. Horodecki, "Universal Quantum Information Compression", *Phys. Rev. Letters* Vol. 81, No. 8, 1714-1717 (1998)

[44] S. Kakutani, "A Generalization of Brouwer's Fixed Point Theorem", *Duke Math. J.*, Volume 8, Number 3, 457-459 (1941)

[45] J. Kiefer, J. Wolfowitz, "Channels with arbitrarily varying channel probability functions", *Information and Control* 5, 44-54 (1962)

[46] A.Y. Kitaev, A.H. Shen, M.N. Vyalyi, *Classical and Quantum Computation*, Graduate Studies in Mathematics 47, American Mathematical Society, Providence, Rhode Island 2002

[47] R. Klesse, "Approximate Quantum Error Correction, Random Codes, and Quantum Channel Capacity", *Phys. Rev. A* 75, 062315 (2007)

[48] E. Knill, R. Laflamme, "Theory of quantum error-correcting codes", *Phys. Rev. A* Vol. 55, No. 2, 900-911 (1997)

[49] J. Körner, A. Orlitsky, "Zero-error Information Theory", *IEEE Trans. Inf. Theory* Vol. 44, No. 6, 2207-2229 (1998)

[50] D. Kretschmann, R.F. Werner, "Tema con variazioni: quantum channel capacity", *New Journal of Physics* Vol. 6, 26-59 (2004)

[51] D. Leung, G. Smith, "Continuity of quantum channel capacities", *Commun. Math. Phys.* 292, 201-215, (2009)

[52] E.H. Lieb and M.B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy", *J. Math. Phys.* 14, 1938 (1973)

[53] S. Lloyd, "Capacity of the noisy quantum channel ", *Phys. Rev. A* Vol. 55, No. 3 1613-1622 (1997)

[54] J. Matousek, *Lectures on Discrete Geometry*, Graduate Texts in Mathematics, Vol. 212, Springer 2002

[55] V.D. Milman, G. Schechtman *Asymptotic Theory of Finite Dimensional Normed Spaces*, Lecture Notes in Mathematics vol. 1200, Springer-Verlag 1986

[56] M.A. Nielsen, I. Chuang, *Quantum Information and Computation*, Cambridge University Press, Cambridge, UK, 2000

[57] T. Ogawa, H. Nagaoka, "Strong converse to the quantum channel coding theorem", *IEEE Trans. Inf. Th.* Vol. 45, No. 7, 2486-2489 (1999)

[58] V. Paulsen, *Completely Bounded Maps and Operator Algebras*, Cambridge Studies in Advanced Mathematics vol. 78, Cambridge University Press 2002

[59] B. Schumacher, "Sending entanglement through noisy quantum channels", *Phys. Rev. A* Vol. 54, No. 4, 2614 (1996)

[60] B. Schumacher, M.A. Nielsen, "Quantum data processing and error correction." *Phys. Rev. A* Vol. 54, No. 4, 2629 (1996)

[61] B. Schumacher, M.D. Westmoreland, "Sending classical information via noisy quantum channels", *Phys. Rev. A* Vol. 56, No. 1, 131-138, (1997)

[62] B. Schumacher, M.D. Westmoreland, "Approximate quantum error correction", *Quant. Inf. Proc.* Vol. 1, 5-12 (2002)

[63] C. E. Shannon, "The zero error capacity of a noisy channel". *IRE Trans. Information Theory IT-2*, 8-19 (1956)

[64] P.C. Shields, *The Ergodic Theory of Discrete Sample Paths*, Graduate Studies in Mathematics Vol. 13 (American Mathematical Society 1996)

[65] P. Shor, unpublished talk manuscript. Available at:
http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/

[66] A. Uhlmann "The "Transition Probability" in the State Space of a *-Algebra", *Rep. Math. Phys.* 9 273 - 279 (1976)

[67] J. von Neumann, "Zur Theorie der Gesellschaftsspiele", *Math. Ann.* Vol. 100, 295-320 (1928)

[68] R. Webster, *Convexity*, Oxford University Press 1994

[69] R.F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model", *Phys. Rev. A* Vol. 40, No. 8, 4277-4281 (1989)

[70] A. Winter, "Coding theorem and strong converse for quantum channels", *IEEE Trans. Inf. Th.* Vol. 45, No. 7, 2481-2485 (1999)

[71] M. M. Wolf, J.I. Cirac, "Dividing Quantum Channels" *Commun. Math. Phys.* 279, 147-168 (2008)

[72] J. Wolfowitz, "Simultaneous channels", *Arch. Rational Mech. Anal.* Vol. 4, No. 4, 371-386 (1960)

[73] J. Wolfowitz, *Coding Theorems of Information Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete 31, 3. Edition, Springer-Verlag, Berlin, Germany, 1978

[74] J. Yard, I. Devetak, P. Hayden, "Capacity theorems for quantum multiple access channels: Classical-quantum and quantum-quantum capacity Regions", *IEEE Trans. Inf. Theory* 54, 3091 (2008) e-print arXiv:quant-ph/0501045.