Technische Universität München
TUM School of Computation, Information and Technology

Universal Codes for Broadcast Systems with Physical Layer Security and Their Algorithmic Computability in Quantum Information Theory

Sajad Saeedinaeeni

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zu Erlangung des akademischen Grades eines:

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr.-Ing. Wolfgang Kellerer

Prüfer der Dissertation:

1. Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
2. Prof. Dr. Robert König

Die Dissertation wurde am 26.04.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 06.12.2022 angenommen.

Abstract

When implementing communication protocols in the real world, the idealized conditions assumed at the theoretical level often do not hold. At the practical level, one is interested in controlling aspects of the communication that might be considered ”negligible” in the mathematical modeling of the given protocol. This effect often renders a large class of protocols with such assumptions unreliable in practice. In this thesis, we attempt to address practical questions that undermine known communication protocols, by either re-modeling the protocols in such a way that they would be robust against practical uncertainties, or demonstrating that they face logical or feasibility problems prior to their implementation. There are two categories of such questions that are addressed here. The first one concerns the uncertainty in the communication system parameters, in particular the communicating channel. This affects security as well as reliability measures. Reliable transmission of messages in protocols that assume the state of the channel to be perfectly known relies fundamentally on this assumption. On the other hand, secure transmission of messages, depends on the powers ascribed to the eavesdropping and jamming parties. In the realm of quantum information theory, these powers are significantly higher. We start by giving results that have destructive implications for some established protocols in quantum key distribution in presence of a *quantum jammer*. We then offer a remodeling of communication protocols, that makes them robust to attacks at physical layer, as well as robust to system parameter uncertainties. Of particular interest are integrated services that are possible in quantum communication. Time-sharing between known coding strategies to perform multiple tasks is often sub-optimal. We derive universal codes that achieve full *capacity regions* of the quantum channel for communicating quantum, public and private messages. The notion of privacy or security in this work is an information theoretic one, that is different from its cryptographic counterpart. Here, we consider strategies that are implemented at the physical layer and are therefore only limited by the physical properties of the system[1]. Given that our coding strategies are as mentioned, robust to uncertainties of the physical properties of the system, they offer a significant advantage to cryptographic methods in practice. The second category of practical questions concerns *computability* of the known protocols. Here, we specifically address the algorithmic computability of capacity functions. When capacity theorems are theoretically derived, a legitimate practical question is whether or not these functions can be fed to a machine via an algorithm. Asked differently, are optimal protocols that achieve capacities

[1] In contrast, known cryptographic methods, rely on computational infeasiblity of eavesdropping.

of channels for different communication tasks, implementable through algorithms? The value of these functions is derived from achievability and converse theorems that ideally establish converging computable sequences from below and above the capacity at any given point. A more relaxed requirement is the question of *decidability* of such problems. We end this thesis by giving examples of protocols that cannot in general be turned into algorithms that can be computed by *Turing machines*. We further show that the bounds for capacities achieved by these protocols are in general undecidable.

Acknowledgments

I am most grateful to my supervisor Prof. Dr. Holger Boche for his unwavering support and guidance in the course of this thesis. His excitement for the included topics was an indispensable source of motivation to pursue unanswered questions and to see the beauty of the unknown and uncertainty; a vision that is of necessity for doing science. I am honored to have been a part of his research group and to have gotten to know like-minded people who love solving puzzles. I would like to especially express my gratitude to Dr. Gisbert Janßen for being a great mentor and a friend. This thesis would not have been possible without his flare for accuracy and deep understanding in Quantum Information Theory. I am thankful to Dr. Ezra Tampubolon for numerous discussions and great times both inside and outside the institute.

After the original submission of this thesis on 26.04.2022, Prof. Dr. König and Prof. Dr. Kellerer accepted to be the respective second supervisor and the chair of examination committee, for which I am most thankful.

I thank my parents Elaheh Assari and Amir Saeedi Naeeni, for being patient and encouraging throughout years of intolerable distance. I thought about them and my family in Iran every day and took heart from knowing that they take pride in my accomplishments as much as I do in theirs.

I thank my friends who made Germany an incredible second home for me, accepted me with open hearts and generosity. I thank Thomas Alexander Campbell, Tilman Höing, Julius Nückl, Marco Lorenz and Malik Maghames for all the great times that I was lucky enough to experience with them.

Last but not least, I would like to thank my little sister Maryam Saeedi Naeeni for being the primary source of love and light in my life.

Contents

1. Introduction

The ever strengthening belief in the future of quantum technologies that accompanies their commercial advent, has brought about increasing interest in implementability of the proposed quantum information and computation protocols. Algorithms by Shor and Grover [Grover(1996)] to perform computations that are extremely hard and provably impossible on any classical computer could be considered as some of the first protocols that demonstrated use cases for quantum information processing. The advantages of this new model of information were also demonstrated in realization of communication tasks such as *quantum teleportation* and *dense coding* [Bennett et al.(1993)Bennett, Brassard, Crépeau, Jozsa, Peres, and Wootters]. Determining the asymptotic capacities of quantum channels has been studied in the past decades as one of the most prominent subjects in quantum Shannon theory [Wilde(2017)]. There, quite like the classical Shannon theory, information transmission ability of a sender and a receiver, connected via a stochastic channel is examined, while both may use the channel infinitely many times. These protocols in their inception are normally based on theoretical idealizations that provide prototypes for further developments in quantum software, but at implementation, face statistical uncertainties introduced by system parameters. This thesis considers reducing such idealizations in modeling of integrated services that involve simultaneous accomplishment of more than one information processing task, in favor of their real-world implementation. We present this introduction in the following order. We start by giving an introduction to the main content of the thesis, including the studied channel models, information processing tasks and indeed what brings them together to form the theme of this thesis. We do this, bearing in mind that every chapter has a more detailed introduction on its own. We proceed by introducing the statistical model used in this work. This model falls within the boundaries of *Quantum Statistics.* This is done with the intention of expanding the audience of the present work, by getting across the idea that for all intents and purposes here, a basic understanding of probability theory and linear algebra is sufficient for comprehending the results. Finally, we end this introduction by fixing the mathematical notation used in this work.

1.1. Introduction to the topics of the present work

This thesis studies optimal coding strategies for performing integrated services that are possible using quantum channels under real-world assumptions. Here, we briefly introduce these services and their integration into parent protocols that are implemented on the *physical layer* of the communication system. Given the dependence of these protocols on the physical properties of the system, we then motivate the main area of contribution of this thesis, by introducing channel models that include physical-parameter uncertainties that communicating parties inevitably face in practice.

The quantum channel should be understood as the most general way in which noisy evolution and transmission of information is modeled in information theory. Two quantum systems can have a very useful correlation with each other known as entanglement [Wilde(2017)]. This correlation is a valuable resource for public as well as private communication. Given that a quantum channel preserves all correlations of a system with other systems, it has the capacity to transmit and generate entanglement, allowing the use of this resource for communication in the first place. In addition to transmitting messages between senders and receivers of different permissions and priorities therefore, quantum channels have the capacity for tasks that go beyond those possible by classical (today's commercial) communication systems. One may consider the capacity of the channel for public ([Holevo(1998)], [Schumacher and Westmoreland(1997)]) or private ([Devetak(2005)], [Cai et al.(2004)Cai, Winter, and Yeung]) message transmission, entanglement transmission or entanglement generation ([Devetak(2005)]) to name a few.

Simultaneous (integrated) transmission of different types of messages can take place between two parties. The body of research in physical layer integrated services is only interesting where capacity regions beyond those achievable by simple time-sharing between the tasks (Figure 1.1) are achieved. As more tasks are possible using quantum systems, the importance such optimal strategies and their advantage over separate performance of individual tasks becomes clearer. Such integrated (or simultaneous) coding has been considered in the classical realm and for perfectly known channels in quantum information theory. For instance, simultaneous transmission of classical and quantum messages, the subject of Chapter 3, has been of interest([Devetak and Shor(2005)]) for the case where the parameters of the communicating channel are perfectly known to the sender and receiver. This includes scenarios where the communication parties would like to enhance their classical message transmission when having quantum information primarily at their disposal or vice versa([Bennett et al.(1999)Bennett, Shor, Smolin, and Thapliyal], [Hsieh and Wilde(2010a)], [Hsieh and Wilde(2010b)]). In communication systems, the physical layer is determined by the stochastic channel that connects the communicating parties. This approach however could also be considered in quantum computers, where the design of the processor is given by the separate implementation of different layers that are placed on top of the physical layer ([Jones et al.(2012)Jones, Van Meter, Fowler, McMahon, Kim, Ladd, and Yamamoto]). The physical layer consists of hardware apparatus including qubits and control operations. The data storage will then be subject to error correction and logical programming to perform the desired algorithms at the interface on the higher levels. Integrating error correction at the physical layer would require simulta-

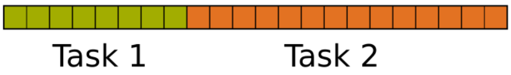

Figure 1.1.: Time sharing between two tasks

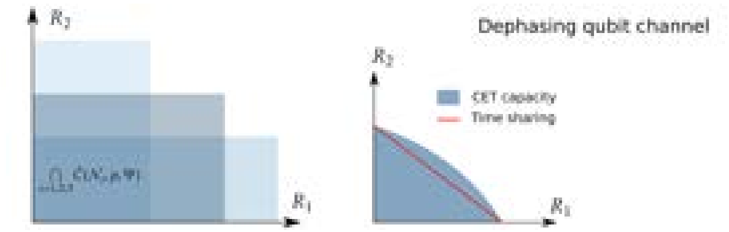

Figure 1.2.: Left: Capacity regions for classically enhanced entanglement transmission. Right: going beyond time-sharing with dephasing qubit channel [Wilde et al.(2012)Wilde, Hayden, and Guha]

neous implementation of these services that possibly improve time-sharing (Figure 1.2).

In today's communication systems, issues such as authentication and privacy of message identification and transmission protocols are handled in system's upper layers using variations of private or public key cryptographic methods (RSA, AES). These methods rely on computational limitations of illegal parties and hence, are becoming increasingly unreliable [Schaefer and Boche(2014a)]. This concern has motivated much of the research on the alternative concept of physical layer integration and more specifically, information theoretic security [Liang et al.(2009)Liang, Poor, and Shamai], [Jorswieck et al.(2010)Jorswieck, Wolf, and Gerbracht], [Liu and Trappe(2010)], [Bloch and Barros(2011)]. Information theoretic security is modelled by the *wiretap channel* (Figure 1.1) that connects the sender to two receivers, one legal and the other wiretapper. The secrecy from the wiretapper is then achieved via a stochastic encoding procedure. The encoder first uses random codes designed for message transmission and then uses part of these codes to confuse the wiretapper. This procedure known as equivocation, makes sure that the outcome of the channel at the wiretapper's end is arbitrarily close to a fixed state, independent of the encoding. This gives a positive rate of secure messages transmitted to the legal receiver, in case the channel connecting the legal parties is better (less noisy) than the one between the sender and the wiretapper. Wyner [Wyner(1975)] introduced the classical wiretap channel, and considered a subclass of channels known

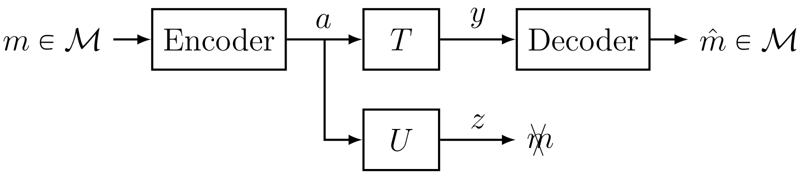

Figure 1.3.: Wiretap channel

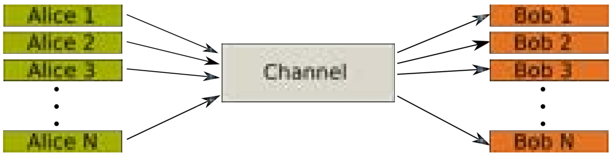

Figure 1.4.: Multi-user channel models

as the degraded wiretap channels, before Csiszár and Körner [Csiszár and Körner(1978)] addressed the general case. The model can be described by two channels from the sender (”Alice”) to the legal receiver (”Bob”) and to the eavesdropper (”Eve”), respectively. In transmission theory the goal is to send messages to the legal receiver, while the wiretapper is to be kept ignorant. The wiretap channel was generalized to the setting of quantum information theory in [Cai et al.(2004)Cai, Winter, and Yeung, Devetak(2005)].

Real-world communication usually involves more communication parties than just one sender and one receiver (Figure 1.4). A very basic situation is when two or more sending parties are connected to a receiver via a multiple-access channel (MAC). A sample use case of this model is when two senders share the same fiber transmission line to a receiver, while both independently aim to achieve individual transmission goals. Developing coding schemes for such situations is of technological importance, since presuming availability of a ”dark fibre” for performing a transmission protocol is rarely feasible. This fact already became apparent as a limiting factor in recent attempts to use commercial fibre lines for quantum key distribution ([Dynes et al.(2016)Dynes, Tam, Plews, Fröhlich, Sharpe, Lucamarini, et al.], [Jacak et al.(2016)Jacak, Melniczuk, Jacak, Janutka, Jóźwiak, Gruber, and Jóźwiak])-commercial fibre lines are usually a valuable resource being shared by many users. Consequently, the rate as well as the performance each of the sending parties can achieve is in general strongly connected to the signal characteristics of other parties. Finding code constructions that asymptotically achieve the optimal rate regions in the Shannon-theoretic sense is a highly nontrivial task ([Boche et al.(2019a)Boche, Janßen, and Saeedinaeeni]). Another important channel model that allows access to more than two parties is the *broadcast channel* (Figure 1.1), in which one sender is connected to two receivers. Here, the sender might wish to communicate one public message received by both, and another private message that is only meant for one receiver and kept secret from the other. Similar to the wiretap channel, the notion of secrecy here is that assured by upper-bounding the mutual information between the sender and the receiver from whom the message is to be kept secret. This results in information theoretic security that only depends on the physical properties of the communicating channel, in contrast with cryptographic security, that depends on the computational limitations of illegal parties [Schaefer and Boche(2014a)]. As outlined in Chapter 4, sophisticated coding strategies are needed to achieve the optimal capacity of the channel for the integrated task in which a message by a sender contains public and private information.

As the coding strategies outlined above all depend on intrinsic and physical properties of the system, a challenge facing their practical implementation is when the physical parameters of the communication system in question are unknown to the communicating parties. In fact, assuming these parameters to be known is unrealistic. In the example of the wiretap channel for instance, the channel to the eavesdropping party is rarely perfectly known. Therefore, taking a step closer to real-world implementation of the mentioned tasks, one needs to consider *channel uncertainty*. In real world communication using quantum or classical systems, the parameter determining the channel in use may belong to an uncertainty set, rendering the protocols that assume the channel to be perfectly known practically obsolete. Given such uncertainty, when using the channel many times, as done in Shannon theoretic information processing tasks, assuming the channel to be memoryless or fully stationary is not realistic. In this thesis we consider three models that include channel uncertainty without attempting to reduce it via techniques such as channel identification or tomography. We refer to these models as the compound, arbitrarily varying and fully quantum arbitrarily varying channel models.

Informally, the first two channel models consist of a set of quantum channels $\{\mathcal{N}_s\}_{s\in S}$ known to the communicating parties. In the compound model, communication is done under the assumption that asymptotically, one of the channels from this set (unknown to the parties) is used in a memoryless fashion (Figure 1.6). The codes used in this model therefore have to be reliable for the whole family $\{\mathcal{N}_s^{\otimes l}\}_{s\in S}$ of memoryless channels for large enough values of $l \in \mathbb{N}$.

In the arbitrarily varying model, given a number of channel uses l, an adversarial party (jammer) chooses the sequence $s^l = (s_1, \dots, s_l) \in S^l$, unknown to the communication parties, to yield the channel $\mathcal{N}_{s^l} := \bigotimes_{i=1}^{l} = \mathcal{N}_{s_i}$. The adversary may choose this sequence knowing the encoding procedure used by the sender. The code in use therefore has to be reliable for the whole family $\{\mathcal{N}_{s^l}\}_{s^l\in S^l}$ of memoryless channels (Figure 1.7). Finally, in the third channel model, namely that of the fully quantum arbitrarily varying, the assumption of memoryless communication is dropped. Here, the adversary may choose channel states that are not necessarily of the product form mentioned in the previous model (Figure 1.8). The size of this uncertainty set S, depends on the strategy and physical resources used for channel estimation, and under real-life physical communication conditions, will in general be infinite.

In the two arbitrarily varying channel models, we refer to the state chose by the adversary as a jamming attack. We consider attacks that are performed directly at the physical layer with the aim of disrupting the physical transmission itself. Such attacks can target a specific single user within the system, but also the overall system itself. Reliable communication between legitimate users is the indispensable basis for any information processing. In the worst case, the jammer is able to perform a denial-of-service (DoS) attack which means that no communication is possible at all. In [H. Boche(2020a)] it was shown that it is impossible to algorithmically detect such fundamental physical jamming attacks. The undetectability of DoS attacks has crucial implications and consequences

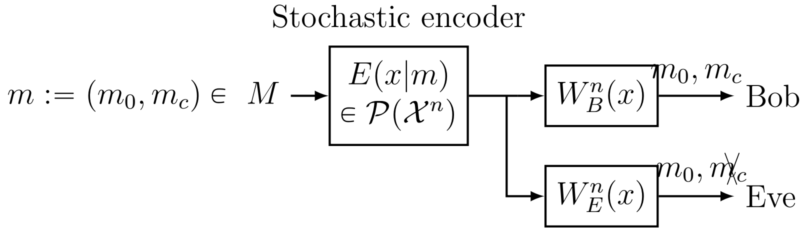

Figure 1.5.: Broadcast channel model with one public and one confidential message

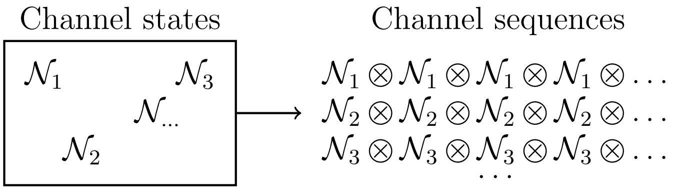
Channel states
$\mathcal{N}_1$
$\mathcal{N}_3$
$\mathcal{N}_{\dots}$
$\mathcal{N}_2$
Channel sequences
$\mathcal{N}_1 \otimes \mathcal{N}_1 \otimes \mathcal{N}_1 \otimes \mathcal{N}_1 \otimes \dots$
$\mathcal{N}_2 \otimes \mathcal{N}_2 \otimes \mathcal{N}_2 \otimes \mathcal{N}_2 \otimes \dots$
$\mathcal{N}_3 \otimes \mathcal{N}_3 \otimes \mathcal{N}_3 \otimes \mathcal{N}_3 \otimes \dots$
$\dots$

Figure 1.6.: Compound channel model.

on higher layers of communication systems. It was discussed in [H. Boche(2020a)] that it is possible to obtain resilience by design, by invoking additional resources to stabilize the communication directly at the physical layer. Techniques to achieve resilience by design have been analyzed in [Ahlswede et al.(2013)Ahlswede, Bjelakovic, Boche, and Nötzel, H. Boche(2014), H. Boche(2019), H. Boche(2020b)].
Relaxing the assumption of the perfectly known channel, requires coding strategies that work for all channels belonging to a set of possibly infinite cardinality and are hence, significantly more sophisticated. A case in point is the coding strategy established in [Boche et al.(2019b)Boche, Janßen, and Saeedinaeeni] to derive capacity results for simultaneous transmission of classical (public) messages and quantum information over the quantum channel, given that those developed for the perfectly known channel in [Devetak and Shor(2005)] did not provide the structure needed to deal with channel uncertainty. Optimal codes derived for the compound model, can be used to derive optimal codes for the arbitrarily varying models [Ahlswede(1978)]. This fact further emphasizes the theoretic importance of the compound model.

We finish this thesis by analyzing the achievability and converse bounds (comprising a coding theorem) from a fundamental, algorithmic point of view by studying whether or not such bounds can be computed algorithmically in principle (without putting any constraints on the computational complexity of such algorithms). For this purpose, the concept of *Turing machines* is used which provides the fundamental performance limits of digital computers. A Turing machine is a mathematical model of an abstract machine that manipulates symbols on a strip of tape according to certain given rules. It can simulate any given algorithm and therewith provides a simple but very powerful model of computation. Turing machines have no limitations on computational complexity, unlimited computing capacity and storage, and execute programs completely error-free. They are further equivalent to the von Neumann-architecture without hardware limitations

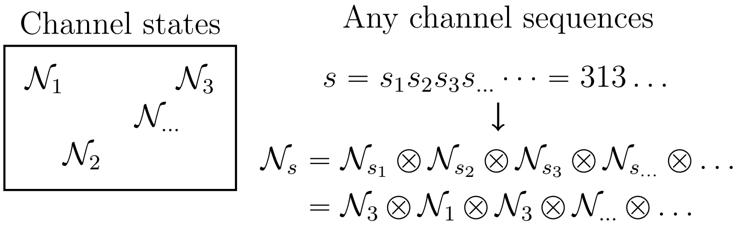
Channel states
$\mathcal{N}_1$
$\mathcal{N}_3$
$\mathcal{N}_{\dots}$
$\mathcal{N}_2$
Any channel sequences
$s = s_1 s_2 s_3 s_{\dots} \cdots = 313\dots$
$\downarrow$
$\mathcal{N}_s = \mathcal{N}_{s_1} \otimes \mathcal{N}_{s_2} \otimes \mathcal{N}_{s_3} \otimes \mathcal{N}_{s_{\dots}} \otimes \dots$
$= \mathcal{N}_3 \otimes \mathcal{N}_1 \otimes \mathcal{N}_3 \otimes \mathcal{N}_{\dots} \otimes \dots$

Figure 1.7.: Arbitrarily varying channel model.

in the theory of recursive functions, cf. also [Avigad and Brattka(2014), Gödel(1930), Gödel(1934), Kleene(1952), Minsky(1961)]. Accordingly, Turing machines provide fundamental performance limits for today's digital computers. Since bounds on the capacity are usually evaluated and often plotted on digital computers, Turing machines are the ideal concept to study whether or not such upper and lower bounds can be found algorithmically in principle. Subsequently, these findings are applied to two different open problems. The first one is the ϵ-capacity of compound channels which is unknown to date. It is shown that either the achievability or converse must yield a non-computable bound. This is demonstrated for the capacity as a function of the error input. We also consider the less restrictive condition of *decidability*. The crucial consequence is that the ϵ-capacity cannot be characterized by a finite-letter entropic expression and is not in general a decidable problem. The second application are asymptotic bounds for tasks involving pre-shared resources such as common randomness and entanglement. We demonstrate using our computability results that such resources can offer advantages in the asymptotic regime.
In Chapter 2 following this introduction, we motivate the main idea of this thesis further by giving an example where dealing with general attacks (fully quantum jammer Fig 1.8) calls for more sophisticated coding strategies than those suggested by cryptographic methods. We demonstrate this by showing that a quantum jammer's power cannot be approximated by a classical one. This is shown in the case of the well-known quantum secret key distribution protocol. We do this by demonstrating that the protocols that try to approximate the fully quantum jammer by a classical one using the known de Finetti approximation, must use more information theoretic resource in form of common randomness that they yield. This observation is an example showing the non-trivial nature of quantum generalization of the arbitrary varying channel model.
In Chapter 3, based on results from [Boche et al.(2019b)Boche, Janßen, and Saeedinaeeni], we consider an integrated task in which the communicating parties wish to transmit classical messages and entanglement under the channel uncertainty models mentioned above. Precise definitions of the protocols will be given therein. Clearly, the resulting capacity-region achieving codes here will reduce to those appropriate for each of these two tasks, when only one dimension of the region is considered. In Chapter 4, based on results from [Boche et al.(2019c)Boche, Janßen, and Saeedinaeeni], we consider an integrated task where the communicating parties wish to establish secure and public communica-

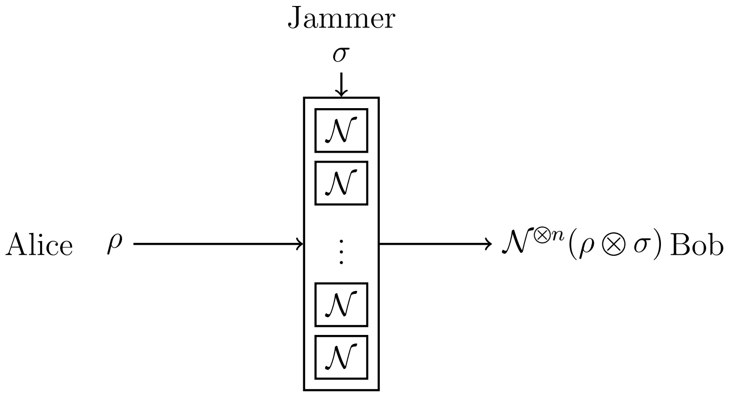

Figure 1.8.: Fully quantum arbitrarily varying channel model.

tion. Here, our channel may be appropriately named a compound classical-quantum broadcast channel. Again our codes will achieve the channel's two dimensional capacity region that also contains the capacity of the channel for each task. In this chapter, we leave out the capacity under assumptions of the arbitrarily varying model, and instead delve deeper into the compound model. Information theoretically, the compound model has yielded intriguing properties. One of the interesting information theoretic properties of the compound channel is that in general, a strong converse cannot be established on the capacity of the compound channel for message transmission when upper-bounding of the average decoding error is considered. This holds even for finite uncertainty sets [Ahlswede(1967a), Ahlswede and Wolfowitz(1969), Bjelaković et al.(2013)Bjelaković, Boche, Janßen, and Nötzel]. This observation implies that a second order capacity theorem cannot be developed in this case. Further, calculation of the so-called ϵ-capacity of the compound channel under the average error criterion is still an open question. We note however, that determining a second order ϵ-capacity for the compound channel is not possible, due to the observation that there are examples of the compound channel where the optimistic ϵ-capacity is strictly larger than its pessimistic one (see [Boche et al.(2018a)Boche, Schaefer, and Poor] Remark 13).

We consider the computational properties of the ϵ-capacity of the compound channel in Chapter 5 and based on results from [Boche et al.(2022)Boche, S. Saeedinaeeni, and Poor]. Therein, we consider examples where the ϵ-capacity of the compound channel is not Turing computable or less restrictively, decidable. Since classical channels are a specific example of classical-quantum and quantum channels, in this chapter we consider classical channels that give rise to non-computable capacities. Therein we also consider assisted scenarios where communicating parties have at their disposal, pre-shared entanglement and correlation. In the following two sections, we introduce the underlying statistical theory appropriate for the information processing tasks and the mathematical notion considered in this thesis.

1.2. Introduction to the statistical model

The prevalence of quantum information and computation in performing tasks that are not possible in the *classical* realm both at the software and hardware levels, can be attributed to the often counter-intuitive behaviour of the fundamental particles of the universe, using which quantum protocols and algorithms are performed. The intersection of this thesis with the *strange* picture of quantum mechanics, is however the very well understood statistical model that was born out of the theoretical needs, and later, axiomatic treatment of the physical theory. A practical model of statistics, provides the scientist with the necessary mathematical tools to record accounts of a given experiment. This includes a set of states, possible transformations and measurements that the system can take, undergo and be observed with, respectively. There are natural requirements on any such model ([Holevo(2012)], [Ludwig(1983)]). For instance, one requires the set of states to be statistically convex. This means that a statistical mixture of possible configurations (preparations) must be a configuration permitted by the set of states. A measurement with finitely many outcomes, is then a device or more mathematically put, an affine map that takes in a state from the state space, and outputs a probability distribution, on the set of its outcomes. Given the structure that such natural requirements impose on the set of states, here we intend to give an explanation as to why a new statistical model was called into necessity by discoveries in physics.

The need for a more general model of statistics presented itself in the 1920s, as physicists were trying to explain phenomena such as Bose-Einstein condensation and stability of atoms with even numbers of electrons (Fermions). These observations resulted in a theoretical departure from what is now referred to as classical physics. The problem with the existing statistical models was their insufficiency to account for *indistinguishable* particles. Here, we must specify that by two indistinguishable systems, we mean two that cannot be told apart by any statistical test (measurement). More specifically and in the context of statistical mechanics, two particles in the position-momentum phase-space are indistinguishable if their position and momentum cannot be observed simultaneously or via a *joint measurement*. The usefulness of treating constituting particles of a system as indistinguishable originates in a paper by S.N Bose published 1924 [Bose(1924)]. In an attempt to describe electromagnetic radiation in the framework of statistical mechanics, the author suggested that the number of distinct phase-space micro-states of an ideal gas made up of photons, was significantly lower than the one predicted by Maxwell-Boltzmann statistics. This new model that predicted the experimental results more accurately, treated micro-states that were obtained by exchange of the particles and yielded the same macro-state as indistinguishable. This notion was later reinforced by Heisenberg's uncertainty principle [Compton and Heisenberg(1984)]. In classical mechanics, identical particles, namely those that share their intrinsic properties such as mass, electric charge or size, can always be told apart given their position. This possibility is overruled by the wave-like behaviour of quantum particles. According to Heisenberg's uncertainty principle, one cannot determine the position of a particle with arbitrary precision, resulting in indistinguishability of identical particles in close enough proximity. In the following we show that in the formalism of classical statistics, every pair of observables are jointly measurable, resulting in turn in distinguishability of classical systems. We argue that this is due to the *simplex* structure of the convex set of states in classical statistics, represented by probability distributions [1].

Definition 1 *Let $\mathfrak{S}$ be the convex set of states of a given system. A finite-valued measurement with outcomes in $\mathcal{X}, |\mathcal{X}| < \infty$, is an affine map $M : \mathfrak{S} \to \mathcal{P}(\mathcal{X})$ from the set of states to the set of all probability distributions on $\mathcal{X}$. For $\rho \in \mathfrak{S}$, we denote the map by $\rho \mapsto \mu_\rho^M$.*

The affinity condition on measurement is an axiomatic one. As such, for states $\rho_1, \rho_2 \in \mathfrak{S}$ and constant $0 \leqslant \lambda \leqslant 1$, there exists some state $\rho \in \mathfrak{S}$ such that $\mu_\rho^M = \lambda\mu_{\rho_1}^M + (1-\lambda)\mu_{\rho_2}^M$ for all $M \in \mathfrak{M}$. If $\mathfrak{S}$ is indeed a simplex, it is uniquely given by the convex hull of some fixed set of (pure) states $\mathfrak{B} = \{\rho_1 \ldots, \rho_d\}$ with $d := \dim(\mathfrak{S})$. $\mathfrak{B}$ is the set of extremal points of the simplex (see [Heinosaari et al.(2016)Heinosaari, Miyadera, and Ziman] for a more detailed account of the terms used). We will see that this condition dictates that all observables in classical statistics should be *jointly measurable* or equivalently *compatible.*

Definition 2 *Let $M : \mathfrak{S} \to \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ be a measurement defined by $\rho \mapsto \mu_\rho^M, \rho \in \mathfrak{S}$. Then $M_1 : \mathfrak{S} \to \mathcal{P}(\mathcal{X}_1)$, $\rho \mapsto \mu_\rho^{M_1}$ is a marginal measurement of M, if for all $\rho \in \mathfrak{S}, x \in \mathcal{X}_1$ we have*

$$\sum_{y \in \mathcal{X}_2} \mu_\rho^M(x, y) = \mu_\rho^{M_1}(x). \tag{1.1}$$

We can now define joint measurablity.

Definition 3 *Two measurements $M_1 : \mathfrak{S} \to \mathcal{P}(\mathcal{X}_1)$ and $M_2 : \mathfrak{S} \to \mathcal{P}(\mathcal{X}_2)$ are called jointly measurable if there exists measurement $M : \mathfrak{S} \to \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ such that M_1 and M_2 are marginal measurements of M.*
M_1 and M_2 are called compatible, if there exists a transition probability $\Pi : \mathcal{X}_1 \to \mathcal{P}(\mathcal{X}_2)$ such that for all $\rho \in \mathfrak{S}, y \in \mathcal{X}_2$

$$\mu_\rho^{M_2}(y) = \sum_{x \in \mathcal{X}_1} \Pi(y|x)\mu_\rho^{M_1}(x).$$

Joint measurablity and compatibility are equivalent (see e.g. [Filippov et al.(2017)Filippov, Heinosaari, and Leppäjärvi]).

[1] Here, we define the classical statistical model as one where the set of states is given by a simplex, and deduct compatibility of observables as the implication of this definition. An equivalent approach (taken e.g. by [Holevo(2012)]), is to define the classical statistical model as one where all measurements are compatible, and then prove a one to one affine map between the phase-space and the set of all probability distributions on a finite set.

The following statement exhibits the necessity for a statistical model with a broader class of states.

Proposition 4 *[Plávala(2016)] Let $\mathfrak{S}$ be a simplex and let $\mathfrak{B} := \{\rho_1, \ldots, \rho_d\}$ be the set of its extremal points. Then every measurement on $\mathfrak{S}$ is jointly measurable with every other measurement on $\mathfrak{S}$.*

Any statistical and information theoretic model that allows incompatible measurements, will have the following list of *impossible machines* (see e.g [Werner(2001)] Chapter 2 or [Wold(2012)]). Let $\mathfrak{S}^{classical}$ be the set of classical states and $\mathfrak{S}^{non-classical}$ be the set of states of a non-classical statistical model that allows incompatible measurements.

- Classical teleportation, whereby an unknown state $\rho_0 \in \mathfrak{S}^{non-classical}$ is mapped to some $\rho_1 \in \mathfrak{S}^{classical}$ and then converted to some $\rho_2 \in \mathfrak{S}^{non-classical}$ such that no statistical test could distinguish between ρ_0 and ρ_2.

- Cloning, whereby an unknown state $\rho_0 \in \mathfrak{S}^{non-classical}$ is taken as input, and two indistinguishable copies of ρ_0 are put out.

- Measurement without disturbing the system, whereby generally measurements leave the state of the system unchanged.

The above gives a hierarchy of machines, in the sense that existence of one enables the next. Existence of incompatible measurements implies existence of measurements that disturb the state of the system. In fact it can be shown that commutativity of measurements imply that they are jointly measurable ([Heinosaari et al.(2016)Heinosaari, Miyadera, and Ziman]). In turn, the fact that there are measurements that disturb the state of the system, imply that in general cloning is not possible. It is also evident that if classical teleportation were possible, one could clone unknown states by repeating the process. A natural requirement on an acceptable non-classical statistical model is that it reduces to the classical model when the set of states is reduced to a simplex. Quantum statistics is an example of a non-classical model where incompatible measurements are permitted. In fact in this sense, quantum statistics is a fairly general model ([Wolf et al.(2009)Wolf, Perez-Garcia, and Fernandez]). A state is described by a density operator, which can be represented by a Hermitian square matrix whose eigenvalues form a probability distribution (all positive semi-definite and adding up to unity). The convex set of states in quantum statistics is not a simplex. There are infinitely many decompositions into pure states for any given mixed state and after the state is prepared, there is no way of finding out which of the pure state ensembles were used in preparation ([Werner(2001)])[2]. These non-commutative statistics, as is readily obvious, call for a new information theoretic analysis by offering new possibilities. One of the first information-theoretic tasks

[2] This is based on the assumption that the statistical model is a *non-signaling* one. As such, another impossible machine is the *mixed state analyzer* whereby one determines the actual or refined pure state ensemble of a given mixed state.

that were considered using quantum statistics was hypothesis testing (see [Holevo(1973)]. We intend to use this statistical model in its elementary form (discrete and finite dimensional) to generalize and replace certain classical tasks, in the spirit of *Quantum Shannon Theory.*

1.3. Notations and conventions

All Hilbert spaces are assumed to have finite dimensions and are over the field $\mathbb{C}$. All alphabets are also assumed to have finite dimensions. We denote the set of states by $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{L}(\mathcal{H}) : \rho \geqslant 0, \mathrm{tr}(\rho) = 1\}$. Pure states are given by projections onto one-dimensional subspaces. To each subspace $\mathcal{F} \subset \mathcal{H}$, we can associate a 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.

$$\pi_{\mathcal{F}} := \frac{q_{\mathcal{F}}}{\mathrm{tr}(q_F)}. \tag{1.2}$$

$\mathcal{C}^{\downarrow}(\mathcal{H}_A, \mathcal{H}_B)$ stands for the set of completely positive trace non-increasing maps between $\mathcal{L}(\mathcal{H}_A)$ and $\mathcal{L}(\mathcal{H}_B)$. In what follows, $\mathcal{U}(\mathcal{H})$ will denote the group of unitary operators acting on $\mathcal{H}$. For a Hilbert space $\mathcal{G} \subset \mathcal{H}$, we will always identify $\mathcal{U}(\mathcal{G})$ with a subgroup of $\mathcal{U}(\mathcal{H})$. For any projection $q \in \mathcal{L}(\mathcal{H})$ we set $q^{\perp} := 1_{\mathcal{H}} - q$.

Each projection $q \in \mathcal{L}(\mathcal{H})$ defines a completely positive trace non-increasing map Q given by $Q(a) := qaq$ for all $a \in \mathcal{L}(\mathcal{H})$. In a similar fashion, any $U \in \mathcal{U}(\mathcal{H})$ defines a $\mathcal{U} \in \mathcal{C}(\mathcal{H}, \mathcal{H})$ by $\mathcal{U}(a) := UaU^{\dagger}$ for $a \in \mathcal{L}(\mathcal{H})$. The coherent information for $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ and $\rho \in \mathcal{S}(\mathcal{H}_A)$ is defined by

$$I_c(\rho, \mathcal{N}) := S(\mathcal{N}(\rho)) - S((\mathrm{id}_{\mathcal{H}_A} \otimes \mathcal{N})(|\psi\rangle\langle\psi|)), \tag{1.3}$$

where $\psi \in \mathcal{H}_A \otimes \mathcal{H}_A$ is an arbitrary purification of the state ρ. A short-hand notation $S_e(\rho, \mathcal{N}) := S((\mathrm{id}_{\mathcal{H}_A} \otimes \mathcal{N})(|\psi\rangle\langle\psi|))$ to denote entropy exchange is also used in the literature. A useful equivalent definition of $I_c(\rho, \mathcal{N})$ is given in terms of $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ and any complementary channel $\hat{\mathcal{N}} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_e)$ where $\mathcal{H}_e$ denotes the Hilbert space of the environment. Due to Stinespring's dilation theorem (see [Hsieh and Wilde(2009)]), $\mathcal{N}$ can be represented as

$$\mathcal{N}(\rho) = \mathrm{tr}_{\mathcal{H}_e}(v\rho v^*) \tag{1.4}$$

for $\rho \in \mathcal{S}(\mathcal{H}_A)$, where $v : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_e$ is a linear isometry. The complementary channel $\hat{\mathcal{N}} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_e)$ of $\mathcal{N}$ is given by

$$\hat{\mathcal{N}}(\rho) := \mathrm{tr}_{\mathcal{H}_B}(v\rho v^*). \tag{1.5}$$

The coherent information can then be written as

$$I_c(\rho, \mathcal{N}) = S(\mathcal{N}(\rho)) - S(\hat{\mathcal{N}}(\rho)). \tag{1.6}$$

This quantity can also be defined in terms of the bipartite state $\sigma \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with

$$\sigma := \mathrm{id}_{\mathcal{H}_A} \otimes \mathcal{N}(|\psi\rangle\langle\psi|) \tag{1.7}$$

as

$$I(A \rangle B, \sigma) := S(\sigma^B) - S(\sigma), \tag{1.8}$$

where σ^B is the marginal state given by $\sigma^B := \mathrm{tr}_A(\sigma)$ and we have the identity

$$I_c(\rho, \mathcal{N}) = I(A \rangle B, \sigma). \tag{1.9}$$

For the approximation of arbitrary compound channels (introduced in the next section) by finite ones, we use the diamond norm $\| \cdot \|_\diamond$, given for any $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ by

$$\| \mathcal{N} \|_\diamond := \sup_{n \in \mathbb{N}} \max_{a \in \mathcal{L}(\mathbb{C}^n \otimes \mathcal{H}), \|a\|_1 = 1} \| (\mathrm{id}_n \otimes \mathcal{N})(a) \|_1, \tag{1.10}$$

where $\mathrm{id}_n : \mathcal{L}(\mathbb{C}^n) \to \mathcal{L}(\mathbb{C}^n)$ is the identity channel. We state the following facts about $|| \cdot ||_\diamond$ (see e.g [Kitaev et al.(2002)Kitaev, Shen, and Vyalyi]). First, $||\mathcal{N}||_\diamond = 1$ for all $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$. Thus, $\mathcal{C}(\mathcal{H}_A, \mathcal{H}_B) \subset S_\diamond$, where $S_\diamond$ denotes the unit sphere of the normed space $(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B), || \cdot ||_\diamond)$. Moreover, $||\mathcal{N}_1 \otimes \mathcal{N}_2||_\diamond = ||\mathcal{N}_1||_\diamond ||\mathcal{N}_2||_\diamond$ for arbitrary linear maps $\mathcal{N}_1, \mathcal{N}_2 : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$. Throughout this section we have made use of the idea of nets to approximate arbitrary compound quantum channels using ones with finite uncertainty sets. This idea is presented in Appendix B.

The set of probability distributions on the finite alphabet $\mathcal{X}$ of cardinality $|\mathcal{X}|$ will be denoted by $\mathcal{P}(\mathcal{X})$. For $n \in \mathbb{N}$, we define $\mathcal{X}^n := \quad (x_1, \ldots, x_n) : x_i \in \mathcal{X}, \forall i \in \{1, \ldots, n\}\big\}$. The sequence $\mathbf{x}$ will denote elements of $\mathcal{X}^n$. Also, we use bold letters to denote vectors (sequences with more that one element). The probability distribution $p^{\otimes n} \in \mathcal{P}(\mathcal{X}^n)$ will be given by the n-fold product of $p \in \mathcal{P}(\mathcal{X})$, namely $p^{\otimes n}(\mathbf{x}) = p(x_1) \ldots p(x_n)$ with $\mathbf{x} = (x_1, \ldots, x_n)$. For any number $M \in \mathbb{N}$, we use $[M] := \{1, \ldots, M\}$.

The classical quantum (cq) channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ is a completely positive trace preserving map from alphabet $\mathcal{X}$ to the set of states on Hilbert space $\mathcal{H}$. We denote the set of all such maps by $\mathcal{CQ}(\mathcal{X}, \mathcal{H})$. This set is equipped with the norm $\| \cdot \|_{CQ}$ defined for $W \in CQ(\mathcal{X}, \mathcal{H})$ by

$$\| W \|_{CQ} := \max_{x \in \mathcal{X}} \| W(x) \|_1, \tag{1.11}$$

where $\| \cdot \|_1$ is the trace norm on $\mathcal{L}(\mathcal{H})$. We use the term cqq channel for map $V \in CQ(\mathcal{X}, \mathcal{H}_1 \otimes \mathcal{H}_2)$ with two outcomes in two sets of states on two Hilbert spaces. With a slight abuse of notation, we write $a^c := \mathbb{1}_{\mathcal{H}} - a$ for $a \in \mathcal{L}(\mathcal{H})$.

We use $\epsilon_n \to 0$ exponentially as $n \to \infty$ or we say ϵ_n approaches (goes to) zero exponentially, if $-\frac{1}{n} \log \epsilon_n$ is a strictly positive constant. For $\epsilon_{1,n}$ and $\epsilon_{2,n}$ both approaching zero exponentially, we use $\epsilon_{1,n} \geqslant \epsilon_{2,n}$ if $-\frac{1}{n} \log \epsilon_{1,n} \leqslant -\frac{1}{n} \log \epsilon_{2,n}$. We use $\mathrm{cl}(A)$ to denote the closure of set A and finally, we use $\mathfrak{S}_n$ to denote the group of permutations on n elements such that $\alpha(s^n) = (s_{\alpha(1)}, \ldots, s_{\alpha(n)})$ for each $\alpha \in \mathfrak{S}_n$ and $s^n = (s_1, \ldots, s_n) \in S^n$.

A measurement or a positive operator valued measure (POVM) with $M \in \mathbb{N}$ outcomes on Hilbert space $\mathcal{H}$, is given by an M-tuple $(D_1, \ldots, D_M) : D_i \geqslant 0, \forall i \in [M]$ and $\sum_{i \in [M]} D_i = \mathbb{1}_{\mathcal{H}}$. With slight abuse of notation, we write $a^c := \mathbb{1}_{\mathcal{H}} - a$ for $a \in \mathcal{L}(\mathcal{H})$.

Given the state $\omega_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, a closely related quantity to coherent information is the mutual information that is given by

$$I(A; B, \omega) := S(A, \omega) + S(B, \omega) - S(AB, \omega),$$

where $S(\gamma, \omega)$, indicates the von Neumann entropy of the state ω_γ, the marginal state of ω. Consider the ensemble $\{p(x), \omega_{AB}^x\}$ with $\omega_{AB}^x \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $p \in \mathcal{P}(\mathcal{X})$. We can define a classical-quantum (cq) state $\omega_{XAB} \in \mathcal{S}(\mathbb{C}^{|\mathcal{X}|} \otimes \mathcal{H}_A \otimes \mathcal{H}_B)$, given some ONB $\{e_x\}_{x \in \mathcal{X}} \in \mathbb{C}^{|\mathcal{X}|}$ as

$$\omega_{XAB} := \sum_{x \in \mathcal{X}} p(x) \left|e_x\right\rangle\!\left\langle e_x\right|^X \otimes \omega_{AB}^x \tag{1.12}$$

Note that we have used the suffix X to label the Hilbert space corresponding to alphabet $\mathcal{X}$. The conditional mutual information is then defined by

$$I(A; B|X, \omega_{XB}) := \sum_{x \in \mathcal{X}} I(A; B, \omega_{AB}^x). \tag{1.13}$$

2. Randomness cost of symmetric twirling

In this chapter, we study random unitary channels which reproduce the action of the twirling channel corresponding to the representation of the symmetric group on an n-fold tensor product. We derive upper and lower bounds on the randomness cost of implementing such a map which depend exponentially on the number of systems. Consequently, symmetric twirling can be regarded as a reasonable Shannon-theoretic protocol. On the other hand, such protocols are disqualified by their resource-inefficiency in situations where randomness is a costly resource.

2.1. Introduction

When designing communication protocols, the quantum information theorist has a vast and steadily growing toolbox of approved protocol parts at hand. Especially useful are universal protocols, which perform a certain task regardless of the preparation of the system.

As a prominent example of this class we mention the quantum teleportation protocol which allows noiseless transmission of an unknown qubit state by using a pure maximally entangled qubit pair and two bits of noiseless forward communication. The fact that the teleportation protocol perfectly accomplishes this goal is completely independent of the state to be transmitted, motivates modular use in larger protocols without further adjustment of the protocol. In this chapter, we address symmetric twirling, which perfectly and universally transforms each state on a given n-party system to a permutation invariant one. This is accomplished by applying a unitary U^π which exchanges the subsystems according to a permutation π which is chosen randomly according to the equidistribution on the group S_n of permutations on n elements, i.e. the quantum channel

$$\mathcal{U}(\cdot) := \frac{1}{|S_n|} \sum_{\pi \in S_n} U^\pi (\cdot) U^{\pi \dagger}.$$

is applied.

This protocol is very useful in in situations where the system is demanded to be permutationally invariant for further processing. An example of such a situation is, where $\mathcal{U}$ is performed to make a system ready for applying an instance of the quantum de Finetti Theorem (see for example [Christandl et al.(2009)Christandl, König, and Renner].) While application of $\mathcal{U}$ makes all states on the underlying systems perfectly permutation-invariant, the protocol is highly demanding regarding its randomness cost. Since $n!$ grows super-exponentially with the number n of systems, the randomness cost of the protocol is not bounded by any rate. This fact prevents $\mathcal{U}$ from being a reasonable protocol in situations where randomness is at all counted as a resource. However, the equidistributed choice out of all permutations obviously bares some redundancies, such that a less randomness consuming way of choosing permutations to emulate $\mathcal{U}$ seems possible.

In Section 2.2 of this chapter, we derive upper and lower bounds on the randomness needed to perform the symmetric twirling channel $\mathcal{U}$, both of which lie on the exponential scale. We also show, that the lower bound essentially remains valid under the weakened condition, that the action of the twirling is simulated only approximately well.

In Section 2.3, we discuss the consequences of our findings for communication theory. The upper bounds derived show that the action of symmetric twirling indeed can be accomplished universally by a protocol with rate-bounded randomness demands. This fact is important in situations where randomness is not a free resource (e.g. when the random permutations have to be applied by two or more parties in a coordinated way.) On the other hand, the lower bounds derived show, that the randomness needed is close to the maximum randomness which can be generated from that system. Therefore, symmetric twirling is too expensive in some situations. Such situations arise especially, when the random choice of permutations has to be kept private from additional adversarial communication parties.

The twirling and determination of randomness needed to perform such was extensively studied in the work [Gross et al.(2007)Gross, Audenaert, and Eisert] in case of the group of unitary transformations on a given Hilbert space. Therein, the notion of a *unitary design* was introduced, a terminology which we extend to the symmetric group in this work.

Recently, Wakakuwa [Wakakuwa(2017)] determined the asymptotic randomness cost of symmetrizing a given quantum state in case of tensor product representations of an arbitrary given group. We point out, that the focus set in [Wakakuwa(2017)] is different from this work. Namely, the twirling we consider does not arise from a tensor product representation and is therefore out of the scope of [Wakakuwa(2017)]. Moreover, we are focused on protocols which emulate the twirling operation *universally*, while the mentioned work rather asks for the randomness cost of simulating the action of a twirling for a fixed state.

2.2. Bounds for symmetric designs

Considerations briefly explained in the introduction, have brought up the question of whether it is possible to render the average over a group using only a subset of its elements. In particular, S_n, the group of all permutations on n elements (or equivalently all bijections over $\{1,...,n\}$), will be of interest in the remainder of our work. We consider the unitary representation $\{U^\pi\}_{\pi\in S_n}$ of this group acting on $(\mathbb{C}^d)^{\otimes n}$, defined by the action:

$$U^\pi x_1 \otimes ... \otimes x_n = x_{\pi(1)} \otimes ... \otimes x_{\pi(n)} \tag{2.1}$$

for each $\pi \in S_n$ and $x_1, \ldots, x_n \in \mathbb{C}^d$. We prove bounds for all weighted subsets of S_n, the unitary representations of which produce the group's average. We refer to such subsets as *symmetric weighted designs*, as they are the analogous objects to spherical or unitary designs ([Gross et al.(2007)Gross, Audenaert, and Eisert]).

Definition 5 *Let $X \subset S_n$ and $\omega : X \to \mathbb{R}^+$ be a weight function (i.e. $\omega > 0$ and $\sum_{\pi\in X} \omega(\pi) = 1$). The pair (X, ω) is a symmetric weighted design (or a weighted design for S_n), if:*

$$\frac{1}{n!}\sum_{\pi\in S_n} U^\pi \eta U^{\pi\dagger} = \sum_{\pi\in X} \omega(\pi) U^\pi \eta U^{\pi\dagger} \tag{2.2}$$

for all $\eta \in \mathcal{L}(\mathcal{H}^{\otimes n})$, where $\mathcal{H} := \mathbb{C}^d$.

To prove the upper bound on the cardinality of the designs we use the following theorem from convex analysis. A proof can be found in e.g. [Barvinok(2003)], Theorem 2.3.

Theorem 6 (Carathéodory's Theorem) *Let $S \subset \mathbb{R}^d$ be a set. Then every point $x \in$ $\mathrm{conv}(S)$ can be represented as a convex combination of $d+1$ points from S, i.e. there exist $\alpha_1, \ldots, \alpha_{d+1} \geqslant 0$, $\sum_{i=1}^{d+1} \alpha_i = 1$, and $y_1, \ldots, y_n \in S$ such that*

$$x = \alpha_1 y_1 + \cdots + \alpha_{d+1} y_{d+1} \tag{2.3}$$

holds.

Theorem 7 *There exists a symmetric weighted design (X, ω) with cardinality of X upper-bounded as:*

$$|X| \leqslant d^{4n} + 1 \tag{2.4}$$

Proof 8 *Let $B := \{|e_x\rangle : x \in [d]\}$ be the standard basis for $\mathbb{C}^d$. We will use the notation*

$$|e_{\mathbf{x}}\rangle := |e_{x_1}\rangle \otimes \cdots \otimes |e_{x_n}\rangle \tag{2.5}$$

for each $\mathbf{x} := (x_1, \ldots, x_n) \in \mathcal{X}^n$. Writing the left and right hand sides of (2.2) in terms *of matrix entries of U^π and $U^{\pi\dagger}$ ($u^{\pi\dagger}_{ij} : \langle e_i| U^{\pi\dagger} |e_j\rangle, u^\pi_{ij} : \langle e_i| U^\pi |e_j\rangle$) we obtain:*

$$L := \frac{1}{n!} \sum_{\pi \in S_n} U^\pi \eta U^{\pi\dagger} =$$

$$\frac{1}{n!} \sum_{\pi \in S_n} \sum_{\mathbf{w},\mathbf{x},\mathbf{y},\mathbf{z} \in [d^n]} a_{\mathbf{xy}} u^\pi_{\mathbf{wx}} u^{\pi\dagger}_{\mathbf{yz}} |e_{\mathbf{w}}\rangle\langle e_{\mathbf{z}}| \tag{2.6}$$

and

$$R := \sum_{\pi \in X} \omega(\pi) \sum_{\mathbf{w},\mathbf{x},\mathbf{y},\mathbf{z} \in [d^n]} a_{\mathbf{xy}} u^\pi_{\mathbf{wx}} u^{\pi\dagger}_{\mathbf{yz}} |e_{\mathbf{w}}\rangle\langle e_{\mathbf{z}}| \tag{2.7}$$

where $a_{\mathbf{xy}} := \langle e_{\mathbf{x}}| \eta |e_{\mathbf{y}}\rangle$. Since $a_{\mathbf{xy}}$ only depends on η, it can be observed that $R = L$ (and hence (X, ω) is a symmetric weighted design) if we have:

$$\frac{1}{n!} \sum_{\pi \in S_n} u^\pi_{\mathbf{wx}} u^{\pi\dagger}_{\mathbf{yz}} = \sum_{\pi \in X} \omega(\pi) u^\pi_{\mathbf{wx}} u^{\pi\dagger}_{\mathbf{yz}} \tag{2.8}$$

for all $\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z} \in [d^n] := \{1, ..., d\}^n$. Define the vector $\nu_\pi := (u^\pi_{\mathbf{wx}} u^{\pi\dagger}_{\mathbf{yz}} : \mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z} \in [d^n])$. We observe that $\nu_\pi \in \mathbb{R}^{d^{4n}}$, as the entries of U^π are either equal to zero or one:

$$U^\pi |e_{\mathbf{x}}\rangle = |e_{\pi(\mathbf{x})}\rangle$$

and hence:

$$\langle e_{\mathbf{y}}| U^\pi |e_{\mathbf{x}}\rangle = 1 \textit{ if } \pi(\mathbf{x}) = \mathbf{y} \textit{ and } 0 \textit{ otherwise}$$

Define the set $\Omega_A := \{\nu_\pi : \pi \in A\}$ for some $A \subset S_n$. The point $p := \frac{1}{n!} \sum_{\pi \in S_n} \nu_\pi$ is in the convex hull of Ω_{S_n}:

$$p \in \textit{ conv } (\Omega_{S_n}) \tag{2.9}$$

where

$$\textit{conv } (\Omega_{S_n}) := \{ \sum_{\pi \in S_n} \alpha_\pi |\nu_\pi\rangle : \forall \alpha_\pi \geqslant 0, \sum_{\pi \in S_n} \alpha_\pi = 1\}$$

At this point, we can apply the Carathéodory's theorem stated above, to complete our proof. We observe that $\Omega_{S_n} \subset \mathbb{R}^{d^{4n}}$, and hence by Carathéodory's theorem, there exists a subset $X \subset S_n$ such that $p \in conv(\Omega_X)$ and $|X| \leqslant d^{4n} + 1$. Therefore there exists a weight function ω on X such that:

$$\sum_{\pi \in X} \omega(\pi) \nu_\pi = p \tag{2.10}$$

which fulfills (2.8).

The above stated bound can be also formulated in terms of entropies. Let $H(p)$ denote the Shannon entropy of the probability distribution p on an alphabet $\mathcal{X}$, i.e.

$$H(p) := -\sum_{x\in\mathcal{X}} p(x)\log p(x),$$

where we use the convention that log denotes base 2 logarithms. The cardinality bound in Theorem 7 implies, that we find a weighted symmetric design (X,ω) with

$$\frac{1}{n}H(\omega) \leqslant 4\log(d+1) \tag{2.11}$$

Next, we will prove a lower bound on the Shannon entropy of symmetric designs which complements the upper bound from Eq. (2.11).

Remark 9 *The weight function ω over X as defined in Def.5 is a special case of a probability distribution over S_n. This can be observed by setting $\omega(\pi) = 0$ for all $\pi \notin X$. When dealing with entropies, we consider ω to be a probability distribution over S_n, and hence derive a lower bound on entropy of any convex combination of permutation unitaries that produces the desired average over the group.*

In what follows, we set $\mathcal{U}_\pi(\cdot) := U_\pi(\cdot)U_\pi^\dagger$ $(\pi \in S_n)$.

Theorem 10 *Let (X,ω) be a symmetric weighted design. Then:*

$$\frac{1}{n}H(\omega) \geqslant \log(d) - 2d\frac{\log(n+1)}{n} \tag{2.12}$$

where $H(\omega)$ is the Shannon entropy of the weight function.

Proposition 11 (Almost-convexity of the von Neumann entropy) *Let p be a probability distribution on $\mathcal{X}$, $|\mathcal{X}| < \infty$, ρ_x be a density matrix on $\mathcal{H}$ where* $\dim(\mathcal{H}) = d$, *for each $x \in \mathcal{X}$, and set $\overline{\rho}_p := \sum_{x\in\mathcal{X}} p(x)\rho_x$. It holds*

$$S(\overline{\rho}) \leqslant \sum_{x\in\mathcal{X}} p(x)S(\rho_x) + H(p). \tag{2.13}$$

Proof 12 *See e.g. [Nielsen and Chuang(2010)], Theorem 11.10.*

Proof 13 (Proof of Theorem10) *Fix $n \in \mathbb{N}$, and set $\mathcal{X} := \{1,\dots,d\}$ and let μ be a type of sequences in $\mathcal{X}^n$ with*

$$H(\mu) \geqslant \log d - d\frac{\log(n+1)}{n}. \tag{2.14}$$

Notice that existence of such a type is guaranteed by Lemma 191(see Appendix A, where more definitions and statements on frequency typical sets can also be found). Define the *projection p_μ by*

$$p_\mu := \sum_{\mathbf{x}\in T_\mu^n} |e_\mathbf{x}\rangle\langle e_\mathbf{x}|\,.$$

First, we notice, that for each μ-typical word $\mathbf{x}$,

$$\frac{1}{n!}\sum_{\pi\in S_n} \mathcal{U}_\pi(|e_\mathbf{x}\rangle\langle e_\mathbf{x}|) = \frac{1}{|T_\mu^n|}p_\mu \tag{2.15}$$

holds (on the r.h.s. of the above equality, we find the maximally mixed state on the subspace of $\mathcal{H}^{\otimes n}$ belonging to the type class T_μ^n). We fix a μ-typical word $\mathbf{x}$ and set $E : |e_\mathbf{x}\rangle\langle e_\mathbf{x}|$. We can bound the Shannon entropy of ω by

$$\begin{aligned} H(\omega) &\geqslant S\left(\sum_{\pi\in S_n}\omega(\pi)\mathcal{U}_\pi(E)\right) - \sum_{\pi\in S_n}\omega(\pi)S(\mathcal{U}_\pi(E)) \\ &= S\left(\frac{1}{n!}\sum_{\pi\in S_n}\mathcal{U}_\pi(E)\right) \\ &= S\left(\frac{1}{|T_\mu^n|}p_\mu\right) \\ &= \log|T_\mu^n|. \end{aligned}$$

The inequality above is by Proposition 11. The first equality is by the fact, that $\mathcal{U}_\pi(E)$ is a pure state for each $\pi \in S_n$ combined with the hypothesis of the lemma that (X,ω) is a weighted design defined by (2.2). The second equality is by (2.15). We conclude

$$H(\omega) \geqslant n\cdot H(\mu) - \log(n+1)^d \geqslant n\cdot H(\mu) - 2\cdot\log(n+1)^d.$$

The left inequality above is by the standard type bound

$$|T_\mu^n| \geqslant \frac{1}{(n+1)^d}\cdot 2^{nH(\mu)},$$

while the second is by choice of μ, i.e. by the bound from (2.14). We are done.

The above reasoning can be extended to derive a bound for averages of permutations which approximately simulate the action of the uniform average over S_n. To formulate such an assertion, we use the diamond norm $\|\cdot\|_\diamond$ on the set of quantum channels on a Hilbert space $\mathcal{K}$. We define

$$\|\mathcal{N}\|_\diamond = \sup_{n\in\mathbb{N}} \max_{\substack{a\in\mathcal{L}(\mathbb{C}^n\otimes\mathcal{K})\\ \|a\|_1=1}} \|\mathrm{id}_{\mathbb{C}^n}\otimes\mathcal{N}(a)\|_1 \tag{2.16}$$

for each c.p.t.p. map on $\mathcal{K}$. We define c.p.t.p. maps

$$\overline{\mathcal{U}}(b) := \frac{1}{n!}\sum_{\pi\in S_n}\mathcal{U}_\pi(b) \qquad (b\in\mathcal{L}(\mathcal{H}^{\otimes n})), \tag{2.17}$$

and

$$\overline{\mathcal{U}}_q(b) := \sum_{\pi\in S_n} q(\pi)\mathcal{U}_\pi(b) \qquad (b\in\mathcal{L}(\mathcal{H}^{\otimes n})) \tag{2.18}$$

for each probability distribution q on S_n. We prove

Theorem 14 *It holds*

$$\frac{1}{n}H(q) \geqslant \log(d) - 2d\frac{\log(n+1)}{n} - \tfrac{1}{n}f(\|\overline{\mathcal{U}} - \overline{\mathcal{U}}_q\|_\diamond) \tag{2.19}$$

for each probability distribution q on S_n, where $\frac{1}{n}f(x)\to 0$, $(x\to 0)$. More specifically, $f(x) := 2x\log(d-1) + 2H_2(x)$ where $H_2(x)$ is the binary entropy and d is the dimension of the underlying Hilbert space.

Proof 15 *The proof is by minor extension of the argument given to prove Theorem 10. Note, that with $E := |e_{\mathbf{x}}\rangle\langle e_{\mathbf{x}}|$ as in the proof of Theorem 10*

$$\epsilon := \|\overline{\mathcal{U}} - \overline{\mathcal{U}}_q\|_\diamond \geqslant \|(\overline{\mathcal{U}} - \overline{\mathcal{U}}_q)(E)\|_1 \tag{2.20}$$

holds. By a sharp version of Fannes' inequality due to Audenaert ([Audenaert(2007)]), we have

$$S(\overline{\mathcal{U}}_q) \geqslant S(\overline{\mathcal{U}}) - f(\epsilon) \tag{2.21}$$

with a function fulfilling $\frac{1}{n}f(\epsilon)\to 0$ $(\epsilon\to 0)$. We repeat the line of reasoning from the proof of Theorem 10 including the above tradeoff to the inequalities and get

$$\begin{aligned} H(q) &\geqslant S(\overline{\mathcal{U}}_q(E)) &(2.22)\\ &\geqslant S(\overline{\mathcal{U}}(E)) - f(\epsilon) &(2.23)\\ &\geqslant n\log d - 2d(n+1) - f(\epsilon). &(2.24)\end{aligned}$$

The bounds obtained so far directly imply corresponding bounds for completely positive and trace preserving (c.p.t.p.) matrices.

Theorem 16 *Let $\dim\mathcal{K} := d_{\mathcal{K}}$, $\dim\mathcal{H} := d_{\mathcal{H}}$, and $\mathcal{U}_\pi(\cdot) := U^\pi(\cdot)(U^\pi)^*$, $\mathcal{V}_\pi(\cdot) := V^\pi(\cdot)(V^\pi)^*$ be the c.p.t.p. maps permuting the tensor factors on $\mathcal{L}(\mathcal{H})^{\otimes n}$ resp. $\mathcal{L}(\mathcal{K})^{\otimes n}$*

according to π for each $\pi \in S_n$. If

$$\frac{1}{n!}\sum_{\pi \in S_n} \mathcal{U}_\pi \circ \mathcal{N} \circ \mathcal{V}_{\pi^{-1}} = \sum_{\pi \in S_n} \omega(\pi)\mathcal{U}_\pi \circ \mathcal{N} \circ \mathcal{V}_{\pi^{-1}} \tag{2.25}$$

for each c.p.t.p. map $\mathcal{N} : \mathcal{L}(\mathcal{H}^{\otimes n}) \to \mathcal{L}(\mathcal{K}^{\otimes n})$, then

$$\frac{1}{n}H(\omega) \geqslant \log(d_\mathcal{K} d_\mathcal{H}) - 2d_\mathcal{K} d_\mathcal{H}\frac{\log(n+1)}{n} \tag{2.26}$$

Proof 17 *The proof of the above assertion almost immediately follows from Theorem 10 combined with the Jamiołkowski isomorphism (see e.g. [Wilde(2017)])*

$$\mathcal{N} \mapsto \sigma_\mathcal{N} := \mathcal{N} \otimes \mathrm{id}_{\mathcal{H}^{\otimes n}}(|\Phi\rangle\langle\Phi|), \tag{2.27}$$

where $|\Phi\rangle$ is defined by

$$|\Phi\rangle := \frac{1}{d^n}\sum_{\mathbf{x}\in\mathcal{X}^n} |e_\mathbf{x}\rangle \otimes |e_\mathbf{x}\rangle \tag{2.28}$$

Indeed, for each c.p.t.p. map $\mathcal{N} : \mathcal{L}(\mathcal{H})^{\otimes n} \to \mathcal{L}(\mathcal{K})^{\otimes n}$, it holds

$$\begin{aligned} \sigma_{\mathcal{U}_\pi \circ \mathcal{N} \circ \mathcal{V}_{\pi^{-1}}} &= \mathcal{U}_\pi \circ \mathcal{N} \circ \mathcal{V}_{\pi^{-1}} \otimes \mathrm{id}_{\mathcal{H}^{\otimes n}}(|\Phi\rangle\langle\Phi|) && (2.29) \\ &= \mathcal{U}_\pi \otimes \mathcal{V}_\pi(\sigma_\mathcal{N}). && (2.30) \end{aligned}$$

A lower bound on the cardinality of designs (and 2-designs by a straightforward extension) can be readily established from Theorem 10. We finish this section, however, by remarking a relation between vectors belonging to the symmetric subspace and permutation invariant states, that in turn enables us to derive a lower bound on the cardinality of designs.

It can be observed that permutation invariant matrices are not in general supported on $\mathrm{sym}^{(n)}(\mathcal{H})$, the subspace defined by:

$$\mathrm{sym}^{(n)}(\mathcal{H}) := \mathrm{span}(|\nu\rangle : U^\pi |\nu\rangle = |\nu\rangle \forall \pi \in S_n)$$

An example to the point is $M = |e_{01}\rangle\langle e_{01}| + |e_{10}\rangle\langle e_{10}|$ where $|e_{ij}\rangle = |e_i\rangle \otimes |e_j\rangle$. The following lemma from [Renner(2005)] can be used to establish a relation between permutation invariant states and vectors on $\mathrm{sym}^{(n)}(\mathcal{H})$:

Lemma 18 ([Renner(2005)], Lemma 4.2.2) *Let the state $\rho_n \in S(\mathcal{H}^{\otimes n})$ be permutation invariant and have the following spectral decomposition:*

$$\rho_n := \sum_i \lambda_i |\nu_i\rangle\langle\nu_i|$$

where we have included the zero eigenvalues. Then $|\psi\rangle := \sum_i \sqrt{\lambda_i}\, |\nu_i\rangle \otimes |\nu_i\rangle \in sym^{(n)}(\mathcal{H})$.

Using this lemma, we prove a lower bound on the cardinality of symmetric weighted designs:

Theorem 19 *Let* (X, ω) *be a weighted design for* S_n*. Then we have:*

$$|X| \geqslant d^n - \binom{d+n-1}{d-1} \tag{2.31}$$

Proof 20 *Consider* $|\nu\rangle \in sym^{(n)}(\mathcal{H})^{\perp}$, *where the superscript indicates the orthogonal compliment. It can be observed that* $U^{\pi}|\nu\rangle \in sym^{(n)}(\mathcal{H})^{\perp} \forall \pi \in S_n$*. To see this, we notice that* $\forall |\psi\rangle \in sym^{(n)}(\mathcal{H})$ *we have* $\langle\psi| U^{\pi} |\nu\rangle = \langle\psi|\nu\rangle = 0$*. The second equality is due to the fact that* $|\psi\rangle$ *is permutation invariant and absorbs* U^{π}*. Consider the set* $V := \{U^{\pi}|\nu\rangle\}_{\pi\in X}$ *for some* $X \subset S_n$*. If* $|X| < \dim(sym^{(n)}(\mathcal{H})^{\perp})$,*we can orthonormalize this set via Gram-Schmidt process and obtain* $V' := \{|\nu^{\pi}\rangle\}_{\pi\in X}$*.* V' *would then be an ONB for a subspace of* $sym^{(n)}(\mathcal{H})^{\perp}$*. Finally, define* $\tilde{V} := \{|\nu^{\pi}\rangle \otimes |\nu^{\pi}\rangle\}_{\pi\in X}$*. It can be observed that* $|\nu^{\pi}\rangle \otimes |\nu^{\pi}\rangle \in sym^{(n)}(\mathcal{H}\otimes\mathcal{H})^{\perp}$*. There are two possibilities for any linear combination with non-zero multiples of elements in* $\tilde{V}$*: for any set* $\{\lambda^{\pi} \neq 0, \pi \in X\}$ *either:*

$$1. \sum_{\pi\in X} \lambda^{\pi} |\nu^{\pi}\rangle \otimes |\nu^{\pi}\rangle = 0$$

or

$$2. \sum_{\pi\in X} \lambda^{\pi} |\nu^{\pi}\rangle \otimes |\nu^{\pi}\rangle \neq 0 \text{ and } \in sym^{(n)}(\mathcal{H}\otimes\mathcal{H})^{\perp}$$

But the first case cannot be, as $\{|\nu^{\pi}\rangle \otimes |\nu^{\pi}\rangle\}_{\pi\in X}$ *is linearly independent for* $|X| < \dim(sym^{(n)}(\mathcal{H})^{\perp})$*. The second case, by Lemma 18 implies that the state* $\sigma := \sum_{\pi\in X} (\lambda^{\pi})^2 |\nu^{\pi}\rangle\langle\nu^{\pi}|$ *cannot be permutation invariant. Since* λ^{π} *is any non-zero number, this is true for all linear combinations of states* $|\nu^{\pi}\rangle\langle\nu^{\pi}|$ *as long as* $|X| < \dim(sym^{(n)}(\mathcal{H})^{\perp})$*. But what does this imply for linear combinations of states* $U^{\pi}|\nu\rangle\langle\nu| U^{\pi\dagger}$ *for* $\pi \in X$*. For any such state*

$$\mu := \sum_{\pi\in X} \omega^{\pi} U^{\pi} |\nu\rangle\langle\nu| U^{\pi\dagger}$$

we have:

$$\mu = \sum_{\pi\tilde{\pi}\in X} \gamma^{\pi\tilde{\pi}} |\nu^{\pi}\rangle\langle\nu^{\tilde{\pi}}|$$

In the ONB given by V'*, the right hand side can be decomposed into a diagonal matrix* Q *and an off-diagonal matrix* R*. The diagonal matrix is a linear combination of states* $|\nu^{\pi}\rangle\langle\nu^{\pi}|$ *and hence cannot be permutation invariant by arguments given above. But for* $\mu = Q + R$ *to be permutation invariant, both* Q *and* R *have to be permutation invariant, as application of any unitary on* μ *will produce a diagonal matrix and an off-diagonal one, cancelling out* Q *and* R *respectively when considering* $\mu - U^{\pi}\mu U^{\pi\dagger}$*.*

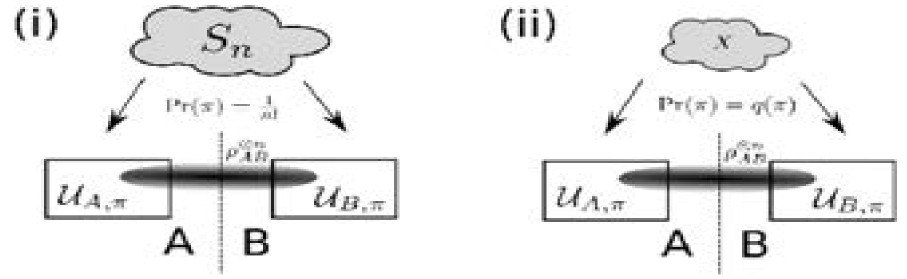

Figure 2.1.: (i) Implementation of $\overline{\mathcal{U}}$ by equidistributed and correlated random choice of permutations. (ii) Simulation of $\overline{\mathcal{U}}$ by correlated choice of a random permutation from a smaller set X according to probability distribution q

2.3. Communication-theoretic implications of the results

In this section, we discuss some consequences of the technical results from the the last section. From the upper and lower bounds derived there, some remarkable conceptual implications in communication theory can be drawn.

Assuming $\mathcal{H}$ as the underlying Hilbert space of the system under consideration, the quantum channel

$$\overline{\mathcal{U}}(\cdot) := \frac{1}{n!}\sum_{\pi\in S_n}\mathcal{U}_\pi(\cdot), \quad \mathcal{U}_\pi(\cdot) := U^\pi(\cdot)U^{\pi\dagger} \qquad (\pi \in S_n)$$

is usually regarded as the standard protocol applied to universally map each state on $\mathcal{H}^{\otimes n}$ to a permutation invariant one. [1]

To zest the discussion, we consider $\mathcal{H} := \mathcal{H}_A \otimes \mathcal{H}_B$ the space of a bipartite system shared by distant communication parties A and B. The corresponding map $\overline{\mathcal{U}}$ on $\mathcal{H}$ has the form

$$\overline{\mathcal{U}}(\cdot) = \frac{1}{n!}\sum_{\pi\in S_n}\mathcal{U}_{A,\pi}\otimes\mathcal{U}_{B,\pi}(\cdot), \tag{2.32}$$

where $\mathcal{U}_{A,\pi}$ and $\mathcal{U}_{B,\pi}$ are the channels exchanging the subsystems of $\mathcal{H}_A^{\otimes n}$ respectively $\mathcal{H}_B^{\otimes n}$ according to permutation π. To implement $\overline{\mathcal{U}}$ as a communication protocol, A and B have to agree on a permutation which is chosen randomly from the symmetric group S_n on n letters (see Figure 2.1).

Applying $\mathcal{U}$ as a communication protocol (or as a part of a greater protocol) consequently amounts in consuming shared equidistributed randomness (*common randomness* as it is called usually in the information theory literature) at rate

$$R_n = \frac{1}{n}\log n! \tag{2.33}$$

[1] In this section, we restrict ourselves to discussion of the consequences of the derived bounds for quantum states. Similar observation regarding quantum channels easily follow from our bounds regarding quantum channels.

bits per block length. The observation, that the rates R_n grow unbounded in the asymptotic limit $n \to \infty$ disqualifies $\mathcal{U}$ as a protocol in situations, where shared randomness is not a free resource, but instead does count to the resource trade-off.

In this context, Theorem 7 proven in the preceding section provides an uplifting message. A weighted symmetric design (on $\mathcal{H}^{\otimes n}$) as introduced in Definition 5 exactly simulates the action of $\overline{\mathcal{U}}$. Theorem 7 therefore shows, that we always can equivalently replace $\overline{\mathcal{U}}$ by a protocol which demands (not necessarily equidistributed randomness) at a rate

$$R'_n \leqslant 4 \cdot \log \dim(\mathcal{H}_A \otimes \mathcal{H}_B). \tag{2.34}$$

We have shown, that the brute-force evenly distributed random selection out of all permutations can be replaced by random selection from a much smaller set of permutations (which amounts to rate-bounded coordinated randomness demands.)

Opposite to the consequences discussed so far, our results also enforce some conclusions of the more disillusioning type. Having established protocols for enforcing permutation-invariance which are reasonable regarding their randomness consumption, they may be too expensive in randomness consumption sometimes.

As a consequence of the well-known Holevo bound, we obtain the inequality

$$I(X_{A^n}; Y_{B^n}) \;\leqslant\; n \cdot \log \dim \mathcal{H}_A \otimes \mathcal{H}_B \tag{2.35}$$

which provides a principal bound for the mutual information of a bipartite random variable (X_{A^n}, Y_{B^n}) produced by local measurements on the A and B subsystems of any bipartite quantum system with underlying Hilbert space $(\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes n}$. When regarding resource trade-offs, comparing the bounds in (2.35) and the one given by

$$H(p) \;\geqslant\; n \log \dim \mathcal{H}_A \otimes \mathcal{H}_B \tag{2.36}$$

for Shannon entropy of any probability distribution producing a symmetric design given by

$$\overline{\mathcal{U}}_q \;:=\; \sum_{\pi \in S_n} q(\pi) \cdot \mathcal{U}_{A,\pi} \otimes \mathcal{U}_{B,\pi} \tag{2.37}$$

we notice that permutation-symmetrization, costs at least as much shared randomness as could be produced at all (in a perfect situation) by local measurements on a system.

While the preceding observation may have no consequences in communication situations where shared randomness is a cheap resource, there are other situations, where the communication demands are critical to an extent, that the introduced protocol class is disqualified.

A special instance of such a situation is faced, when in addition to A and B (which we call henceforth *legitimate users*) a third, malicious party E takes part in the communication.

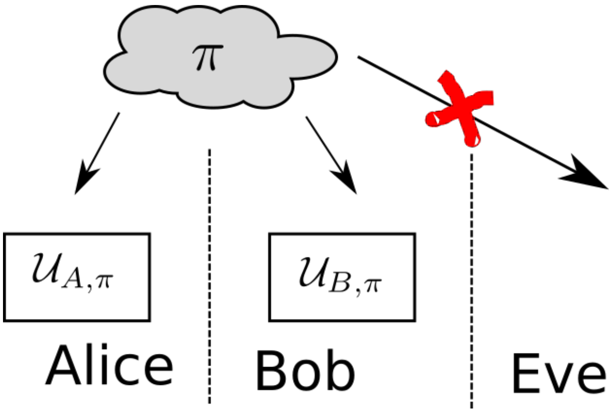

Figure 2.2.: Choice of random permutation π for implementation of $\overline{\mathcal{U}}$ has to be coordinated between legitimate parties A and B but protected from knowledge by adversarial party E.

Let the underlying space of the system be $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$. In this case, it is usually not enough to perform the random choice in a way that it is coordinated between the legitimate parties A and B. Moreover, it has to be secure in the sense, that the malicious party E has no knowledge of the permutation π chosen (see Figure 2.2.)

An example where the correlation shared not just by the legitimate but also the adversarial communication parties is useless, is given in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel]. Therein it is proven, that the secrecy capacity of an arbitrarily varying wiretap classical quantum channel (AVWQC) under assistance of common randomness which is secure against the jamming adversarial party sometimes strictly exceeds the corresponding capacity of the AVWC under assistance of public (non-private) common randomness. Additionally, in the case where the randomness is also accessible to the jamming adversary, the corresponding capacity equals the capacity without any common randomness assistance. Common randomness is useless for secret message transmission if it is also known to the (active) adversary.

3. Simultaneous transmission of classical and quantum information under channel uncertainty

In this chapter we derive universal codes for simultaneous transmission of classical messages and entanglement through quantum channels, possibly under attack of a malignant third party. These codes are robust to different kinds of channel uncertainty. To construct such universal codes, we invoke and generalize properties of random codes for classical and quantum message transmission through quantum channels. We show these codes to be optimal by giving a multi-letter characterization of regions corresponding to capacity of compound quantum channels for simultaneously transmitting and generating entanglement with classical messages. Also, we give dichotomy statements in which we characterize the capacity of arbitrarily varying quantum channels for simultaneous transmission of classical messages and entanglement. These include cases where the malignant jammer present in the arbitrarily varying channel model is classical (chooses channel states of product form) and fully quantum (is capable of general attacks not necessarily of product form).

3.1. Introduction

Simultaneous transmission of classical messages and entanglement is a nontrivial problem even if capacity achieving codes for the corresponding univariate transmission goals are at hand. It was already observed in [Devetak and Shor(2005)] for perfectly known quantum channels that the naive time sharing strategy is generally insufficient to achieve the full capacity region. Examples of channels where coding beyond time-sharing is indispensable does not depend on constructing pathologies. They are readily found even within the standard arsenal of qubit quantum channels, e.g. the dephasing qubit channels [Devetak and Shor(2005)].

We derive codes for simultaneous transmission of classical messages and entanglement that are robust to the three types of uncertainty mentioned above. The codes used here for the compound model, are different from those used for the point to point communication in [Devetak and Shor(2005)] when considering the special case of $|S| = 1$. Given that the input state approximation techniques used therein prove insufficient in presence of channel state uncertainty, in this thesis we use the decoupling approach first established in [Klesse(2007)]. We combine robust random codes for classical message transmission from [Mosonyi(2015)] and a generalization of (decoupling based) entanglement transmission codes from [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] to construct appropriate simultaneous codes for compound quantum channels under the maximal error criterion. We show that these codes are optimal by giving a multi-letter characterization of the capacity of compound quantum channels with no assumption on, the size of the underlying uncertainty set. We use the asymptotic equivalence of the two tasks of entanglement transmission and entanglement generation to include the capacity region corresponding to simultaneous transmission of classical messages and generation of entanglement between the two parties.

Next, we convert the codes derived for the compound channel, using Ahlswede's robustification and elimination techniques ([Ahlswede(1978)]) to derive suitable codes for arbitrarily varying quantum channels. This is possible given that the error functions associated with codes corresponding to the compound model decay to zero exponentially. We derive a dichotomy statement ([Ahlswede(1978)]), for the simultaneous classical message and entanglement transmission through AVQCs under the average error criterion. This dichotomy is observed when considering two scenarios where the communicating parties do and do not have access to unlimited common randomness, yielding the common-randomness and deterministic capacity regions of the channel model respectively. Therefore, we show that firstly, the common-randomness capacity region of the arbitrarily varying channel is equal to that of the compound channel $\mathrm{conv}(\mathcal{J})$, namely the compound channel generated by the convex hull of the uncertainty set of channels $\mathcal{J}$. Secondly, if the deterministic capacity of the arbitrarily varying channel is not the point $(0, 0)$, it is equal to the common-randomness capacity of the channel.

We give a necessary and sufficient condition for the deterministic capacity region to be be the point $(0, 0)$. This condition is known as symmetrizablity of the channel (see [Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel] and [Boche and Nötzel(2014)]). Finally, we show that the codes derived here, can be used for fully quantum AVCs where the jammer is not restricted to product states, but can use general quantum states to parametrize the channel used many times. This model has been introduced in Section 3.6 along with the main result and related work for fully quantum AVCs and hence here, we avoid further explanation of the techniques used there.

The task of simultaneous transmission of classical messages and entanglement was first considered by Devetak and Shor in [Devetak and Shor(2005)] in case of a memoryless quantum channel under assumption that the channels state is perfectly known to its users. The authors derived a multi-letter characterization of the capacity region in this setting which also classified the naïve time-sharing approach as being suboptimal for simultaneous transmission. A code construction sufficient to achieve also the rate pairs lying outside the time-sharing region was derived using a "piggy-backing" technique. A specialized construction introduced in [Devetak(2005)] allows to encode the identity of the classical message into the coding states of an underlying entanglement transmission code. The mentioned strategy to optimally combine different communication tasks in quantum channel coding was afterwards used and further developed in different directions. We explicitly mention subsequent research activity by Hsieh and Wilde [Hsieh and Wilde(2010a), Hsieh and Wilde(2010b)] where the idea of "piggy backing" classical messages onto quantum codes was extended to include entanglement assistance. The resulting code construction being sufficient to achieve each point in the three-dimensional rate region for entanglement-assisted classical/quantum simultaneous transmission leads to a full (multi-letter) characterization of the "Quantum dynamic capacity" of a (perfectly known) quantum channel [Wilde and Hsieh(2011a)] (see the textbook [Wilde(2017)] for an up-to-date pedagocial presentation of the mentioned results).

In order to derive classically enhanced quantum codes being robust against channel uncertainty, we refine the construction entanglement transmission codes for compound quantum channels from [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel,Boche et al.(2018b) Boche, Deppe, Nötzel, and Winter] instead of elaborating on the usual approach building up on codes from [Devetak(2005)]. In fact, it was noticed earlier that deriving entanglement generation codes from secure classical message transmission codes (the strategy which the arguments in [Devetak(2005)] follow) seems to be not suitable when the channel is a compound quantum channel.

In the first section following this introduction, we introduce the notation used in this work. Precise definitions of the channel models, codes used in different scenarios along with capacity regions and finally the main results in form of Theorem 25 and Theorem 32, are given in Section 3.2. In Section 3.3, we present preliminary coding results for entanglement transmission (Section 3.3.1) and classical message transmission (Section 3.3.2). The entanglement transmission codes introduced in this section are a generalization of the random codes in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] and [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] to accommodate conditional typicality of the input on words from many copies of an alphabet. The classical message transmission codes are those from [Mosonyi(2015)] that prove sufficient for our simultaneous coding purposes.

Equipped with these results, we move on to Section 3.4, to prove the coding results for the compound channel model. In this section, after proving a converse for the capacity region in Theorem 25, we prove the direct part in two steps. In the first step, we show that capacity regions that correspond to the case where the sender is restricted to inputting maximally entangled pure states are achieved. In the second step, we prove achievablity of capacity regions corresponding to general inputs, using elementary methods that are less involved that the usual BSST type results used for this generalization in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] and [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter].

In Section 3.5, after proving a converse for the capacity region under the arbitrarily varying channel model, we prove coding results in this model by converting the compound channel model codes using Ahlswede's robustification method. This, assumes unlimited common randomness available to the legal parties. We then use an instance of elimination to show that if the deterministic capacity region is not the point $(0, 0)$, negligible amount of common randomness per use of the channel is sufficient to achieve the same capacity region. Also in this section, we prove necessity and sufficiency of symmetrizablity condition for the case where the deterministic capacity region is the point $(0, 0)$. Finally, in Section 3.6, m these results to the case of quantum jammer by proving Theorem 65.

3.2. Basic definitions and main results

We consider two channel models of compound and arbitrarily varying quantum channels. They are both generated by an uncertainty set of CPTP maps. For the purposes of the present work, when considering the arbitrarily varying channel model, we assume finiteness of the generating uncertainty set. This assumption is absent in the case of the compound channel model.

3.2.1. The compound quantum channel

Here, we consider quantum compound channels. Let $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ be a set of CPTP maps. The compound quantum channel generated by $\mathcal{J}$ is given by family $\{\mathcal{N}^{\otimes n} : \mathcal{N} \in \mathcal{J}\}_{n=1}^{\infty}$. In other words, using n instances of the compound channel is equivalent to using n instances of one of the channels from the uncertainty set. The users of this channel may or may not have access to the Channel State Information (CSI). We will often use the set S to index members of $\mathcal{J}$. A compound channel is used $n \in \mathbb{N}$ times by the sender Alice, to convey classical messages from a set $[M_{1,n}] := \{1, ..., M_{1,n}\}$ to a receiver Bob. At the same time, the parties would like to communicate quantum information. Here, we consider two scenarios in which quantum information can be communicated between the parties.

Classically Enhanced Entanglement Transmission (CET): While transmitting classical messages using $n \in \mathbb{N}$ instances of the compound channel, the sender wishes to transmit the maximally entangled state in her control to the receiver. The subspace $\mathcal{F}_{A,n}$ with $\mathcal{F}_{A,n} \subset \mathcal{H}_A^{\otimes n}$ and $M_{2,n} := \dim(\mathcal{F}_{A,n})$, quantifies the amount of quantum information transmitted. More precisely:

Definition 21 *An $(n, M_{1,n}, M_{2,n})$ CET code for $\mathcal{J} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$, is a family $\mathcal{C}_{CET} := (\mathcal{P}_m, \mathcal{R}_m)_{m\in[M_{1,n}]}$ with*

- $\mathcal{P}_m \in \mathcal{C}(\mathcal{F}_{A,n}, \mathcal{H}_A^{\otimes n})$,

- $\mathcal{R}_m \in \mathcal{C}^{\downarrow}(\mathcal{H}_A^{\otimes n}, \mathcal{F}_{B,n})$ *with* $\mathcal{F}_{A,n} \subset \mathcal{F}_{B,n}$ *and*
- $\sum_{m \in [M_{1,n}]} \mathcal{R}_m \in \mathcal{C}(\mathcal{H}_B^{\otimes n}, \mathcal{F}_{B,n})$.

Remark 22 *We remark that as defined above, for each* $m \in [M_{1,n}]$ *we have a* $(n, M_{2,n})$ *entanglement transmission code for* $\mathcal{J}$.

For every $m \in M_{1,n}$ and $s \in S$, we define the following performance function for this communication scenario when $n \in \mathbb{N}$ instances of the channel have been used,

$$P(\mathcal{C}_{CET}, \mathcal{N}_s^{\otimes n}, m) := F(|m\rangle\langle m| \otimes \Phi^{AB}, \mathrm{id}_{\mathcal{F}_{A,n}} \otimes \mathcal{R} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_m(\Phi^{AA})),$$

where Φ^{XY} is a maximally entangled state on $\mathcal{F}_{X,n} \otimes \mathcal{F}_{Y,n}$ and

$$\mathcal{R} := \sum_{m \in [M_{1,n}]} |m\rangle\langle m| \otimes \mathcal{R}_m.$$

Classically Enhanced Entanglement Generation (CEG): In this scenario, while transmitting classical messages, Alice wishes to establish a pure state shared between her and Bob. As the maximally entangled pure state shared between the parties is an instance of such a pure state, it can be proven that the previous task achieved in CET, achieves the task laid out by this one, but the opposite is not necessarily true. More precisely:

Definition 23 *An* $(n, M_{1,n}, M_{2,n})$ *CEG code for* $\mathcal{J} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$, *is a family* $\mathcal{C}_{CEG} := (\Psi_m, \mathcal{R}_m)_{m=1}^{M_{1,n}}$, *where* Ψ_m *is a pure state on* $\mathcal{F}_{A,n} \otimes \mathcal{H}_A^{\otimes n}$ *and*

- $\mathcal{R}_m \in \mathcal{C}^{\downarrow}(\mathcal{H}_B^{\otimes n}, \mathcal{F}_{B,n})$ *with* $\mathcal{F}_{A,n} \subset \mathcal{F}_{B,n}$ *and*
- $\sum_{m \in [M_{1,n}]} \mathcal{R}_m \in \mathcal{C}(\mathcal{H}_B^{\otimes n}, \mathcal{F}_{B,n})$.

The relevant performance functions for this task, for every $m \in [M_{1,n}]$ and $s \in S$, are

$$P(\mathcal{C}_{CEG}, \mathcal{N}_s^{\otimes n}, m) := F(|m\rangle\langle m| \otimes \Phi, \mathrm{id}_{\mathcal{F}_{A,n}} \otimes \mathcal{R} \circ \mathcal{N}_s^{\otimes n}(\Psi_m)), \tag{3.1}$$

with Φ maximally entangled on $\mathcal{F}_{A,n} \otimes \mathcal{F}_{B,n}$.
Averaging over the message set $[M_{1,n}]$, will give us the corresponding average performance functions for each $s \in S$,

$$\overline{P}(\mathcal{C}_X, \mathcal{N}_s^{\otimes n}) := \frac{1}{M_{1,n}} \sum_{m \in [M_{1,n}]} P(\mathcal{C}_X, \mathcal{N}_s^{\otimes n}, m),$$

for $X \in \{CET, CEG\}$. For each scenario, we define the achievable rates.

Definition 24 *Let* $X \in \{CET, CEG\}$. *A pair* (R_1, R_2) *of non-negative numbers is called an achievable X rate for the compound channel* $\mathcal{J}$, *if for each* $\epsilon, \delta > 0$ *exists a number* $n_0 = n_0(\epsilon, \delta)$, *such that for each* $n > n_0$ *we find and* $(n, M_{1,n}, M_{2,n})$ *X code* $\mathcal{C}_X$ *such that*

1. $\frac{1}{n} \log M_{i,n} \geqslant R_i - \delta$ *for* $i \in \{1, 2\}$,

2. $\inf_{s\in S}\min_{m\in M_{1,n}} P(\mathcal{C}_X, \mathcal{N}_s^{\otimes n}, m) \geqslant 1-\epsilon$

are simultaneously fulfilled. We also define X "average-error-rates" by averaging the performance functions in the last condition over $m \in [M_{1,n}]$. *We define the X capacity region of* $\mathcal{J}$ *by*

$$C_X(\mathcal{J}) := \{(R_1, R_2) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : (R_1, R_2) \textit{ is achievable } X \textit{ rate for } \mathcal{J}\}. \tag{3.2}$$

Also the capacity region corresponding to average error criteria is defined as

$$\overline{C}_X(\mathcal{J}) := \{(R_1, R_2) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : (R_1, R_2) \textit{ is achievable } X \textit{ average-error-rate for } \mathcal{J}\}. \tag{3.3}$$

Moreover, let $\mathcal{X}$ be an alphabet, $\mathcal{M} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ $\forall s \in S$, $p \in \mathcal{P}(\mathcal{X})$ and Ψ_x be a pure state for all $x \in \mathcal{X}$. Given the state

$$\omega(\mathcal{M}, p, \Psi) := \sum_{x\in\mathcal{X}} p(x)\, |x\rangle\langle x| \otimes \mathrm{id}_{\mathcal{H}_A} \otimes \mathcal{M}(\Psi_x), \tag{3.4}$$

we introduce the following set,

$$\hat{C}(\mathcal{N}_s, p, \Psi) := \{(R_1, R_2) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : R_1 \leqslant I(X;B, \omega(\mathcal{N}_s, p, \Psi)) \wedge R_2 \leqslant I(A\rangle BX, \omega(\mathcal{N}_s, p, \Psi))\}$$

with Ψ denoting $(\Psi_x : x \in \mathcal{X})$ collectively. We will also use

$$\frac{1}{l}A := \{(\frac{1}{l}x_1, \frac{1}{l}x_2) : (x_1, x_2) \in A\}.$$

The following statement is the first main result of this chapter.

Theorem 25 *Let* $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ *be any compound quantum channel. Then*

$$C_{CET}(\mathcal{J}) = \overline{C}_{CET}(\mathcal{J}) = C_{CEG}(\mathcal{J}) = \overline{C}_{CEG}(\mathcal{J}) = \mathrm{cl}\Bigg(\bigcup_{l=1}^{\infty} \frac{1}{l} \bigcup_{p,\Psi} \bigcap_{s\in S} \hat{C}(\mathcal{N}_s^{\otimes l}, p, \Psi)\Bigg)$$

holds.

This theorem is proven in the following steps. In Section 3.4.1, we prove that $\overline{C}_{CEG}(\mathcal{J})$ is a subset of the set on the rightmost set in the above equalities. In Section 3.4.2, we prove that the rightmost set is a subset of $C_{CET}(\mathcal{J})$. Together with the operational inclusions

$$C_{CET}(\mathcal{J}) \subset C_{CEG}(\mathcal{J})$$

and

$$C_X(\mathcal{J}) \subset \overline{C}_X(\mathcal{J})$$

for $X \in \{CEG, CET\}$, we conclude the equalities in the statement of the theorem.

3.2.2. The arbitrarily varying quantum channel

The arbitrarily varying quantum channel generated by a set $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S}$ of CPTP maps with input Hilbert space $\mathcal{H}_A$ and output Hilbert space $\mathcal{H}_B$, is given by family of CPTP maps $\{\mathcal{N}_{s^l} : \mathcal{L}(\mathcal{H}_A^{\otimes l}) \to \mathcal{L}(\mathcal{H}_B^{\otimes l}), s^l \in S^l, l \in \mathbb{N}\}_{l=1}^{\infty}$, where

$$\mathcal{N}_{s^l} := \mathcal{N}_{s_1} \otimes \dots \mathcal{N}_{s_l} \quad (s^l \in S^l).$$

We use $\mathcal{J}$ to denote the AVQC generated by $\mathcal{J}$. To avoid further technicalities, we always assume $|S| < \infty$ for the AVQC generating sets appearing in this chapter. Most of the results in this chapter may be generalized to the case of general sets by clever use of approximation techniques from convex analysis together with continuity properties of the entropic quantities which appear in the capacity characterizations (see [Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel]).

Definition 26 *An $(l, M_{1,l}, M_{2,l})$ random CET code for $\mathcal{J}$ is a probability measure μ_l on $(\mathcal{C}(\mathcal{F}_{A,l}, \mathcal{H}_A^{\otimes l})^{M_{1,l}} \times \Omega_l, \sigma_l)$, where*

- $\Omega_l := \{(\mathcal{R}^{(1)}, \dots, \mathcal{R}^{(M_{1,l})}), \sum_{m\in[M_{1,l}]} \mathcal{R}^{(m)} \in \mathcal{C}(\mathcal{H}_B^{\otimes l}, \mathcal{F}_{B,l})\}$,
- $\dim(\mathcal{F}_{A,l}) = M_{2,l}, \mathcal{F}_{X,l} \subset \mathcal{H}_X^{\otimes l}, \quad (X \in \{A, B\})$.
- *The sigma-algebra σ_l is chosen such that the function*

$$g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)}) := F(|m\rangle\langle m| \otimes \Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{R} \circ \mathcal{N}_{s^l} \circ \mathcal{P}^{(m)}(\Phi^{AA})) \tag{3.5}$$

  is measurable with respect to μ_l, for all $m \in [M_{1,l}], s^l \in S^l$. In (3.5), Φ^{XY} is a maximally entangled state on $\mathcal{F}_{X,l} \otimes \mathcal{F}_{Y,l}$ and $\mathcal{R} := \sum_{m\in[M_{1,l}]} |m\rangle\langle m| \otimes \mathcal{R}^{(m)}$.

- *We further require that σ_l contains all the singleton sets. The case where μ_l is deterministic, namely is equal to unity on a singleton set and zero otherwise, gives us a deterministic $(l, M_{1,l}, M_{2,l})$ CET codes for $\mathcal{J}$. Abusing the terminology, we also refer to the singleton sets as deterministic codes.*

Definition 27 *A non-negative pair of real numbers (R_1, R_2) is called an achievable CET rate pair for $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S}$ with random codes and average error criterion, if there exists a random CET code μ_l for $\mathcal{J}$ with members of singleton sets notified by $(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]}$ such that*

1. $\liminf_{l\to\infty} \frac{1}{l} \log M_{i,l} \geqslant R_i \ (i \in \{1, 2\})$,
2. $\lim_{l\to\infty} \inf_{s^l\in S^l} \int \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)}) \ d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m=1}^{M_{1,l}} = 1$.

The random CET capacity region with average error criterion of $\mathcal{J}$ is defined by

$$\overline{\mathcal{A}}_{r,CET}(\mathcal{J}) := \{(R_1, R_2) : (R_1, R_2) \ is \ achievable \ CET \ rate \ pair \ for \ \mathcal{J}$$
$$with \ random \ codes \ and \ average \ error \ criterion\}.$$

Definition 28 *A non-negative pair of real numbers (R_1, R_2) is called an achievable deterministic CET rate for $\mathcal{J}$ with average error criterion, if there exists a deterministic $(l, M_{1,l}, M_{2,l})$ CET code $(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]}$ for $\mathcal{J}$ with*

1. $\liminf_{l\to\infty} \frac{1}{l}\log M_{i,l} \geqslant R_i$ $(i \in \{1,2\})$,
2. $\lim_{l\to\infty} \inf_{s^l\in S^l} \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)}) = 1$

Correspondingly we define the following capacity region,

$$\overline{\mathcal{A}}_{d,CET}(\mathcal{J}) := \{(R_1,R_2) : (R_1,R_2) \text{ is achievable deterministic} \\ \text{CET rate pair for } \mathcal{J} \text{ with average error criterion}\}.$$

The deterministic CET codes defined here, are entanglement transmission codes for each $m \in [M_{1,l}]$. More precisely we have the following definition.

Definition 29 *An (n, M), $n, M \in \mathbb{N}$, entanglement transmission code for AVQC $\mathcal{J} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ is a pair $(\mathcal{P}, \mathcal{R})$ with $\mathcal{P} \in \mathcal{C}(\mathcal{F}_{A,n}, \mathcal{H}_A^{\otimes n}), \mathcal{R} \in \mathcal{C}(\mathcal{H}_B^{\otimes n}, \mathcal{F}_{B,n})$ with $\mathcal{F}_{A,n} \subset \mathcal{F}_{B,n} \subset \mathcal{H}_A^{\otimes n}$ and $\dim(\mathcal{F}_{A,n}) = M$. The corresponding performance function for this task is*

$$F(\Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes n}} \otimes \mathcal{R} \circ \mathcal{N}_{s^n} \circ \mathcal{P}(\Phi^{AA})), \quad s^n \in S^n.$$

Essential to the statement of our results is the concept of symmetrizablity defined in the following.

Definition 30 *Let $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ with $|S| < \infty$ be an AVQC.*

1. *$\mathcal{J}$ is called l-symmetrizable for $l \in \mathbb{N}$, if for each finite set $\{\rho_1, \dots, \rho_K\} \subset \mathcal{S}(\mathcal{H}_A^{\otimes l})$ with $K \in \mathbb{N}$, there is a map $p : \{\rho_1, \dots, \rho_K\} \to \mathcal{P}(S^l)$ such that for all $i, j \in \{1, \dots, K\}$*
$$\sum_{s^l\in S^l} p(\rho_i)(s^l)\mathcal{N}_{s^l}(\rho_j) = \sum_{s^l\in S^l} p(\rho_j)(s^l)\mathcal{N}_{s^l}(\rho_i). \tag{3.6}$$
2. *We call $\mathcal{J}$ symmetrizable if it is l-symmetrizable for all $l \in \mathbb{N}$.*

Remark 31 *The above definition for symmetrizablity was first established in [Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel], generalizing the concept of symmetrization for classical AVQCs from [Ericson(1985)]. This definition for symmetrizablity was meaningfully simplified in [Boche and Nötzel(2014)], to require checking of the condition (3.6) for two input states only (K=2).*

We prove the following result to be the second main result of this chapter.

Theorem 32 *Let $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ with $|S| < \infty$ be an AVQC. The following hold.*

1. *$\overline{\mathcal{A}}_{d,CET}(\mathcal{J}) \neq \{(0,0)\}$ implies*

$$\overline{\mathcal{A}}_{d,CET}(\mathcal{J}) = \overline{\mathcal{A}}_{r,CET}(\mathcal{J}) = \overline{C}_{CET}(\text{conv}(\mathcal{J})), \tag{3.7}$$

where $\overline{C}_{CET}(\mathcal{M})$ is the CET capacity of compound channel $\mathcal{M}$ with average error criterion defined in the previous section and

$$\text{conv}(\mathcal{J}) := \{\mathcal{N}_q : \mathcal{N}_q := \sum_{s \in S} q(s)\mathcal{N}_s, q \in \mathcal{P}(S)\}.$$

2. *$\overline{\mathcal{A}}_{d,CET}(\mathcal{J}) = \{(0,0)\}$ if and only if $\mathcal{J}$ is symmetrizable.*

3.3. Universal random codes for quantum channels

In this section we prove universal random coding results for entanglement transmission and classical message transmission over quantum channels. Most of the statements below, are implicitly contained in the literature. We state some properties of these codes that stem from their random nature and prove useful when deriving CET codes stated in Section 3.4.

Before proceeding with the following two sections in which we introduce appropriate entanglement transmission and classical message transmission coding results and for the reader's convenience, we present briefly the concept of types used in the remainder of this section. For more information on the concept of types, see e.g. [Wilde(2017)].

For $l \in \mathbb{N}$, the word $x^l \in \mathcal{X}^l$ that is a string of letters $x \in \mathcal{X}$ and the state ρ with spectral decomposition $\rho := \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x|$, we define the δ-typical (frequency typical) projection

$$q_{\delta,l}(\rho) := \sum_{x^l \in T^l_{p,\delta}} |x^l\rangle\langle x^l| ,$$

where $T^l_{p,\delta}$ is the set of δ-typical sequences in $\mathcal{X}^l$, defined by

$$T^l_{p,\delta} := \{x^l : \forall x \in \mathcal{X}, |\frac{1}{l}N(x|x^l) - p(x)| \leqslant \delta \ \wedge \ p(x) = 0 \iff N(x|x^l) = 0\} \tag{3.8}$$

where $N(x|x^l)$ is the number of occurrences of letter x in word x^l.

For each $l \in \mathbb{N}$, we consider the set of types over alphabet $\mathcal{X}$, $\mathcal{T}(\mathcal{X}, l)$ defined as

$$\mathcal{T}(\mathcal{X}, l) := \{\lambda : T^l_\lambda \neq \varnothing\},$$

where $T^l_\lambda = T^l_{\lambda,0}$ $(\delta = 0)$.

3.3.1. Entanglement transmission codes

In this section, we prove universal entanglement transmission coding results that are to be combined with suitable classical message transmission codes introduced in the next section. The following lemma is a generalization of random entanglement transmission codes obtained in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] and [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter], where a in turn generalization of the decoupling lemma from [Klesse(2007)] has been obtained. As stated in the following lemma, there are two points to be remarked about these codes. First, the random nature of these codes gives us an encoding state (outcome of the random encoding operation) with a tensor product structure, that is of interest for the present work. Therefore at this stage, we skip the de-randomization step that seemed natural in the original work. Secondly, the integration over unitary groups with respect to the normalized Haar measure done in the random encoding operation therein, is replaced here by an average over the elements of discrete and finite subsets of representations of the unitary group known as unitary designs (see e.g. [Gross et al.(2007)Gross, Audenaert, and Eisert]).
The product structure of the encoding state can be used for an instance of channel coding stated later on. This becomes clear when the tensor product structure of the average state is used to accommodate typicality. For $p \in \mathcal{P}(\mathcal{X})$ where $\mathcal{X}$ is some finite alphabet, $\delta > 0$ and $x^l \in \mathcal{X}^l$, we introduce the following notation. For the tuple $x^l := (x_1, \dots, x_l)$ where $x_i \in \mathcal{X}$ for $i = 1, \dots, l$, we define

$$\mathcal{G}_{x^l} := \mathcal{G}_{x_1} \otimes \dots \otimes \mathcal{G}_{x_l},$$

where $\mathcal{G}_{x_i} \subset \mathcal{H}_A$ and clearly, $\mathcal{G}_{x^l} \subset \mathcal{H}_A^{\otimes l}$. Then $\pi_{x^l} := \pi_{\mathcal{G}_{x^l}}$ denotes the maximally mixed state on $\mathcal{G}_{x^l}$ (correspondingly π_x denotes the maximally mixed state on $\mathcal{G}_x$ for $x \in \mathcal{X}$), $\varPhi_{x^l}$ a purification of $\pi_{\mathcal{G}_{x^l}}$ (correspondingly $\varPhi_x$ denotes a purification of π_x) and X_{x^l} is a unitary design (see Theorem 38) for $\mathcal{U}(\mathcal{G}_{x^l})$. The following lemma reduces to Theorem 5 of [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] when $|\mathcal{X}| = 1$.

Lemma 33 *Let $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ be any compound quantum channel and alphabet $\mathcal{X}$ be given. For subspaces $(\mathcal{G}_x)_{x\in\mathcal{X}}$ with $\mathcal{G}_x \subset \mathcal{H}_A, x \in \mathcal{X}$, probability distribution $p \in \mathcal{P}(\mathcal{X})$ and $\delta > 0$, there exists $l_0 \in \mathbb{N}$, such that for all $l \geqslant l_0$, we find for each $x^l \in T^l_{p,\delta}$, a subspace $\mathcal{F}_{A,l} \subset \mathcal{G}_{x^l}$ and a family $(\mathcal{P}_i, \mathcal{R}_i)_{i=1}^{|X_{x^l}|}$ of $(l, \dim(\mathcal{F}_{A,l}))$ entanglement transmission codes with $|X_{x^l}| < \infty$ and*

1. $\frac{1}{l}\log\dim(\mathcal{F}_{A,l}) \geqslant \inf_{s\in S} I(A\rangle BX, \omega(\mathcal{N}_s, p, \varPhi)) - \delta$ *, with $\omega(\mathcal{N}_s, p, \varPhi)$ defined in (3.4) for $\varPhi := (\varPhi_x : x \in \mathcal{X})$,*

2. $\forall s \in S \quad \frac{1}{|X_{x^l}|}\sum_{i=1}^{|X_{x^l}|} F_e(\pi_{\mathcal{F}_{A,l}}, \mathcal{R}_i \circ \mathcal{N}_s^{\otimes l} \circ \mathcal{P}_i) \geqslant 1 - \epsilon_l$ *with $\epsilon_l \to 0$ exponentially as $l \to \infty$,*

3. $\frac{1}{|X_{x^l}|}\sum_{i=1}^{|X_{x^l}|} \mathcal{P}_i(\pi_{\mathcal{F}_{A,l}}) = \pi_{x^l}.$

The ingredients to prove this lemma are presented here in form of two lemmas prior to the main proof. The following two lemmas reduce to Lemma 5 and 6 from [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel][1] when $|\mathcal{X}| = 1$. Following these lemmas, we state Theorem 38 based on which we replace the integration with respect to Haar measure, with an average over a subset of the unitary groups called unitary designs. In short, the entanglement transmission codes in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] were derived given a number $l \in \mathbb{N}$ and subspace $\mathcal{G}^{\otimes l} \subset \mathcal{H}^{\otimes l}$. Here, we derive codes for a subspace $\mathcal{G}_{x^l}$, with a tensor product structure determined by word x^l (see the description above Lemma 33).

Lemma 34 *Let $(\lambda_x)_{x\in\mathcal{A}}$ be a probability distribution with $\lambda_x > 0, \forall x \in \mathcal{A}$ on an alphabet $\mathcal{A}$. For $\rho_{x^l} := \bigotimes_{x\in\mathcal{A}} \rho_x^{\otimes N_x}, N_x := \lambda_x \cdot l \in \mathbb{N}, \rho_x \in \mathcal{S}(\mathcal{H}) \;\; \forall x \in \mathcal{A}$ and $\delta \in (0, 1/2)$, there exist a real number $\tilde{c} > 0$, functions $h : \mathbb{N} \to \mathbb{R}^+$, $\phi : (0, 1/2) \to \mathbb{R}^+$ with $\lim_{l\to\infty} h(l) = 0$ and $\lim_{\delta\to 0} \phi(\delta) = 0$ and an orthogonal projection $q_{\delta,l}$ satisfying*

1. $\operatorname{tr}(\rho_{x^l} q_{\delta,l}) \geqslant 1 - |\mathcal{A}| 2^{-l(\tilde{c}\delta^2 - h(l))}$
2. $q_{\delta,l}\rho_{x^l}q_{\delta,l} \leqslant 2^{-(S(\rho_{x^l}) - l\phi(\delta))} q_{\delta,l}$.

The last inequality implies

$$\| q_{\delta,l}\rho_{x^l}q_{\delta,l} \|_2^2 \leqslant 2^{-(S(\rho_{x^l}) - l\phi(\delta))}.$$

Proof 35 *Let for each $x \in \mathcal{A}$, $q^{(x)}_{\delta,N_x}$ be the frequency typical projection associated with state $\rho_x^{\otimes N_x}$ in terms of Lemma 201. We show that the projection operator $q_{\delta,l} := \bigotimes_{x\in\mathcal{A}} q^{(x)}_{\delta,N_x}$ has the properties listed in the statement above. We have*

$$\begin{aligned}\operatorname{tr}(\rho_{x^l} q_{\delta,l}) &= \operatorname{tr}(\bigotimes_{x\in\mathcal{A}} \rho_x^{\otimes N_x} q^{(x)}_{\delta,N_x}) = \prod_{x\in\mathcal{A}} \operatorname{tr}(\rho_x^{\otimes N_x} q^{(x)}_{\delta,N_x}) \\ &\geqslant \prod_{x\in\mathcal{A}} (1 - 2^{-N_x(\bar{c}\delta^2 - h'(N_x))}) \\ &\geqslant (1 - 2^{-c_0 l(\bar{c}\delta^2 - h'(c_0 l))})^{|\mathcal{A}|} \geqslant 1 - |\mathcal{A}| 2^{-c_0 l(\bar{c}\delta^2 - h'(c_0 l))},\end{aligned}$$

where $c_0 := \min_{x\in\mathcal{A}} \lambda_x$. Setting $\tilde{c} = c_0\bar{c}$ and $h(l) = c_0 h'(c_0 l)$, we have the first claim. To see the second claim, we observe that

$$\begin{aligned}q_{\delta,l}\rho_{x^l}q_{\delta,l} &\leqslant \bigotimes_{x\in\mathcal{A}} q^{(x)}_{\delta,N_x} \rho_x^{\otimes N_x} q^{(x)}_{\delta,N_x} \\ &\leqslant \prod_{x\in\mathcal{A}} 2^{-(S(\rho_x^{\otimes N_x}) - N_x\phi(\delta))} \bigotimes_{x\in\mathcal{A}} q^{(x)}_{\delta,N_x} \\ &= 2^{-(S(\rho_{x^l}) - l(\sum_{x\in\mathcal{A}} \lambda_x \phi(\delta))} q_{\delta,l},\end{aligned}$$

where in the last equality, we have used additivity of von Neumann entropy. We are done.

[1] see Lemmas 201 and 202 for the statements.

Lemma 36 *Let $(\lambda_x)_{x\in\mathcal{A}}$ be a probability distribution with $\lambda_x > 0, \forall x \in \mathcal{A}$ on an alphabet $\mathcal{A}$. For each $\mathcal{N} \in \mathcal{C}(\mathcal{H},\mathcal{K})$, $\delta \in (0,1/2)$, and maximally mixed state $\pi_{x^l} := \bigotimes_{x\in\mathcal{A}} \pi_x^{\otimes N_x}, N_x = \lambda_x \cdot l \in \mathbb{N}$ on some $\mathcal{G}_{x^l} \subset \mathcal{H}^{\otimes l}$, there are functions $\gamma : (0,1/2) \to \mathbb{R}^+$ and $h : \mathbb{N} \to \mathbb{R}^+$ satisfying $\lim_{\delta\to 0} \gamma(\delta) = 0$ and $h(l) \searrow 0$ and an operation $\mathcal{N}_{\delta,l} \in \mathcal{C}^{\downarrow}(\mathcal{H}^{\otimes l},\mathcal{K}^{\otimes l})$, called the reduced operation with respect to $\mathcal{N}$ and π_{x^l}, such that*

1. $\operatorname{tr}(\mathcal{N}_{\delta,l}(\pi_{x^l})) \geqslant 1 - |\mathcal{A}|2^{-l(\hat{c}\delta^2 - h(l))}$, *with constant $\hat{c} > 0$.*
2. *$\mathcal{N}_{\delta,l}$ has a Kraus representation with at most $n_{\delta,l} \leqslant 2^{S_e(\pi_{x^l},\mathcal{N}^{\otimes l}) + l(\gamma(\delta)+\check{c}h(l))}$ Kraus operators with constant $\check{c} > 0$.*
3. *For every state $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$ and every two channels $\mathcal{M} \in \mathcal{C}^{\downarrow}(\mathcal{H}^{\otimes l},\mathcal{H}^{\otimes l})$ and $\mathcal{L} \in \mathcal{C}^{\downarrow}(\mathcal{K}^{\otimes l},\mathcal{H}^{\otimes l})$, the inequality*
$$F_e(\rho, \mathcal{L} \circ \mathcal{N}_{\delta,l} \circ \mathcal{M}) \leqslant F_e(\rho, \mathcal{L} \circ \mathcal{N}^{\otimes l} \circ \mathcal{M})$$
is fulfilled.
4. *As the set of Kraus operators of $\mathcal{N}_{\delta,l}$ is a subset of the set of Kraus operators of $\mathcal{N}^{\otimes l}$ for each $l \in \mathbb{N}$, we have*
$$\mathcal{N}_{\delta,l}(\sigma) \leqslant \mathcal{N}^{\otimes l}(\sigma) \quad \forall \sigma \in \mathcal{S}(\mathcal{H}^{\otimes l}).$$

Proof 37 *Let for $x \in \mathcal{A}$, $\mathcal{N}^{(x)}_{\delta,N_x}$ be the reduced operation for $\pi_x^{\otimes N_x}$ in terms of Lemma 202. We show that $\mathcal{N}_{\delta,l} = \bigotimes_{x\in\mathcal{A}} \mathcal{N}^{(x)}_{\delta,N_x}$ has the properties mentioned above. We have*

$$\begin{aligned} tr(\mathcal{N}_{\delta,l}(\pi_{x^l})) &= \prod_{x\in\mathcal{A}} tr(\mathcal{N}_{\delta,N_x}(\pi_x^{\otimes N_x})) \geqslant \prod_{x\in\mathcal{A}} (1 - 2^{-N_x(c'\delta^2 - h'(N_x))}) \\ &\geqslant (1 - 2^{-c_0 l(c'\delta^2 - h'(c_0 l))})^{|\mathcal{A}|} \geqslant 1 - |\mathcal{A}|2^{-c_0 l(c'\delta^2 - h'(c_0 l))}, \end{aligned}$$

where $c_0 := \min_{x\in\mathcal{A}} \lambda_x$. Setting $h(l) = c_0 h'(c_0 l)$ and $\hat{c} = c_0 c'$ we conclude the first claim. Also the following holds for $n_{\delta,l}$, the number of Kraus operators of $\mathcal{N}_{\delta,l}$.

$$\begin{aligned} n_{\delta,l} = \bigotimes_{x\in\mathcal{A}} n_{\delta,N_x} &\leqslant \prod_{x\in\mathcal{A}} 2^{(S_e(\pi_x^{\otimes N_x},\mathcal{N}^{\otimes N_x}) + N_x\gamma(\delta) + N_x h'(N_x))} \\ &\leqslant 2^{(S_e(\pi_{x^l},\mathcal{N}^{\otimes l}) + l(\sum_{x\in\mathcal{A}} \lambda_x \gamma(\delta) + \frac{\lambda_x}{c_0} h(l))} \\ &= 2^{(S_e(\pi_{x^l},\mathcal{N}^{\otimes l}) + l(\gamma(\delta) + \frac{1}{c_0} h(l))}, \end{aligned}$$

where in the second line we have used additivity of the entropy exchange S_e. Finally, the last property comes from multiplicativity of the trace and entanglement fidelity function with respect to tensor products of its arguments.

We now have generalized statements of Lemmas 5 and 6 from [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel]. In the statement of Lemma 33, we have used unitary designs to mimic the average over the unitary group with respect to Haar measure. The following theorem contains a definition of unitary designs.

Theorem 38 *(See e.g. [Gross et al.(2007)Gross, Audenaert, and Eisert]) Let $\mathcal{G}$ be a Hilbert space. For unitaries $U \in \mathcal{U}(\mathcal{G})$, there exists a finite set $X \subset \mathcal{U}(\mathcal{G})$ with $|X| \leqslant \dim(\mathcal{G})^4$ such that*

$$\int_{U \in \mathcal{U}(\mathcal{G})} (U \otimes U)(\cdot)(U \otimes U)^\dagger dU = \frac{1}{|X|} \sum_{U \in X} (U \otimes U)(\cdot)(U \otimes U)^\dagger \tag{3.9}$$

where the integration is with respect to the normalized Haar measure. From this definition it is clear that for X we also have,

$$\int_{U \in \mathcal{U}(\mathcal{G})} U(\cdot)U^\dagger dU = \frac{1}{|X|} \sum_{U \in X} U(\cdot)U^\dagger. \tag{3.10}$$

We refer to the set X as a unitary design. We proceed with the proof.
The expected fidelity function present in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] and [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] is achieved by averaging the fidelity function over unitary group with respect to the Haar measure. Here we show that we can replace this by an expected value achieved by taking the average over the unitaries from the relevant unitary design. This brings us to the final statement needed to prove Lemma 33, that is an implication of Lemma 204. We take the average of both sides of (C.1) with respect to the unitary design introduced in Theorem 38, to arrive at the desired expression for the expected fidelity lower-bounded. This result is essentially stated in the proof of Theorem 3.2 [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter], to which we refer for more information. In the statement, we will also use the following notation.

$$F_{c,e}(\rho, \mathcal{N}) := \max_{\mathcal{R} \in \mathcal{C}(\mathcal{H}_B, \mathcal{H}_A)} F_e(\rho, \mathcal{R} \circ \mathcal{N}), \tag{3.11}$$

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

Lemma 39 *Let X be a unitary design in $\mathcal{G}$ and $\mathcal{F} \subset \mathcal{G}$. With quantities defined as in Lemma 204, we have*

$$\begin{aligned} \mathbb{E}F_{c,e}(U\pi_{\mathcal{F}}U^\dagger, \overline{\mathcal{N}}) :=& \frac{1}{|X|} \sum_{U \in X} F_{c,e}(U\pi_{\mathcal{F}}U^\dagger, \overline{\mathcal{N}}) \\ &\geqslant tr(\overline{\mathcal{N}}(\pi_{\mathcal{G}})) - 2\sum_{j=1}^{|S|} \sqrt{kn_j} \parallel \mathcal{N}_j(\pi_{\mathcal{G}}) \parallel_2 . \end{aligned}$$

Proof 40 *In the first and more straight forward step, we take the average of first term* *on the right hand side of (C.1), namely* $w_U = tr(\overline{\mathcal{N}}(U\pi_{\mathcal{F}}U^\dagger))$;

$$\frac{1}{|X|}\sum_{U\in X} tr(\overline{\mathcal{N}}(U\pi_{\mathcal{F}}U^\dagger)) = tr(\overline{\mathcal{N}}(\frac{1}{|X|}\sum_{U\in X} U\pi_{\mathcal{F}}U^\dagger)) = \overline{\mathcal{N}}(\pi_{\mathcal{G}}). \tag{3.12}$$

What remains is the expected value of $\| D(kU\pi_{\mathcal{F}}U^\dagger) \|_1$. *To make the calculation easier we consider averaging of an upper bound on this term in terms of the* 2*-norm. From [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] we know that*

$$\| D(kU\pi_{\mathcal{F}}U^\dagger) \|_1 \leqslant \sum_{j,l=1}^{|S|} \frac{1}{|S|}\sqrt{k\min\{n_j,n_l\} \| D_{j,l}(kU\pi_{\mathcal{F}}U^\dagger) \|_2^2}.$$

Using the concavity of square root function and Jensen's inequality we have

$$\mathbb{E}(\| D(kU\pi_{\mathcal{F}}U^\dagger) \|_1) \leqslant \sum_{j,l=1}^{|S|} \frac{1}{|S|}\sqrt{k\min\{n_j,n_l\}\mathbb{E}(\| D_{j,l}(kU\pi_{\mathcal{F}}U^\dagger) \|_2^2)},$$

where the expectation is taken over the unitaries belonging to the design. To use Klesse's [Klesse(2007)] argument as done in proof of Theorem 3.2 of [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter], we must invoke the unitary invariance of $\mathbb{E}(\| D_{j,l}(kU\pi_{\mathcal{F}}U^\dagger) \|_2^2)$ *with respect to all* $U \in \mathcal{U}(\mathcal{G})$. *To see this unitary invariance, we observe that (see [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter])*

$$\| D_{j,l}(p) \|_2^2 = \frac{1}{k^2}\sum_{i=1,r=1}^{n_j,n_l} tr(p(a_{j,i}^\dagger a_{l,r})^\dagger p a_{j,i}^\dagger a_{l,r}) - |tr(p a_{j,i}^\dagger a_{l,r})|^2. \tag{3.13}$$

The unitary invariance of the expectation of the first summand is clear due to linearity of the trace function. For the expectation of the second summand we have

$$\begin{aligned}\frac{1}{|X|}\sum_{U\in X}|tr(UpU^\dagger a_{j,i}^\dagger a_{l,r})|^2 &= \frac{1}{|X|}\sum_{U\in X} tr(UpU^\dagger a_{j,i}^\dagger a_{l,r})tr(UpU^\dagger a_{l,r}^\dagger a_{j,i})\\ &= \frac{1}{|X|}\sum_{U\in X} tr(UpU^\dagger a_{j,i}^\dagger a_{l,r}\otimes UpU^\dagger a_{l,r}^\dagger a_{j,i})\\ &= \frac{1}{|X|}\sum_{U\in X} tr(U\otimes U(p\otimes p)(U\otimes U)^\dagger(A_{jilr}\otimes A_{jilr}^\dagger))\\ &= tr(\frac{1}{|X|}\sum_{U\in X} U\otimes U(p\otimes p)(U\otimes U)^\dagger(A_{jilr}\otimes A_{jilr}^\dagger)),\end{aligned}$$

where $A_{jilr} := a_{j,i}^\dagger a_{l,r}$. *From (3.9), we conclude the invariance of second summand in (3.13). Therefore we can conclude that* $\mathbb{E}(\| D_{j,l}(U\pi_{\mathcal{F}}U^\dagger) \|_2^2)$ *is indeed invariant with respect to all* $U \in \mathcal{U}(\mathcal{G})$. *The rest of the proof is exactly the same as the proof of Theorem 3.2 of [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter], yet stated here for reader's* *convenience, as follows. We can use Klesse's argument to conclude*

$$\mathbb{E}(\| D_{j,l}(kU\pi_{\mathcal{F}}U^{\dagger}) \|_2^2) \leqslant tr(\mathcal{N}_j(\pi_{\mathcal{G}})\mathcal{N}_l(\pi_{\mathcal{G}})). \tag{3.14}$$

Using (3.12), (C.1) and (3.14) we conclude

$$\mathbb{E}(F_{c,e}(U\pi_{\mathcal{F}}U^{\dagger},\overline{\mathcal{N}})) \geqslant tr(\overline{\mathcal{N}}(\pi_{\mathcal{G}})) - \sum_{j,l=1}^{|S|} \frac{1}{|S|}\sqrt{L_{jl}D_{jl}}, \tag{3.15}$$

where for $j, l \in \{1, ..., |S|\}$, we introduce abbreviations

$$L_{j,l} = k\min\{n_j, n_l\}$$

and

$$D_{j,l} = tr(\mathcal{N}_j(\pi_{\mathcal{G}})\mathcal{N}_l(\pi_{\mathcal{G}})) = \langle \mathcal{N}_j(\pi_{\mathcal{G}}), \mathcal{N}_l(\pi_{\mathcal{G}})\rangle_{HS},$$

where $\langle \cdot, \cdot \rangle_{HS}$ denotes the Hilbert Schmidt product. It is obvious that

$$L_{jl} \leqslant L_{jj} and L_{lj} \leqslant L_{ll}.$$

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

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

Therefore, an application of Lemma 199 allows us to conclude from (3.15) that

$$\mathbb{E}(F_{c,e}(U\pi_{\mathcal{F}}U^{\dagger},\overline{\mathcal{N}})) \geqslant tr(\overline{\mathcal{N}}(\pi_{\mathcal{G}})) - 2\sum_{j=1}^{|S|} \sqrt{kn_j} \| \mathcal{N}_j(\pi_{\mathcal{G}}) \|_2 .$$

Let for $\delta > 0$, $\mathcal{N}_{\delta,l,j}$ be the reduced operation associated with $\mathcal{N}_j, j \in S, |S| < \infty$ as defined by Lemma 36. Let $q_{\delta,l,j} \in \mathcal{L}(\mathcal{H})$ be the frequency-typical projection of $\mathcal{N}_{\delta,l,j}(\pi_{x^l})$ in terms of Lemma 36. Define

$$\mathcal{N}'_{\delta,l,j} := \mathcal{Q}_{\delta,l,j} \circ \mathcal{N}_{\delta,l,j} \tag{3.16}$$

where $\mathcal{Q}_{\delta,l,j}(\cdot) = q_{\delta,l,j}(\cdot)q_{\delta,l,j}$. Also define

$$\overline{\mathcal{N}}_{\delta,l} := \frac{1}{|S|}\sum_{j=1}^{|S|}\mathcal{N}'_{\delta,l,j}$$

Applying Lemma 39 on $\{\mathcal{N}'_{\delta,l,j}\}_{j\in S}$, with expectation taken over unitaries from a unitary design on $\mathcal{U}(\mathcal{G}_{x^l})$ we obtain

$$\mathbb{E}F_{c,e}(U\pi_{\mathcal{F}}U^{\dagger},\overline{\mathcal{N}}_{\delta,l}) \geqslant \operatorname{tr}(\overline{\mathcal{N}}_{\delta,l}(\pi_{x^l})) - 2\sum_{j=1}^{|S|}\sqrt{kn_{\delta,l,j}} \parallel \mathcal{N}'_{\delta,l,j}(\pi_{x^l}) \parallel_2 . \tag{3.17}$$

We may now follow the steps taken in proof of Theorem 5 from [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] to give a lower bound on each of the terms on the right hand side of (3.17) using Lemmas 34 and 36, to derive the following result.

Lemma 41 *Let $\mathfrak{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H},\mathcal{K})$ be a compound channel, $\delta > 0$ and $\lambda \in \mathcal{P}(\mathcal{X})$. For subspaces $(G_x)_{x\in\mathcal{X}}, \mathcal{G}_x \subset \mathcal{H}, x \in \mathcal{X}$, there exists $l_0 \in \mathbb{N}$ such that for each $l \geqslant l_0$ and $x^l \in T^l_\lambda$, we find a subspace $\mathcal{F}_l \subset \mathcal{G}_{x^l}$ and $(l, \dim(\mathcal{F}_l))$ entanglement transmission codes $(\mathcal{P}_i,\mathcal{R}_i)_{i=1}^{|X|}$ with $|X| < \infty$ such that,*

1. $\dim(\mathcal{F}_l) \geqslant 2^{\inf_{s\in S} I_c(\pi_{x^l},\mathcal{N}_s^{\otimes l})-l\delta}$ and

2. $\inf_{s\in S}\frac{1}{|X|}\sum_{i=1}^{|X|}F_e(\pi_{\mathcal{F}_l},\mathcal{R}_i\circ\mathcal{N}_s^{\otimes l}\circ\mathcal{P}_i) \geqslant 1-\epsilon_l$ with $\epsilon_l \to 0$ as $l\to\infty$.

Proof 42 *Let $\mathfrak{J}_\tau$ with index set S_τ be the net associated with $\mathfrak{J}$ in terms of Lemma 195. Choose $\delta' \in (0,1/2)$ and $l_0 \in \mathbb{N}$ satisfying $\gamma(\delta') + \phi(\delta') + \check{c}h(l_0) \leqslant \frac{\delta}{2}$ with functions γ,ϕ,h and constant $\check{c}$ from Lemmas 34 and 36. Now choose for every $l \geqslant l_0$, a subspace $\mathcal{F}_l \subset \mathcal{G}_{x^l}$ such that*

$$\dim(\mathcal{F}_l) := k_l = \lfloor 2^{\min_{s\in S_\tau} I_c(\pi_{x^l},\mathcal{N}_s^{\otimes l})-l\delta}\rfloor. \tag{3.18}$$

This is always possible as $S(\pi_{\mathcal{G}_{x^l}}) \geqslant I_c(\pi_{x^l},\mathcal{N}_s^{\otimes l})$. We have

$$\min_{s\in S_\tau} I_c(\pi_{x^l},\mathcal{N}_s^{\otimes l}) - l\delta - o(l_0) \leqslant \log k_l \leqslant \min_{s\in S_\tau} I_c(\pi_{x^l},\mathcal{N}_s^{\otimes l}) - l\delta. \tag{3.19}$$

We assume for the moment that $x^l \in T^l_\lambda$ is given by concatenation of homogeneous words of size $N_x := N(x|x^l)$. That is, for $\mathcal{A} := \{x \in \mathcal{X} : N_x \neq 0\} \subset \mathcal{X}$, we have $x^l = (x^{N_x})_{x\in\mathcal{A}}$. As such, the hypotheses of Lemma 34 and Lemma 36 apply to to product states indexed by x^l. This assumption however, does not prohibit generality of the proven results, since each word of type λ results from a permutation of the letters of word x^l. Namely, for any word $\tilde{x}^l \in T^l_\lambda$, there exists a permutation mape γ with $\gamma(x^l) = \tilde{x}^l$. Therefore, given codes $(\mathcal{P}_i,\mathcal{R}_i)_{i\in X}$ for x^l with the properties mentioned in the statement of the present lemma, suitable codes for $\tilde{x}^l$ will be given by $(\mathcal{U}_\gamma\circ\mathcal{P}_i\circ\mathcal{U}_\gamma^{-1},\mathcal{U}_\gamma^{-1}\circ\mathcal{R}_i\circ\mathcal{U}_\gamma)$, with $\mathcal{U}_\gamma$ the CPTP map permuting the tensor factors according to γ.

We now give lower bounds for the terms on the right hand side of (3.17).

$$\begin{aligned}\operatorname{tr}(\overline{\mathcal{N}}_{\delta',l}(\pi_{x^l})) &= \frac{1}{|S_\tau|}\sum_{s=1}^{|S_\tau|}\operatorname{tr}(\mathcal{N}'_{\delta',l,s}(\pi_{x^l})) &(3.20)\\ &= \frac{1}{|S_\tau|}\sum_{s=1}^{|S_\tau|}\left[\operatorname{tr}(Q_{\delta',l,s}\circ\mathcal{N}_s^{\otimes l}(\pi_{x^l})) - \operatorname{tr}(Q_{\delta',l,s}\circ[\mathcal{N}^{\otimes l}-\mathcal{N}_{\delta',l,s}](\pi_{x^l}))\right]\\ &\geqslant 1-|\mathcal{X}|(2^{-l(\tilde{c}\delta'^2-h(l))}-2^{-l(c\delta'^2-h(l))}). &(3.21)\end{aligned}$$

In the last inequality we have inserted the bounds from Lemmas 34 and 36, after using $0 \leqslant \operatorname{tr}(Q_{\delta',l,s}\circ[\mathcal{N}^{\otimes l}-\mathcal{N}_{\delta',l,s}](\pi_{x^l})) \leqslant \operatorname{tr}([\mathcal{N}^{\otimes l}-\mathcal{N}_{\delta',l,s}](\pi_{x^l}))$*. Also,*

$$\begin{aligned}\parallel \mathcal{N}'_{\delta',l,s}(\pi_{x^l}) \parallel_2^2 &\leqslant\parallel Q_{\delta',l,s}\circ\mathcal{N}_{\delta',l,s}(\pi_{x^l}) \parallel_2^2 + \parallel Q_{\delta',l,s}\circ(N_s^{\otimes l}-\mathcal{N}_{\delta',l,s})(\pi_{x^l}) \parallel_2^2\\ &\leqslant\parallel Q_{\delta',l,s}\circ\mathcal{N}_s^{\otimes l}(\pi_{x^l}) \parallel_2^2\leqslant 2^{-(S(\pi_{x^l})-l\phi(\delta'))}. \qquad (3.22)\end{aligned}$$

In the second inequality we have used $\parallel A \parallel_2^2 + \parallel B \parallel_2^2\leqslant\parallel A+B \parallel_2^2$ *for non-negative operators* $A, B \in \mathcal{L}(\mathcal{K}^{\otimes l})$ *(see [Klesse(2007)]), and inserted the lower bound from Lemma 34. Inserting the bounds from (3.20) and (3.22) into (3.17) we obtain*

$$\begin{aligned}\mathbb{E}F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger,\overline{\mathcal{N}}_{\delta',l}) &\geqslant 1-|\mathcal{X}|\left[2^{-l(c\delta'^2-h(l))}-2^{-l(\tilde{c}\delta'^2-h(l))}\right]\\ &\quad -2\sum_{s=1}^{|S_\tau|}\sqrt{2^{\log k_l - S(\pi_{x^l})+l\phi(\delta')+S_e(\pi_{x^l},\mathcal{N}_s^{\otimes l})+l(\gamma(\delta')+\check{h}(l))}}\\ &\geqslant 1-|\mathcal{X}|\left[2^{-l(c\delta'^2-h(l))}-2^{-l(\tilde{c}\delta'^2-h(l))}\right]-2|S_\tau|\sqrt{2^{-l(\delta-\phi(\delta')-\gamma(\delta')-\check{h}(l))}}.\end{aligned} \qquad (3.23)$$

In the second inequality above we have inserted the upper bound for k_l *from (3.19). For* $l \geqslant l_0$*, (3.23) gives us an exponential decay of error. Therefore we can write*

$$\mathbb{E}F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger,\overline{\mathcal{N}}_{\delta',l}) \geqslant 1-\epsilon_{1,l}-|S_\tau|\epsilon_{2,l}$$

with $\epsilon_{i,l}\to 0$ *with* $l\to\infty$ *for* $i=1,2$*. From this we conclude*

$$\min_{s\in S_\tau}\mathbb{E}F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger, Q_{\delta',l,s}\circ\mathcal{N}_{\delta',l,s}) \geqslant 1-|S_\tau|\epsilon_{1,l}-|S_\tau|^2\epsilon_{2,l}.$$

From the third property under Lemma 36, the above inequality implies

$$\min_{s\in S_\tau}\mathbb{E}F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger, Q_{\delta',l,s}\circ\mathcal{N}_s^{\otimes l}) \geqslant 1-|S_\tau|\epsilon_{1,l}-|S_\tau|^2\epsilon_{2,l} \geqslant 1-|S_\tau|^2\epsilon_{0,l}, \qquad (3.24)$$

where $\epsilon_{0,l} := \max_{i=1,2}\epsilon_{i,l}$*. Setting shorthand notation* $\beta_{s,U} := 1-F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger, Q_{\delta',l,s}\circ\mathcal{N}_s^{\otimes l})$*, we obtain from Lemma 200,* $F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger,\mathcal{N}_s^{\otimes l}) \geqslant 1-3\beta_{s,U}$*. Hence from (3.24) we* *conclude*

$$\min_{s\in S_\tau} \mathbb{E}F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger, \mathcal{N}_s^{\otimes l}) \geqslant 1 - 3|S_\tau|^2\epsilon_{0,l}. \tag{3.25}$$

By Lemma 195, we have

$$\min_{s\in S} \mathbb{E}F_{c,e}(U\pi_{\mathcal{F}_l}U^\dagger, \mathcal{N}_s^{\otimes l}) \geqslant 1 - 3|S_\tau|^2\epsilon_{0,l} - 2l\tau. \tag{3.26}$$

Given that we find $|S_\tau| \leqslant (\frac{6}{\tau})^{2(d\cdot d')^2}$, choosing $\tau = \epsilon_{0,n}^{\frac{1}{8(d\cdot d')^2}}$, we have the desired exponential decay of error. Also, as $\mathcal{J}_\tau \subset \mathcal{J}$, we obtain the desirable lower bound on the rate.

Proof 43 (Proof of Lemma 33) *Assume $x^l \in T^l_\lambda$. Note that $\forall \lambda \in \mathcal{T}(\mathcal{X}, l)$, either $T^l_\lambda \subset T^l_{p,\delta}$ or $T^l_\lambda \bigcap T^l_{p,\delta} = \emptyset$. Since by assumption of the lemma $x^l \in T^l_{p,\delta}$, we conclude $T^l_\lambda \subset T^l_{p,\delta}$. For each $\tilde{\delta} > 0$, we have from Lemma 41 applied on the compound channel $\mathcal{J} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$, a subspace $\mathcal{F}_{A,l} \subset \mathcal{H}_A^{\otimes l}$ with*

$$\begin{aligned}\frac{1}{l}\log\dim(\mathcal{F}_{A,l}) \geqslant \frac{1}{l}\inf_{s\in S} I_c(\pi_{x^l}, \mathcal{N}_s^l) - \tilde{\delta} &= \inf_{s\in S}\sum_{x\in\mathcal{A}} \frac{1}{l} I_c(\pi_x^{\otimes N_x}, \mathcal{N}_s^{\otimes N_x}) - \tilde{\delta} \\ &= \inf_{s\in S}\sum_{x\in\mathcal{A}} \lambda(x) I_c(\pi_x, \mathcal{N}_s) - \tilde{\delta} \\ &\geqslant \inf_{s\in S}\sum_{x\in\mathcal{A}} p(x)(I_c(\pi_x, \mathcal{N}_s) \\ &- \tilde{\delta}) - |\lambda(x) - p(x)| \cdot (I_c(\pi_x, \mathcal{N}_s) - \tilde{\delta}) \\ &\geqslant \inf_{s\in S} I(A\rangle BX, \omega(\mathcal{N}_s, p, \varPhi)) - \tilde{\delta} - |\mathcal{X}|\overline{c}\tilde{\delta}.\end{aligned}$$

with $\overline{c} := 2\log\dim(\mathcal{H}_A \otimes \mathcal{H}_B)$ and ω_s defined by (3.4). With this rate we obtain exponential decay of error as explained above. Choosing $\tilde{\delta}$ such that $\delta > \tilde{\delta} + |\mathcal{X}|\overline{c}\tilde{\delta}$, we obtain the desired lower bound on the rate. The last property listed under Lemma 33 is clear by averaging property of the Haar measure, reproduced here by the unitary design in $\mathcal{U}(\mathcal{G}_{x^l})$.

3.3.2. Classical message transmission codes

The desired statement of universal codes for c-q channels can be extracted from [Mosonyi(2015)]. Therein, the authors have introduced universal random codes for transmission of classical messages over c-q channels, using properties of Renyi entropies. Based on the same codes, we have derived the following lemma to allow for a faster decay of error while considering only " typical " inputs.

Lemma 44 *Let $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ and $V : \mathcal{X} \to \mathcal{S}(\mathcal{H}_A)$ be a c-q channel. For each $\eta > 0$ and $p \in \mathcal{P}(\mathcal{X})$, there exists a number n_0, such that for $n \geqslant n_0$, there exists a classical encoding map $u : m \to u_m \in \mathcal{X}^n$ and decoding POVM $(\varLambda)_{m\in[M_n]}$ such that*

1. *$\forall m \in [M_n] : u_m \in T^n_{p,\eta}$,*

2. $\inf_{s\in S}\min_{m\in M_n}\operatorname{tr}((\mathcal{N}_s\circ V)^{\otimes n}(u_m)\Lambda_m)\geqslant 1-\epsilon_n$*, with* $\epsilon_n\to 0$ *exponentially as* $n\to\infty$*,*

3. $\frac{1}{n}\log M_n\geqslant\inf_{s\in S} I(X;B,\omega(\mathcal{N}_s,p,\Psi))-c\eta$

with $\omega(\mathcal{N}_s,p,\Psi)$ *defined by (4.4) for* $\Psi:=(\Psi_x : x\in\mathcal{X},\operatorname{tr}_{\mathcal{H}}(\Psi_x)=V(x))$ *and constant* $c>0$*.*

3.4. Proofs for the compound channel

In this section we proceed with the proof of Theorem 25 in two parts. In the following section (converse part), it is also demonstrated that CSI at the decoder does not improve channel's classically enhanced entanglement generation capacity. In the more involved direct part of the proof, we introduce classically enhanced entanglement transmission codes by marrying classical message transmission codes from [Mosonyi(2015)] and a generalization of entanglement transmission codes from [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] and [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] as stated in Section 3.3.

3.4.1. Proof of the converse

In this section we prove the following lemma.

Lemma 45 *Let* $\mathcal{J}:=\{\mathcal{N}_s\}_{s\in S}\subset\mathcal{C}(\mathcal{H}_A,\mathcal{H}_B)$ *be any compound quantum channel. It holds*

$$\overline{C}_{CEG}(\mathcal{J})\subset\operatorname{cl}\left(\bigcup_{l=1}^{\infty}\frac{1}{l}\bigcup_{p,\Psi}\bigcap_{s\in S}\hat{C}(\mathcal{N}_s^{\otimes l},p,\Psi)\right).\tag{3.27}$$

To prove this result, we shall make use of the following lemma (see [Devetak(2005)]).

Lemma 46 *For two states* ρ^{AB} *and* σ^{AB} *on some Hilbert space* $\mathcal{K}_A\otimes\mathcal{K}_B$ *of dimension* r *and fidelity* $f:=F(\rho^{AB},\sigma^{AB})$*, we have*

$$|I(A\rangle B,\rho)-I(A\rangle B,\sigma)|\leqslant\frac{2}{e}+4\log r\sqrt{1-f}.$$

Proof 47 (Proof of Lemma 45) *We prove a more general claim than stated in Lemma 45, allowing the decoder to choose the processing according to the channel state (i.e. the decoder has access to CSI). Let for each* $n\in\mathbb{N}$*,* $\mathcal{C}_{CEG,s}:=(\Psi_m,\mathcal{R}_{m,s})_{m\in M_{1,n}}$ *be an* $(n,M_{1,n},M_{2,n})$ *CEG code with informed decoder*[2]*, such that*

$$\inf_{s\in S}\overline{P}(\mathcal{C}_{CEG,s},\mathcal{N}_s^{\otimes n})\geqslant 1-\epsilon,\tag{3.28}$$

[2] As clear from the notation, these codes are CEG codes for compound channel $\mathcal{J}$, when the decoder has access to CSI.

with $\epsilon < 1$ *holds. Fix* $n \in \mathbb{N}$ *and let* $p_* \in \mathcal{P}(\mathcal{X}^n)$ *be the equidistribution on the message set. Consider the pair* (M_s, M_s') *of random variables with joint distribution:*

$$Pr(M_s = m, M_s' = m') = p_*(m) tr(\mathcal{R}_{m',s} \circ \mathcal{N}_s^{\otimes n}(V(m)))$$

for $(m_s, m_s' \in [M_{1,n}])$ *and* $s \in S$ *with* $V(m) := tr_{\mathcal{F}_{A,n}} \Psi_m$ *for some c-q channel* $V : \mathcal{X}^n \to \mathcal{S}(\mathcal{H}_A^{\otimes n})$*. Note that with these definitions, we have*

$$\mathbb{P}(M_s \neq M_s') \leqslant 1 - \overline{P}(\mathcal{C}_{CEG,s}, \mathcal{N}_s^{\otimes n}) \leqslant \epsilon, \tag{3.29}$$

for $s \in S$*. Fix* s *for the moment. Define the state*

$$\sigma_s' := \sum_{m \in [M_{1,n}]} p_*(m) \left|m\right\rangle\left\langle m\right|^X \otimes (id_{\mathcal{H}_A^{\otimes n}} \otimes \mathcal{R}_{s,m} \circ \mathcal{N}_s^{\otimes n})(\Psi_m)$$

and the shorthand notation

$$\sigma_s := \omega(\mathcal{N}_s^{\otimes n}, p_*, \Psi) = \sum_{m \in [M_{1,n}]} p_*(m) \left|m\right\rangle\left\langle m\right|^X \otimes (id_{\mathcal{H}_A^{\otimes n}} \otimes \mathcal{N}_s^{\otimes n})(\Psi_m).$$

We have

$$\begin{aligned} \log M_{1,n} = H(p_*) = I(M_s; M_s') + H(M_s'|M_s) &\leqslant I(M_s; M_s') + \epsilon \log M_{1,n} + 1 \\ &\leqslant I(X; B, \sigma_s) + \epsilon \log M_{1,n} + 1 \\ &\leqslant I(X; B, \sigma_s) + n\epsilon \log |\mathcal{X}| + 1, \end{aligned} \tag{3.30}$$

where $I(Y; Y')$ *is the mutual information of random variables* Y, Y'*. The first inequality comes from (3.29) and the second is by Holevo bound (see [Wilde(2017)]). For* $s \in S$*, We have*

$$\epsilon \geqslant 1 - \overline{P}(\mathcal{C}_{CEG,s}, \mathcal{N}_s^{\otimes n}) = 1 - F(\Phi, \sigma_s'^{AB}), \tag{3.31}$$

where $\sigma_s'^{AB} := \mathrm{tr}_X(\sigma_s')$*. We have*

$$\begin{aligned} I(A\rangle BX : \sigma_s) &\geqslant I(A\rangle BX, \sigma_s') \\ &\geqslant I(A\rangle B, \sigma_s'^{AB}) \\ &\geqslant I(A\rangle B, \Phi) - \frac{2}{e} - 8n \log \dim \mathcal{H} \sqrt{\epsilon} = \log M_{2,n} - \frac{2}{e} - 8n \log \dim \mathcal{H} \sqrt{\epsilon}. \end{aligned} \tag{3.32}$$

In (3.32), the first inequality comes from the quantum data processing inequality, the second comes from the fact that conditioning does not decrease coherent information. The third inequality comes from Lemma 46 together with (3.31) and finally, in the last line we

have used $I(A\rangle B, \Phi) = \log M_{2,n}$.
Choosing n *such that* $\delta \geqslant \frac{2}{ne} + 8 \log \dim(\mathcal{H})\sqrt{\epsilon}$ *, from (3.32) and (3.30) we obtain*

$$(\frac{1}{n}\log M_{1,n} - \delta, \frac{1}{n}\log M_{2,n} - \delta) \in \frac{1}{n}\hat{C}(\mathcal{N}_s^{\otimes n}, p_*, \Psi).$$

Since $s \in S$ *was arbitrary, we have shown*

$$(\frac{1}{n}\log M_{1,n} - \delta, \frac{1}{n}\log M_{2,n} - \delta) \in \text{cl}[\bigcup_{n=1}^{\infty}\bigcup_{p,\Psi}\bigcap_{s\in S}\frac{1}{n}\hat{C}(\mathcal{N}_s^{\otimes n}, p, \Psi)]. \tag{3.33}$$

3.4.2. Proof of the direct part

In this section we prove the following lemma.

Lemma 48 *Let* $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ *be any compound quantum channel. It holds*

$$\text{cl}\left(\bigcup_{l=1}^{\infty}\frac{1}{l}\bigcup_{p,\Psi}\bigcap_{s\in S}\hat{C}(\mathcal{N}_s^{\otimes l}, p, \Psi)\right) \subset C_{CET}(\mathcal{J}). \tag{3.34}$$

In the first step towards proving the above statement, we restrict the encoder to maximally mixed state inputs. The final result will then be a generalization by way of which we lift this restriction. We state the first instance of the classically enhanced codes, satisfying classical and quantum error criteria in the following lemma.

Lemma 49 *Let* $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ *be any quantum compound channel. For finite alphabet* $\mathcal{X}$*, subspaces* $(\mathcal{G}_x)_{x\in\mathcal{X}}, \mathcal{G}_x \subset \mathcal{H}_A \ \forall x \in \mathcal{X}$*,* $p \in \mathcal{P}(\mathcal{X}), V_\pi : \mathcal{X} \to \mathcal{S}(\mathcal{H}_A)$ *with* $V_\pi(x) = \pi_x, x \in \mathcal{X}$*, each* $\delta > 0$ *and large enough values of* n*, there exists an* $(n, M_{1,n}, M_{2,n})$ *CET code with* $M_{2,n} = \dim(\mathcal{F}_{A,n})$ *such that*

1. $\frac{1}{n}\log M_{2,n} \geqslant \inf_{s\in S} I(A\rangle BX, \omega(\mathcal{N}_s, p, \Phi)) - \delta$,
2. $\frac{1}{n}\log M_{1,n} \geqslant \inf_{s\in S} I(X; B, \omega(\mathcal{N}_s, p, \Phi)) - c\delta$ *with some constant* $c > 0$ *and* $\omega(\mathcal{N}_s, p, \Phi)$ *defined by (4.4) for* $\Phi := (\Phi_x : x \in \mathcal{X})$ *defined as in Section 3.3.1,*
3. $\inf_{s\in S}\min_{m\in[M_{1,n}]} \ P(\mathcal{C}_{CET}, \mathcal{N}_s^{\otimes n}, m) \geqslant 1 - \epsilon_n$*, with* $\epsilon_n \to 0$ *exponentially as* $n \to \infty$.

Proof 50 *Let* $\mathcal{J}_\tau \subset \mathcal{J}$ *be as defined in Appendix B, Lemma 195 with index set* S_τ*. According to Lemma 44, for* $\delta > 0$ *and large enough values of* $n \in \mathbb{N}$*, we find pairs* $(u_m, \Lambda_m)_{m\in[M_{1,n}]}$ *with* $\frac{1}{n}\log M_{1,n} \geqslant \min_{s\in S_\tau} I(X; B, \omega(\mathcal{N}_s, p, \Phi)) - c\delta$*, such that for channel* V_π *we have*

$$\min_{s\in S_\tau}\min_{m\in[M_{1,n}]} tr(\Lambda_m(\mathcal{N}_s \circ V_\pi)^{\otimes n}(u_m)) = \min_{s\in S_\tau}\min_{m\in[M_{1,n}]} tr(\Lambda_m \circ \mathcal{N}_s^{\otimes n}(\pi_{u_m})) \geqslant 1 - \epsilon_{1,n}, \tag{3.35}$$

for $u_m \in T^n_{p,\delta}$ *and* $\epsilon_{1,n}$ *going to zero exponentially. Given* $u_m \in T^n_{p,\delta}$ *for each* m*, according to Lemma 33, there exists a family of entanglement transmission codes* $(\mathcal{P}_i^{(m)}, \tilde{\mathcal{R}}_i^{(m)})_{i=1}^{|X_{u_m}|}$

with rate $\frac{1}{n}\log M_{2,n} \geqslant$
$\min_{s\in S_\tau} I(A\rangle BX, \omega(\mathcal{N}_s, p, \Phi)) - \delta$*, such that* π_{u_m} *is exactly the output of the average of encoding operations (third statement of the lemma) and*

$$\min_{s\in S_\tau}\min_{m\in[M_{1,n}]}\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} F_e(\pi_{F_{A,n}}, \tilde{\mathcal{R}}_i^{(m)}\circ\mathcal{N}_s^{\otimes n}\circ\mathcal{P}_i^{(m)}) \geqslant 1-\epsilon_{2,n} \tag{3.36}$$

with $\epsilon_{2,n}\to 0$ *exponentially. Thus (3.35) yields*

$$\min_{s\in S_\tau}\min_{m\in[M_{1,n}]}\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} tr(\Lambda_m\mathcal{N}_s^{\otimes n}(\mathcal{P}_i^{(m)}(\pi_{\mathcal{F}_{A,n}})) \geqslant 1-\epsilon_{1,n}. \tag{3.37}$$

Following [Devetak and Shor(2005)], the encoding and decoding maps are given by $(\mathcal{P}_i^{(m)}, \mathcal{R}_i^{(m)})_{i=1}^{|X_{u_m}|}$ *with*

$$\mathcal{R}_i^{(m)}(\rho) = \tilde{\mathcal{R}_i}^{(m)}(\sqrt{\Lambda_m}\rho\sqrt{\Lambda_m}).$$

It can be observed that for each i we have $\sum_{m\in[M_{1,n}]}\mathcal{R}_i^{(m)}\in\mathcal{C}(\mathcal{H}_B^{\otimes n},\mathcal{F}_{B,n})$*.*

From (3.37) we obtain

$$\min_{s\in S_\tau}\min_{m\in[M_{1,n}]}\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} tr(\mathcal{R}_i^{(m)}\circ\mathcal{N}_s^{\otimes n}(\mathcal{P}_i^{(m)}(\pi_{\mathcal{F}_{A,n}}))) \geqslant 1-\epsilon_{1,n}. \tag{3.38}$$

We define the following state

$$\chi_{i,s}^{(m)} := [id\otimes(\mathcal{N}_s^{\otimes n}\circ\mathcal{P}_i^{(m)})](\Phi_{\mathcal{F}_{A,n}}),$$

where $\Phi_{\mathcal{F}_{A,n}}$ *is a maximally entangled state given by purification of* $\pi_{\mathcal{F}_{A,n}}$*. From (3.37) we obtain*

$$\min_{s\in S_\tau}\min_{m\in[M_{1,n}]}\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} tr\chi_{i,s}^{(m)}(id\otimes\Lambda_m) \geqslant 1-\epsilon_{1,n}. \tag{3.39}$$

Set $\gamma_{i,s,m} := \mathrm{tr}\chi_{i,s}^{m}(\mathrm{id}\otimes\Lambda_m)$*. It is clear that if* $\gamma_{i,s,m}=0$*, we have (3.41). To prove this equation for the case where* $\gamma_{i,s,m}>0$*, we observe that by the gentle measurement lemma (Lemma 203), we have for all* i,m,s

$$\|\ \frac{(\mathrm{id}\otimes\sqrt{\Lambda_m})(\chi_{i,s}^{m})(\mathrm{id}\otimes\sqrt{\Lambda_m})}{\gamma_{i,s,m}} - \chi_{i,s}^{(m)}\ \|_1 \leqslant 2\sqrt{1-\gamma_{i,s,m}}$$

and hence by monotonicity of trace distance under CPTP maps we obtain

$$\|\ \frac{1}{\gamma_{i,s,m}}(id\otimes\mathcal{R}_i^{(m)})(\chi_i^{(m)}) - (id\otimes\tilde{\mathcal{R}}_i^{(m)})(\chi_i^{(m)})\ \|_1 \leqslant 2\sqrt{1-\gamma_{i,s,m}}. \tag{3.40}$$

Applying Lemma 198 and averaging with respect to index i, the above inequality yields

$$\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} F_e(\pi_{\mathcal{F}_{A,n}}, \mathcal{R}_i^{(m)} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_i^{(m)}) \geqslant \frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} (F_e(\pi_{\mathcal{F}_{A,n}}, \tilde{\mathcal{R}}_i^{(m)} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_i^{(m)}) - 2\sqrt{1-\gamma_{i,s,m}})\gamma_{i,s,m}. \tag{3.41}$$

To give a suitable lower bound for (3.41), we use Lemma 206. We observe that,

$$\begin{aligned}\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} F_e(\pi_{\mathcal{F}_{A,n}}, \tilde{\mathcal{R}}_i^{(m)} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_i^{(m)}) - 2\sqrt{1-\gamma_{i,s,m}} \geqslant \\ \frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} F_e(\pi_{\mathcal{F}_{A,n}}, \tilde{\mathcal{R}}_i^{(m)} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_i^{(m)}) - 2\sqrt{1-\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|}\gamma_{i,s,m}} \\ \geqslant 1-\epsilon_{2,n}-2\sqrt{\epsilon_{1,n}},\end{aligned} \tag{3.42}$$

where in the first inequality we have used concavity of the square function along with Jensen's inequality, and in the second one we have used the bounds from (3.36) and (3.39). Setting $\epsilon_{3,n} := \max\{\epsilon_{2,n} - 2\sqrt{\epsilon_{1,n}}, \epsilon_{1,n}\}$, by Lemma 206, (3.42), (3.39) and (3.41) imply

$$\frac{1}{|X_{u_m}|}\sum_{i=1}^{|X_{u_m}|} F_e(\pi_{\mathcal{F}_{A,n}}, \mathcal{R}_i^{(m)} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_i^{(m)}) \geqslant 1-2\epsilon_{3,n}. \tag{3.43}$$

This means that for each m there exists a value $i(m)$ such that:

$$\frac{1}{|S_\tau|}\sum_{s\in S_\tau} F_e(\pi_{\mathcal{F}_{A,n}}, \mathcal{R}_{i(m)}^{(m)} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_{i(m)}^{(m)}) \geqslant 1-2\epsilon_{3,n}.$$

Therefore setting $\mathcal{R} := \sum_{m\in[M_{1,n}]} |m\rangle\langle m| \otimes \mathcal{R}_{i(m)}^{(m)}$ and $\mathcal{P}_m := \mathcal{P}_{i(m)}^{(m)}$ for all $m \in [M_{1,n}]$ for all $s \in S_\tau$ and $m \in [M_{1,n}]$, we have for $\mathcal{C}_{CET} := (\mathcal{P}_m, \mathcal{R}_m)_{m\in[M_{1,n}]}$ with

$$\begin{aligned}P(\mathcal{C}_{CET}, \mathcal{N}_s^{\otimes n}, m) &= F(|m\rangle\langle m| \otimes \Phi^{AB}, \mathrm{id}_{\mathcal{F}_{A,n}} \otimes \mathcal{R} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_m) \\ &= F_e(\pi_{\mathcal{F}_{A,n}}, \mathcal{R}_{i(m)}^{(m)} \circ \mathcal{N}_s^{\otimes n} \circ \mathcal{P}_{i(m)}^{(m)}) \geqslant 1-2|S_\tau|\epsilon_{3,n}.\end{aligned} \tag{3.44}$$

By the third property of $\mathcal{J}_\tau$ stated under Lemma 195, we have for all $s \in S$ and $m \in [M_{1,n}]$

$$P(\mathcal{C}_{CET}, \mathcal{N}_s^{\otimes n}, m) \geqslant 1-2|S_\tau|\epsilon_{3,n} - 2n\tau,$$

Given that we find $|S_\tau| \leqslant (\frac{6}{\tau})^{2(d\cdot d')^2}$, choosing $\tau = \epsilon_{3,n}^{\frac{1}{4(d\cdot d')^2}}$, we have the desired exponential decay of error. Also we obtain the desirable rates as $\mathcal{J}_\tau \subset \mathcal{J}$.

We now run an instance of concatenation upon codes from Lemma 49, to achieve suitable codes without the restriction imposed by V_π. The method used here for lifting this restriction is rather elementary[3] given that the input state can be decomposed as a convex combination of maximally mixed states.

Lemma 51 *For compound channel $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$, $p \in \mathcal{P}(\mathcal{X})$, $V : \mathcal{X} \to \mathcal{S}(\mathcal{H}_A)$ and large enough values of n, there exists a CET codes, $\mathcal{C}_{CET} := (\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,n}]}$ such that*

1. $\liminf_{n\to\infty} \frac{1}{n} \log M_{2,n} \geqslant \inf_{s\in S} I(A\rangle BX, \omega_s(\mathcal{N}_s, p, \Psi))$,

2. $\liminf_{n\to\infty} \frac{1}{n} \log M_{1,n} \geqslant \inf_{s\in S} I(A;B, \omega_s(\mathcal{N}_s, p, \Psi))$ *hold with $\omega_s(\mathcal{N}_s, p, \Psi)$ defined by (4.4),*

3. $\inf_{s\in S} \min_{m\in[M_{1,n}]}\ P(\mathcal{C}_{CET}, \mathcal{N}^{\otimes n}, m) \geqslant 1 - \epsilon_n$,

with $\epsilon_n \to 0$ exponentially as $n \to \infty$.

Proof 52 *For $x \in \mathcal{X}$, let $V(x)$ have the spectral decomposition*

$$V(x) = \sum_{y\in\mathcal{Y}} q_x(y) \left|\phi_x^y\right\rangle\!\left\langle\phi_x^y\right| ,$$

with $\mathcal{Y}$ an alphabet with $|\mathcal{Y}| = \dim(\mathcal{H}_A)$, $\{|\phi_x^y\rangle\}_{y\in\mathcal{Y}}$ an ONB and $q_x \in \mathcal{P}(\mathcal{Y})$ for each $x \in \mathcal{X}$. It can be seen that for $l \in \mathbb{N}$ and $x^l \in \mathcal{X}^l$ we have

$$V^{\otimes l}(x^l) = \sum_{y^l\in\mathcal{Y}^l} q_{x^l}(y^l) \left|\phi_{x^l}^{y^l}\right\rangle\!\left\langle\phi_{x^l}^{y^l}\right| . \tag{3.45}$$

For each $x^l \in \mathcal{X}^l$ and $\lambda \in \mathcal{T}(\mathcal{X} \times \mathcal{Y}, l)$, define the following sets

$$\mathcal{A}_\lambda(x^l) := \{y^l : (x^l, y^l) \in T_\lambda^l\}. \tag{3.46}$$

Given the properties of typical sets, it can be observed that $\mathcal{A}_\lambda(x^l) \bigcap \mathcal{A}_{\lambda'}(x^l) = \varnothing$ for all pairs (λ, λ') with $\lambda \neq \lambda'$. Also, $\bigcup_{\lambda\in\mathcal{T}(\mathcal{X}\times\mathcal{Y},l)} \mathcal{A}_\lambda(x^l) = \mathcal{Y}^l$. Given these properties, from (3.45) we obtain

$$V^{\otimes l}(x^l) = \sum_{\lambda\in\mathcal{T}(\mathcal{X}\times\mathcal{Y},l)} q_{x^l}(\lambda) \sum_{y^l\in\mathcal{A}_\lambda(x^l)} \left|\phi_{x^l}^{y^l}\right\rangle\!\left\langle\phi_{x^l}^{y^l}\right| = \sum_{\lambda\in\mathcal{T}(\mathcal{X}\times\mathcal{Y},l)} q_{x^l}(\lambda)\pi_{x^l}^\lambda, \tag{3.47}$$

with $\pi_{x^l}^\lambda := \frac{1}{|\mathcal{A}_\lambda(x^l)|} \sum_{y^l\in\mathcal{A}_\lambda(x^l)} \left|\phi_{x^l}^{y^l}\right\rangle\!\left\langle\phi_{x^l}^{y^l}\right|$ and $q_{x^l}(\lambda) = q_{x^l}(y^l)|\mathcal{A}_\lambda(x^l)|$ for any $y^l \in \mathcal{A}_\lambda(x^l)$. The above decomposition therefore comes from the fact that for all $y^l \in \mathcal{A}_\lambda(x^l)$, $q_{x^l}(y^l)$ is constant. Define probability distribution $r \in \mathcal{P}(\mathcal{X}^l, \mathcal{T}(\mathcal{X}\times\mathcal{Y}, l))$ with $r(x^l, \lambda) = p^l(x^l)q_{x^l}(\lambda)$.

[3] Compare with BSST type lemmas used for instance in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel]

Also define the state

$$\sigma_s := \sum_{(x^l,\lambda)\in\mathcal{X}^l\times\mathcal{T}(\mathcal{Y},l)} r(x^l,\lambda)\,|e_{x^l}\rangle\langle e_{x^l}|^X \otimes |e_\lambda\rangle\langle e_\lambda|^T \otimes \mathbb{1}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{N}_s^{\otimes l}(\Phi_{x^l}^\lambda),$$

where $\Phi_{x^l}^\lambda$ *is a purification of* $\pi_{x^l}^\lambda$*, a maximally entangled state on subspace* $\mathcal{G}_{x^l}^\lambda \subset \mathcal{H}_A^{\otimes l}$*. According to Lemma 49, for* $V_\pi : V_\pi(x^l,\lambda) = \pi_{x^l}^\lambda$*, large enough values of* $a \in \mathbb{N}$ *and* $\delta > 0$*, we find a subspace* $\mathcal{F}_{A,a\cdot l} \subset \mathcal{H}_A^{\otimes a\cdot l}$ *with* $\dim(\mathcal{F}_{A,a\cdot l}) = M_{2,a\cdot l}$ *with*

$$\begin{aligned}\frac{1}{a}\log M_{2,a\cdot l} &\geqslant \inf_{s\in S} I(A\rangle BTX,\sigma_s) - \delta &(3.48)\\ &\geqslant \inf_{s\in S} I(AT\rangle BX,\sigma_s) - \delta &(3.49)\\ &\geqslant \inf_{s\in S} I(A\rangle BX,(\sigma_s^{XAB})^{\otimes l}) - S(T)_\sigma - \delta\\ &\geqslant \inf_{s\in S} I(A\rangle BX,(\sigma_s^{XAB})^{\otimes l}) - \dim(\mathcal{H}_A\otimes\mathcal{H}_B)\log(l+1) - \delta. &(3.50)\end{aligned}$$

The first inequality comes from an application of Lemma 49, second and third from well-known inequalities (see e.g. [Wilde(2017)]) between joint and conditional entropies. We have also used $S(T)_\sigma \leqslant \log|\mathcal{T}(\mathcal{X}\times\mathcal{Y},l)| \leqslant \dim(\mathcal{H}_A\otimes\mathcal{H}_B)\log(l+1)$ *and the marginal state*

$$\begin{aligned}(\sigma_s^{XAB})^{\otimes l} &:= \sum_{x^l\in\mathcal{X}^l} p^l(x^l)\,|e_{x^l}\rangle\langle e_{x^l}| \otimes \sum_\lambda q_{x^l}(\lambda)\mathbb{1}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{N}_s^{\otimes l}(\Phi_{x^l}^\lambda)\\ &= \sum_{x^l\in\mathcal{X}^l} p^l(x^l)\,|e_{x^l}\rangle\langle e_{x^l}| \otimes \mathbb{1}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{N}_s^{\otimes l}(\Psi_{x^l})\\ &= \omega_s^{\otimes l}(\mathcal{N}_s,p,\Psi),\end{aligned}$$

where Ψ_{x^l} *is a purification of* $V^{\otimes l}(x^l)$ *and* $\omega_s(\mathcal{N}_s,p,\Psi)$ *is defined by (4.4). We use* ω_s *to denote this state. From (3.48) we have*

$$\begin{aligned}\frac{1}{l\cdot a}\log M_{2,a\cdot l} &\geqslant \frac{1}{l}\inf_{s\in S} I(A\rangle BX,\omega_s^{\otimes l}) - \frac{\dim(\mathcal{H}_A\otimes\mathcal{H}_B)\log(l+1)}{l} - \frac{\delta}{l}\\ &= \inf_{s\in S} I(A\rangle BX,\omega_s) - \frac{\dim(\mathcal{H}_A\otimes\mathcal{H}_B)\log(l+1)}{l} - \frac{\delta}{l}. &(3.51)\end{aligned}$$

Again from Lemma 49 we have for $\delta > 0$*,*

$$\begin{aligned}\frac{1}{a}\log M_{1,a\cdot l} &\geqslant \inf_{s\in S} I(A;B,\sigma_s) - \delta\\ &= \inf_{s\in S} S(\sum_{x^l}\sum_\lambda p^l(x^l)q_{x^l}(\lambda)\mathcal{N}_s^{\otimes l}(\pi_{x^l}^\lambda)) - \sum_{x^l}\sum_\lambda p^l(x^l)q_{x^l}(\lambda)S(\mathcal{N}_s^{\otimes l}(\pi_{x^l}^\lambda)) - \delta\\ &\geqslant \inf_{s} S(\sum_{x^l}\sum_\lambda p^l(x^l)q_{x^l}(\lambda)\mathcal{N}_s^{\otimes l}(\pi_{x^l}^\lambda)) - \sum_{x^l} p^l(x^l)S(\mathcal{N}_s^{\otimes l}(\sum_\lambda q_{x^l}(\lambda)\pi_{x^l}^\lambda)) - \delta\\ &= \inf_{s\in S} I(A;B,\omega_s^{\otimes l}) - \delta\end{aligned}$$

and hence

$$\frac{1}{a \cdot l} \log M_{1,a\cdot l} \geqslant \inf_{s \in S} I(A; B, \omega_s) - \frac{\delta}{l}. \tag{3.52}$$

For any block-length $n \in \mathbb{N}$, we can write $n = a \cdot l + r$ for $a, l, r \in \mathbb{N}$ and $0 \leqslant r < l$. For all $0 \leqslant r < l$, we use the above $(a \cdot l, M_{1,a\cdot l}, M_{2,a\cdot l})$ CET codes to achieve the desired rate, observing that

$$\liminf_{n\to\infty} \frac{1}{n} M_{i,n} \geqslant \liminf_{a\to\infty} \frac{1}{a \cdot l} M_{i,a\cdot l}, \ \ i = 1, 2.$$

and that $P(\mathcal{C}_{CET}, \mathcal{N}^{\otimes n}, m) \geqslant P(\mathcal{C}_{CET}, \mathcal{N}^{\otimes a\cdot l}, m)$ for all $m \in [M_{a\cdot l}]$.

Proof 53 (Proof of Lemma 48) *According to Lemma 51,*

$$(R_1, R_2) \in \bigcup_{p,\Psi} \bigcap_{s \in S} \hat{C}(\mathcal{N}_s, p, \Psi)$$

implies $(R_1, R_2) \in C_{CET}(\mathcal{J})$. Using standard double-blocking arguments, for each $l \in \mathbb{N}$,

$$(R_1, R_2) \in \bigcup_{l=1}^{\infty} \frac{1}{l} \bigcup_{p,\Psi} \bigcap_{s \in S} \hat{C}(\mathcal{N}_s^{\otimes l}, p, \Psi)$$

implies $(R_1, R_2) \in C_{CET}(\mathcal{J})$.

3.5. Proofs for the arbitrarily varying quantum channel

In this section we consider the task of simultaneous entanglement and classical message transmission in the AVQC model. We derive results for the CET capacities of such channels, when the uncertainty set generating the AVQC is finite. After proving the converse part in the following section, we have used Ahlswede's robustification and elimination techniques to derive suitable codes from compound codes developed so far to prove the direct part of the capacity theorem. Also we will remark the relevant positivity conditions based on results from [Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel].

3.5.1. Proof of converse

In this section, we prove the following lemma.

Lemma 54 *Let $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ with $|S| < \infty$ be an AVQC. We have*

$$\overline{\mathcal{A}}_{r,CET}(\mathcal{J}) \subset \overline{C}_{CET}(\mathrm{conv}(\mathcal{J})).$$

Proof 55 *Let $(\mu_l)_{l=1}^{\infty}$ be a sequence of random codes for AVQC generated by $\mathcal{J}$ with*

$$\liminf_{l\to\infty} \inf_{s^l \in S^l} \int \frac{1}{M_{1,l}} \sum_{m \in [M_{1,l}]} g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)}) \ d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]} = 1 \tag{3.53}$$

with function g_{s^l} defined by (3.5) and $(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]}$ denoting the members of the singleton sets from the respective sigma-algebra. On the other hand, for the compound channel $\mathrm{conv}(\mathcal{J})$ *and each* $\mathcal{N}_q \in \mathrm{conv}(\mathcal{J})$ *we have*

$$\int \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} F(|m\rangle\langle m| \otimes \Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{R} \circ \mathcal{N}_q^{\otimes l} \circ \mathcal{P}^{(m)}(\Phi^{AA}))\ d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]} =$$

$$\sum_{s^l\in S^l} q^l(s^l) \int \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})\ d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]} \geqslant$$

$$\inf_{s^l\in S^l} \int \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})\ d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]} \geqslant 1 - \epsilon_l,$$

with $\epsilon_l \searrow 0$*. The last inequality comes from (3.53). This yields*

$$\inf_{q\in\mathcal{P}(S)} \int \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} F(|m\rangle\langle m| \otimes \Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{R} \circ \mathcal{N}_q^{\otimes l} \circ \mathcal{P}^{(m)}(\Phi^{AA}))d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]}$$
$$\geqslant 1 - \epsilon_l.$$

This means

$$\int \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} F(|m\rangle\langle m| \otimes \Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{R}\circ$$
$$\frac{1}{|\mathcal{P}(S)|} \sum_{q\in\mathcal{P}(S)} \mathcal{N}_q^{\otimes l} \circ \mathcal{P}^{(m)}(\Phi^{AA}))d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]} \geqslant 1 - \epsilon_l,$$

that in turn implies the existence of at least one CET code $(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]}$ *for compound channel* $\mathrm{conv}(\mathcal{J})$ *with average error lower-bounded by* $1 - |\mathcal{P}(S)|\epsilon_l$*. We therefore conclude*

$$\overline{\mathcal{A}}_{r,CET}(\mathcal{J}) \subset \overline{C}_{CET}(\mathrm{conv}(\mathcal{J})).$$

3.5.2. Proof of the direct part

In this section, we prove the following two lemmas, that along with the converse shown in the previous section, prove the first part of Theorem 32.

Lemma 56 *Let* $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ *with* $|S| < \infty$ *be an AVQC. We have*

$$\overline{C}_{CET}(\mathrm{conv}(\mathcal{J})) \subset \overline{\mathcal{A}}_{r,CET}(\mathcal{J}). \tag{3.54}$$

Lemma 57 *Let* $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ *with* $|S| < \infty$ *be an AVQC.* $\overline{\mathcal{A}}_{d,CET}(\mathcal{J}) \neq \{(0,0)\}$ *implies* $\overline{\mathcal{A}}_{d,CET}(\mathcal{J}) = \overline{\mathcal{A}}_{r,CET}(\mathcal{J})$*.*

To prove the second part of Theorem 32, we invoke the following result from [Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel].

Theorem 58 *([Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel] Theorem 40) Let $\mathcal{J} = \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$, $|S| < \infty$, be and AVQC. Then $\mathcal{J}$ is symmetrizable if and only if for all $\{\rho_1, \dots, \rho_M\} \subset \mathcal{S}(\mathcal{H}_A^{\otimes l})$, $M, l \in \mathbb{N}$, $M \geqslant 2$, and POVMs $\{D_m\}_{m=1}^M$ on $\mathcal{H}_B^{\otimes l}$,*

$$\inf_{s^l \in S^l} \frac{1}{M} \sum_{m=1}^{M} (1 - \mathrm{tr}(\mathcal{N}_{s^l}(\rho_m) D_m)) \geqslant 1/4$$

holds.

This result, along with the following lemma, prove the second part of Theorem 32.

Lemma 59 *Let $(\mathcal{P}, \mathcal{R})$ be an (M, l) entanglement transmission code for AVQC $\mathcal{J} = \{\mathcal{N}_s\}_{s\in S} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ with*

$$F(\Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{R} \circ \mathcal{N}_{s^l} \circ \mathcal{P}(\Phi^{AA})) \geqslant 1 - \epsilon \quad \forall s^l \in S^l. \tag{3.55}$$

Then, there exist $\{\rho_1, \dots, \rho_M\} \subset \mathcal{S}(\mathcal{H}_A^{\otimes l})$ and POVM $\{D_m\}_{m\in[M]}$ on $\mathcal{H}_B^{\otimes l}$ such that

$$\frac{1}{M} \sum_{m=1}^{M} \mathrm{tr}(D_m \mathcal{N}_{s^l}(\rho_m)) \geqslant 1 - \epsilon \quad \forall s^l \in S^l \tag{3.56}$$

holds.

Proof 60 *The proof follows directly from the convexity of entanglement fidelity in its first input and that*

$$F(\Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}} \otimes \mathcal{R} \circ \mathcal{N}_{s^l} \circ \mathcal{P}(\Phi^{AA})) = F_e(\pi_{\mathcal{F}_{A,l}}, \mathcal{R} \circ \mathcal{N}_{s^l} \circ \mathcal{P}).$$

Defining for each $m \in [M]$, $D_m := \mathcal{R}_*(|m\rangle\langle m|)$ and $\rho_m := \mathcal{P}(|m\rangle\langle m|)$ with $\mathcal{R}_*$ the Hilbert-Schmidt adjoint of channel $\mathcal{R}$ and spectral decomposition $\pi_{\mathcal{F}_{A,l}} = \frac{1}{M} \sum_{m\in[M]} |m\rangle\langle m|$, we carry the lower bound on (3.55) to (3.56).

Lemma 59 and Theorem 58 show that $\mathcal{J}$ is symmetrizable if and only if there exist no CET codes $(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in M}$ with $M \geqslant 2$, such that we have $\inf_{s^l\in S^l} \frac{1}{M} \sum_{m\in[M]} g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)}) \geqslant \frac{3}{4}$. This in turn implies the second part of Theorem 32.

Proof 61 (Proof of Lemma 56) *In Section 3.4.2 (Lemma 48, Lemma 51), it was shown that for large enough values of $l \in \mathbb{N}$ there exists CET codes $(\tilde{\mathcal{P}}^{(m)}, \tilde{\mathcal{R}}^{(m)})_{m\in[M_{1,]}}$ of $(l, M_{1,l}, M_{2,l})$ for compound channel $\mathrm{conv}(\mathcal{J})$ that achieve the optimum capacity region of this channel $\overline{C}_{CET}(\mathrm{conv}(\mathcal{J}))$ with*

$$\inf_{q\in\mathcal{P}(S)} \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} F(|m\rangle\langle m| \otimes \Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}} \otimes \tilde{R} \circ \mathcal{N}_q^{\otimes l} \circ \mathcal{P}^{(m)}(\Phi^{AA})) \geqslant 1 - \epsilon_l \tag{3.57}$$

with $\epsilon_l \to 0$ exponentially. Since

$$\mathcal{N}_q^{\otimes l} = (\sum_{s\in S} q(s)\mathcal{N}_s)^{\otimes l} = \sum_{s^l \in S^l} q^l(s^l)\mathcal{N}_{s^l},$$

from (3.57) we obtain,

$$\inf_{q\in\mathcal{P}(S)} \sum_{s^l\in S^l} q^l(s^l)\frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} F(|m\rangle\langle m|\otimes\Phi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}}\otimes\tilde{R}\circ\mathcal{N}_{s^l}\circ\mathcal{P}^{(m)}(\Phi^{AA})) \geqslant 1-\epsilon_l. \quad (3.58)$$

Defining the function $f : S^l \to [0,1]$ with

$$f(s^l) := \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} g_{s^l}(\tilde{\mathcal{P}}^{(m)}, \tilde{R}^{(m)}),$$

from (3.58) we obtain

$$\inf_{q\in\mathcal{P}(S)} \sum_{s^l\in S^l} q^l(s^l)f(s^l) \geqslant 1-\epsilon_l. \quad (3.59)$$

Therefore the hypothesis of Ahlswede's robustification (Lemma 205) is satisfied and hence

$$\frac{1}{l!}\sum_{\alpha\in\mathfrak{S}_l} \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} g_{s^l}(\mathcal{U}_{A,\alpha}\circ\tilde{\mathcal{P}}^{(m)}, \tilde{R}^{(m)}\circ\mathcal{U}_{B,\alpha}^{-1}) \geqslant 1-(l+1)^{|S|}\epsilon_l, \quad (3.60)$$

where $\mathcal{U}_{X,\alpha}(\cdot) = U_{X,\alpha}(\cdot)U_{X,\alpha}^\dagger$ with $U_{X,\alpha}$ is a unitary on $\mathcal{H}_A^{\otimes l}$, permuting the tensor factors on this Hilbert space according to α, i.e.

$$U_{X,\alpha}x_1\otimes\cdots\otimes x_l = x_{\alpha(1)}\otimes\ldots x_{\alpha(l)}.$$

Therefore the uniform distribution over the set $\{(\mathcal{P}_\alpha^{(m)}, \mathcal{R}_\alpha^{(m)})_{m\in[M_{1,l}]} : \alpha\in\mathfrak{S}_l\}$ with

$$\mathcal{P}_\alpha^{(m)} := \mathcal{U}_{A,\alpha}\circ\tilde{\mathcal{P}}^{(m)}$$

and

$$\mathcal{R}_\alpha^{(m)} := \tilde{\mathcal{R}}^{(m)}\circ\mathcal{U}_{B,\alpha}^{-1},$$

yield the desired random CET code for arbitrarily varying channel generated by $\mathcal{J}$. Hence we conclude that $(R_1,R_2)\in\overline{C}_{CET}(\mathrm{conv}(\mathcal{J}))$ implies $(R_1,R_2)\in\overline{\mathcal{A}}_{r,CET}(\mathcal{J})$.

To prove Lemma 57, we need the following statement.

Lemma 62 *Let $\mathcal{J} := \{\mathcal{N}_s\}_{s\in S}$ with $|S| < \infty$ be an AVQC, $l\in\mathbb{N}$, μ_l an $(l, M_{1,l}, M_{2,l})$ random CET code for $\mathcal{J}$ with*

$$\inf_{s^l\in S^l}\int \frac{1}{M_{1,l}} \sum_{m\in[M_{1,l}]} g_{s^l}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})d\mu_l(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})_{m\in[M_{1,l}]} \geqslant 1-\epsilon_l \quad (3.61)$$

for a sequence $(\epsilon_l)_{l\in\mathbb{N}}$ *such that* $\epsilon_l \searrow 0$*. Then, for* $\epsilon \in (0,1)$ *and sufficiently large* $l \in \mathbb{N}$*, there exist* l^2 $(l, M_{1,l}, M_{2,l})$ *CET codes* $\{(\mathcal{P}_i^{(m)}, \mathcal{R}_i^{(m)})_{m\in[M_{1,l}]}\}$ *with*

$$\frac{1}{l^2}\sum_{i=1}^{l^2}\frac{1}{M_{1,l}}\sum_{m\in[M_{1,l}]} g_{s^l}(\mathcal{P}_i^{(m)}, \mathcal{R}_i^{(m)}) \geqslant 1-\epsilon \ (\forall s^l \in S^l).$$

Proof 63 *Let for* $K \in \mathbb{N}$*,* $(\Lambda_i^{(m)}, \Gamma_i^{(m)})_{m\in[M_{1,n}]}$ *for* $i = 1,\dots,K$ *be independent random variables with values in* $\mathcal{C}(\mathcal{F}_{A,l}, \mathcal{H}_A^{\otimes l})^{M_{1,l}} \times \Omega_l$ *distributed according to* $\mu_l^{\otimes K}$*. We use the shorthand notation*

$$h_{s^l}(i) := \frac{1}{M_{1,l}}\sum_{m\in[M_{1,l}]} g_{s^l}(\Lambda_i^{(m)}, \Gamma_i^{(m)}).$$

For every $s^l \in S^l$*, an application of Markov's inequality for every* $\epsilon \in (0,1)$ *and* $\gamma > 0$ *yields*

$$\mathbb{P}[1-\frac{1}{K}\sum_{i=1}^{K} h_{s^l}(i) \geqslant \epsilon/2] = \mathbb{P}[2^{K\gamma-\gamma\sum_{i=1}^{K} h_{s^l}(i)} \geqslant 2^{K\gamma(\epsilon/2)}] \leqslant 2^{-K\gamma(\epsilon/2)}\mathbb{E}[2^{\gamma(K-\sum_{i=1}^{K} h_{s^l}(i))}]. \tag{3.62}$$

We now upper-bound the expectation in (3.62).

$$\mathbb{E}[2^{\gamma(K-\sum_{i=1}^{K} h_{s^l}(i))}] = (\mathbb{E}[2^{\gamma(1-h_{s^l}(1))}])^K \leqslant (\mathbb{E}[(1+2^{\gamma}(1-h_{s^l}(1)))])^K \leqslant (1+2^{\gamma}\epsilon_l)^K. \tag{3.63}$$

The second inequality is due to the fact that $(\Lambda_i^{(m)}, \Gamma_i^{(m)})_{m\in[M_{1,n}]}$ *are i.i.d for* $i = 1,\dots,K$*, the first inequality comes from* $2^{\gamma t} \leqslant (1-t)2^{0\cdot\gamma} + t2^{\gamma} \leqslant 1 + 2^{\gamma t}$ *for* $t \in [0,1]$ *and last inequality comes from (3.61). For* $K = l^2$ *and* $\gamma = 2$ *therefore, there exists* $l_0(\epsilon) \in \mathbb{N}$ *such that for* $l \geqslant l_0(\epsilon)$

$$(1+2^{\gamma}\epsilon_l)^{l^2} \leqslant 2^{l^2(\epsilon/2)}. \tag{3.64}$$

Therefore we obtain from (3.62), (3.63) and (3.64),

$$\mathbb{P}[1-\frac{1}{l^2}\sum_{i=1}^{l^2} h_{s^l}(i) \geqslant \epsilon/2] \leqslant 2^{-l^2(\epsilon/2)}.$$

Applying the union bound on the last inequality yields

$$\begin{aligned}&\mathbb{P}[\frac{1}{l^2}\sum_{i=1}^{l^2} h_{s^l}(i) > 1-\epsilon/2, \forall s^l \in S^l]\\ &\qquad \geqslant 1 - |S|^l 2^{-l^2(\epsilon/2)},\end{aligned}$$

which implies that there is a realization $(\mathcal{P}_i^{(m)}, \mathcal{R}_i^{(m)})_{m\in[M_{1,l}]}, i = 1, \dots l^2$ *with*

$$\frac{1}{l^2}\sum_{i=1}^{l^2}\frac{1}{M_{1,l}}\sum_{m\in[M_{1,l}}g_{s^l}(\mathcal{P}_i^{(m)}, \mathcal{R}_i^{(m)}) > 1-\epsilon/2 \;\; \forall s^l \in S^l,$$

when $|S|^l 2^{-l^2(\epsilon/2)} < 1$ *which is possible for sufficiently large values of* l.

Proof 64 (Proof of Lemma 57) *By assumption, for* $\epsilon \in (0,1)$ *there exists a* $(r_l, l^2, 1)$ *deterministic CET code* $(\tilde{\mathcal{P}}^{(m)}, \tilde{\mathcal{R}}^{(m)})_{m=1}^{l^2}$ *with*

$$\frac{1}{l^2}\sum_{m=1}^{l^2} g_{s^{r_l}}(\tilde{\mathcal{P}}^{(m)}, \tilde{\mathcal{R}}^{(m)}) \geqslant 1-\epsilon \;\forall s^{r_l} \in S^{r_l}, \tag{3.65}$$

with $r_l = o(l)$. *This is because if the capacity region is not equal to the point* $(0,0)$, R_1 *(intersection of the capacity region with the* x*-axis), is definitely larger than zero (see Lemma 59). On the other hand, let* $(R_1, R_2) \in \overline{\mathcal{A}}_{r,CET}$. *By Lemma 62, this implies the existence of* l^2 $(l, M_{1,l}, M_{2,l})$ *CET codes* $\{\hat{\mathcal{P}}_i^{(m)}, \hat{\mathcal{R}}_i^{(m)} : i \in [l^2]\}$ *of the same rate with*

$$\frac{1}{l^2}\sum_{i=1}^{l^2}\frac{1}{M_{1,l}}\sum_{m=1}^{M_{1,l}} g_{s^l}(\hat{\mathcal{P}}_i^{(m)}, \hat{\mathcal{R}}_i^{(m)}) \geqslant 1-\epsilon \;\forall s^l \in S^l. \tag{3.66}$$

Define CPTP maps

$$\mathcal{P}^{(m)}(a\otimes b) := \frac{1}{l^2}\sum_{i=1}^{l^2}\tilde{\mathcal{P}}^{(i)}(a)\otimes\hat{\mathcal{P}}_i^{(m)}(b),$$

$$\mathcal{R}^{(m)}(c\otimes d) := \sum_{i=1}^{l^2}\tilde{\mathcal{R}}^{(i)}(c)\otimes\hat{\mathcal{R}}_i^{(m)}(d).$$

We have

$$\begin{aligned}
&\frac{1}{M_{1,l}}\sum_{m=1}^{M_{1,l}} g_{s^{r_l+l}}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)}) = \\
&\qquad \frac{1}{l^2\cdot M_{1,l}}\sum_{m=1}^{M_{1,l}} F(\varPhi^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes r_l+l}}\otimes\sum_{i=1}^{l^2}\tilde{\mathcal{R}}^{(i)}\otimes\hat{\mathcal{R}}_i^{(m)}\circ\mathcal{N}_{s^{r_l}}\otimes\mathcal{N}_{s^l}\circ\sum_{j=1}^{l^2}\tilde{\mathcal{P}}^{(j)}\otimes\hat{\mathcal{P}}_i^{(m)}(\varPhi^{AA})) \\
&\qquad \geqslant \frac{1}{l^2}\sum_{i=1}^{l^2}\frac{1}{M_{1,l}}\sum_{m=1}^{M_{1,l}} F(\tilde{\varPhi}^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes r_l}}\otimes\tilde{\mathcal{R}}^{(i)}\circ\mathcal{N}_{s^{r_l}}\circ\tilde{\mathcal{P}}^{(j)}(\tilde{\varPhi}^{AA})) \\
&\qquad \times F(\hat{\varPhi}^{AB}, \mathrm{id}_{\mathcal{H}_A^{\otimes l}}\otimes\hat{\mathcal{R}}_i^{(m)}\circ\otimes\mathcal{N}_{s^l}\circ\hat{\mathcal{P}}_i^{(m)}(\hat{\varPhi}^{AA})),
\end{aligned} \tag{3.67}$$

where $\tilde{\varPhi}^{XY}$ *and* $\hat{\varPhi}^{XY}$ *are maximally entangled states. The inequality above is due to the fact that* $g_{s^{r_l+l}}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)})$ *is non-negative for all* m *and* s^{l+r_l}. *Applying Lemma 206 on*

(3.67), given (3.65) and (3.66) we conclude

$$\frac{1}{M_{1,l}} \sum_{m=1}^{M_{1,l}} g_{s^{r_l+l}}(\mathcal{P}^{(m)}, \mathcal{R}^{(m)}) \geqslant 1 - 2\epsilon.$$

As $r_l = o(l)$, this implies $(R_1, R_2) \in \overline{\mathcal{A}}_{d,CET}(\mathcal{J})$. This in turn implies $\overline{\mathcal{A}}_{r,CET}(\mathcal{J}) \subset \overline{\mathcal{A}}_{d,CET}(\mathcal{J})$. As the inclusion $\overline{\mathcal{A}}_{d,CET}(\mathcal{J}) \subset \overline{\mathcal{A}}_{r,CET}(\mathcal{J})$ is obvious, we are done.

3.6. Simultaneous classical message and entanglement transmission over fully quantum AVCs

In this section, we consider simultaneous transmission of classical messages and entanglement over an an arbitrarily varying quantum channel with a *quantum jammer.* Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A \otimes \mathcal{H}_J, \mathcal{H}_B)$ be a quantum channel whose input space is a tensor product of a Hilbert space $\mathcal{H}_A$ (the legitimate sender's space) and a Hilbert space $\mathcal{H}_J$ which is under control of a quantum jammer. We consider a situation, where for each given block-length n, the jammer may choose any state η on $\mathcal{H}_J^{\otimes n}$ as input in order to disturb the transmission of the legitimate parties.

The *Arbitrarily Varying Quantum Channel (AVQC)* generated by $\mathcal{N}$ is given by the family

$$\mathcal{N}_{n,\sigma}(\cdot) := \mathcal{N}^{\otimes n}(\cdot \otimes \sigma) : \sigma \in \mathcal{S}(\mathcal{H}_J^{\otimes n}), n \in \mathbb{N}\Big\} \tag{3.68}$$

of CPTP maps[4]. The above channel model already has been under consideration in case of univariate transmission goals. Karumanchi et al. [Karumanchi et al.(2016)Karumanchi, Mancini, Winter, and Yang] utilized the postselection technique from [Christandl et al.(2009) Christandl, König, and Renner] to derive correlated random codes for the AVQC from good codes for the compound channel generated by $\mathfrak{I} := \{\mathcal{N}_\sigma := \mathcal{N}(\cdot, \sigma) : \sigma \in \mathcal{S}(\mathcal{H}_J)\}$. This approach turned out to be successful to determine the random entanglement transmission capacity for the AVQC. In recent work [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter], the above mentioned techniques were used to also characterize the random classical message transmission capacity of the AVQC. Going beyond, the authors of [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] introduced a derandomization technique to derive a dichotomy for the entanglement and classical message transmission capacities of the QAVC. *The deterministic capacity is zero or it equals the random capacity.* We show, that the ideas of the mentioned works together with the results derived in this chapter are sufficient to determine the random capacity and establish a partial characterization of the deterministic capacity in terms of a dichotomy also in case of simultaneous transmission of entanglement and classical messages.

[4] Although acronym "AVQC" is also used for the somewhat more restrictive channel model introduced in Section 3.2.2, it should be apparent from context, which of these models is considered.

The definitions for the corresponding capacity regions can be easily extrapolated from the corresponding definitions in Section 3.2.2 using the set of transmission maps in (3.68). We denote the *random CET capacity region* of $\mathcal{N}$ by $\overline{\mathcal{A}}_{r,CET}(\mathcal{N})$ and the *deterministic CET capacity* by $\overline{\mathcal{A}}_{d,CET}(\mathcal{N})$. First, we give a characterization of the random CET capacity $\overline{\mathcal{A}}_{r,CET}(\mathcal{N})$ of the AVQC with fully quantum jammer.

Theorem 65 *Let $\mathcal{N} \in \mathcal{C}(\mathcal{H}_A \otimes \mathcal{H}_J, \mathcal{H}_B)$, and $\mathfrak{I} := \{\mathcal{N}_\sigma : \ \sigma \in \mathcal{S}(\mathcal{H}_J)\}$. It holds*

$$\overline{\mathcal{A}}_{r,CET}(\mathcal{N}) \ = \ \overline{C}_{CET}(\mathfrak{I}) \tag{3.69}$$

The $\supset$ inclusion in (65) is obvious. To show the reverse inclusion, we will invoke the " robustification " statement in Proposition 67 below. In the derivations, the following representation of the permutation group $\mathfrak{S}_n$ on n-fold tensor product spaces plays a key role. Let for each $\pi \in \mathfrak{S}_n$, U_π be the unitary exchanging the factors in $\mathcal{H}^{\otimes n}$, i.e.

$$U_\pi \ x_1 \otimes \cdots \otimes x_n = x_{\pi(1)} \otimes \cdots \otimes x_{\pi(n)}$$

for each $x_1, \dots, x_n \in \mathcal{H}$. We set $\mathcal{U}_\pi(\cdot) := U_\pi(\cdot)U_\pi^*$. In $\mathcal{U}_{A,\pi}$, $\mathcal{U}_{B,\pi}$,$\mathcal{U}_{J,\pi}$ denote the corresponding maps performed on the subsystems under control of A, B, J accordingly. A rather powerful result for states being invariant under permutations of the tensor factors is the following.

Proposition 66 (de Finetti reduction [Christandl et al.(2009)Christandl, König, and Renner]) *Let $\rho \in \mathcal{S}(\mathcal{H}^{\otimes n})$ permutation invariant, i.e. $\mathcal{U}_\pi(\rho) = \rho$ for each $\pi \in \mathfrak{S}_n$. It holds*

$$\rho \ \leqslant \ (n+1)^{(\dim \mathcal{H})^2} \int \sigma^{\otimes n} d\mu(\sigma)$$

with a probability measure μ.

Proposition 67 *Let $\mathcal{C} := (\mathcal{P}_m, \mathcal{R}_m)_{m=1}^{M_1}$ be an (n, M_1, M_2)-CET code such that with $\lambda \in (0,1)$*

$$\inf_{\sigma \in \mathcal{S}(\mathcal{H}_J)} \overline{P}_{CET}(\mathcal{C}, \mathcal{N}_\sigma^{\otimes n})) \ \geqslant 1 - \lambda$$

holds. With $\mathcal{C}_\pi := (\mathcal{U}_{A,\pi} \circ \mathcal{P}_m, \mathcal{R}_m \circ \mathcal{U}_{B,\pi^{-1}})$ for each $\pi \in \mathfrak{S}_n$, it holds

$$\inf_{\tau \in \mathcal{S}(\mathcal{H}_J^{\otimes n})} \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \overline{P}_{CET}(\mathcal{C}_\pi, \mathcal{N}_{n,\tau}) \ \geqslant 1 - (n+1)^{(\dim \mathcal{H}_J)^2} \cdot \lambda.$$

Proof 68 *The proof closely follows the lines of [Karumanchi et al.(2016)Karumanchi, Mancini, Winter, and Yang]. Set $d_J := \dim \mathcal{H}_J$. By permutation invariance of $\mathcal{N}^{\otimes n}$, the equality*

$$\mathcal{U}_{B,\pi^{-1}} \circ \mathcal{N}^{\otimes n} \circ (\mathcal{U}_{A,\pi} \otimes \mathrm{id}_{\mathcal{H}_J}^{\otimes n}) = \mathcal{N}^{\otimes n} \circ (\mathrm{id}_{\mathcal{H}_A}^{\otimes n} \otimes \mathcal{U}_{J,\pi^{-1}}) \tag{3.70}$$

holds for each permutation $\pi \in \mathfrak{S}_n$. Using (3.70) together with the fact, that $\overline{P}_{CET}$ is an affine function of the channel, we obtain

$$\frac{1}{n!}\sum_{\pi\in\mathfrak{S}_n} \overline{P}_{CET}(\mathcal{C}_\pi, \mathcal{N}_{n,\tau}) \ = \ \overline{P}_{CET}(\mathcal{C}, \mathcal{N}_{n,\overline{\tau}})$$

for each $\tau \in \mathcal{S}(\mathcal{H}_J^{\otimes n})$, where $\overline{\tau} := 1/n! \ \sum_{\pi\in\mathfrak{S}_n} \mathcal{U}_\pi(\tau)$. Define $\mathcal{T}_*$ to be the Hilbert-Schmidt adjoint of the map

$$\sigma \ \mapsto \ \frac{1}{M_1}\sum_{m=1}^{M_1} \mathrm{id} \otimes \mathcal{R}_m \circ \mathcal{N}_{n,\sigma} \circ \mathcal{P}_m(\Phi).$$

We write

$$1 - \overline{P}_{CET}(\mathcal{C}, \mathcal{N}_{n,\tau}) \ = \ \mathrm{tr} X\tau, \tag{3.71}$$

with the matrix $X := \mathbb{1} - \mathcal{T}_*(\Phi)$ (note that $0 \leqslant X \leqslant \mathbb{1}$ holds.) Using Proposition 66 together with linearity and monotonicity of the integral, we have

$$\begin{aligned}\mathrm{tr} X\overline{\tau} \ &\leqslant \ (n+1)^{d_J^2} \int \mathrm{tr} X\sigma^{\otimes n} \ d\mu(\sigma) \\ &\leqslant (n+1)^{d_J^2} \sup_{\sigma\in\mathcal{S}(\mathcal{H})} \mathrm{tr} X\sigma^{\otimes n}. \\ &\leqslant (n+1)^{d_J^2} \cdot \lambda.\end{aligned}$$

Which is, by (3.71), the desired bound.

Proof 69 (Proof of Theorem 65 (Direct part)) *The statement $\overline{C}_{CET}(\mathfrak{I}) \subset \overline{\mathcal{A}}_{r,CET}(\mathcal{N})$ directly follows from combining the results from Section 3.4.2 (Lemma 48 and Lemma 51) with Proposition 67. Let $(\mathcal{C}_n)_{n=1}^{\infty}$ be a sequence of $(n, M_{1,n}, M_{2,n})$-CET codes with*

$$\inf_{\sigma\in\mathcal{S}(\mathcal{H}_J)} \overline{P}_{CET}(\mathcal{C}_n, \mathcal{N}_\sigma^{\otimes n}) \geqslant 1 - 2^{-nc}$$

with a constant $c > 0$ for each large enough n. Let $\tilde{\mu}_n$ be the uniform distribution on $\mathfrak{S}_n$, and $f(\pi) := \mathcal{C}_{n,\pi}$. Then $\tilde{\mu}_n \circ f^{-1}$ is an $(n, M_{1,n}, M_{2,n})$ random CET code, such that

$$\mathbb{E}\left[\inf_{\sigma\in\mathcal{S}(\mathcal{H}_J)} \overline{P}_{CET}(\cdot, \mathcal{N}_{n,\sigma})\right] \ \geqslant 1 - (n+1)^{d_J^2} \cdot 2^{-nc}.$$

Since the right hand side of the above inequality tends to one for $n \to \infty$, every rate pair (R_1, R_2) being achievable for the compound channel $\mathfrak{I}$, is also achievable by random codes for the AVQC $\mathcal{N}$.

Next we show, using a derandomization technique introduced in [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter], the following statement.

Theorem 70 (Dichotomy for $\overline{\mathcal{A}}_{d,CET}$) *$\overline{\mathcal{A}}_{d,CET}(\mathcal{N})$ equals $\{(0,0)\}$ or $\overline{\mathcal{A}}_{r,CET}(\mathcal{N})$*

Remark 71 *The above statement quantifies the deterministic capacity region of the AVQC up to a blind spot. It is an open question whether or not there are channels for which $\overline{\mathcal{A}}_{d,CET}(\mathcal{N}) = \{(0,0)\}$ and $\{(0,0)\} \subsetneq \overline{\mathcal{A}}_{r,CET}(\mathcal{N})$ does happen.*

Remark 72 *For a Hermitian matrix $A \in \mathcal{L}(\mathcal{H})$, and $\alpha > 0$, it holds $A \leqslant \alpha \mathbb{1}$ if and only if $\mathrm{tr}\sigma A \ \leqslant \alpha$ for all $\sigma \in \ \mathcal{S}(\mathcal{H})$.*

Proposition 73 ([Ahlswede and Winter(2002)], Theorem 19) *Let $X_1, \ldots, X_T$ be i.i.d. hermitian random matrices with $0 \leqslant X_i \leqslant \mathbb{1}$ a.s. for all $i \in [T]$, and $\mathbb{E}X_1 \ \leqslant \ m\mathbb{1} \leqslant a\mathbb{1} \leqslant A$. Then*

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T} X_t \ \geqslant a\mathbb{1}_{\mathcal{H}}\right) \ \leqslant \ \dim \mathcal{H} \cdot \exp(-T2(a-m)^2)$$

Proof 74 (Proof of Theorem 70) *We consider the non-trivial case $\overline{\mathcal{A}}_{d,CET} \ \neq \{(0,0)\}$. Let $(R_1, R_2) \in \overline{\mathcal{A}}_{r,CET}(\mathcal{N})\backslash\{(0,0)\}$. We aim to show that (R_1, R_2) is also achievable with deterministic codes. Since $\overline{\mathcal{A}}_{d,CET} \neq \{(0,0)\}$, we find, for each large enough blocklength n an $(n, \tilde{M}_1, \tilde{M}_2)$-CET code $\mathcal{C}^{(1)} := (\mathcal{P}_m^{(1)}, \mathcal{R}_m^{(1)})_{m=1}^{\tilde{M}_1}$ with $\tilde{M}_1 \geqslant 2^{l\tilde{R}}$, where $\tilde{R} > 0$ is a constant, and*

$$\inf_{\sigma \in \mathcal{S}(\mathcal{H}_J^{\otimes l})} \overline{P}_{CET}(\mathcal{C}^{(1)}, \mathcal{N}_{n,\sigma}) \geqslant 1 - \epsilon_l \tag{3.72}$$

with $\epsilon_n \to 0$ for $n \to \infty$. Set for each n, $a_n := \lceil 2\log n/\tilde{R}\rceil$, and $b_n := n - a_n$, i.e. $n = a_n + b_n$. If n is large enough, we have a random (n, M_1, M_2)-CET code μ_{b_n} such that $\frac{1}{b_n}\log M_i \geqslant R_i - \delta$, for $i = 1, 2$, and

$$\mathbb{E}_{\mu_{b_n}} \inf_{\sigma \in \mathcal{S}(\mathcal{H}_J^{\otimes b_n})} \overline{P}_{CET}(\cdot, \mathcal{N}_{n,\sigma}) \ \geqslant 1 - 2^{-b_n c}. \tag{3.73}$$

For simplicity, we assume μ_{b_n} to be finitely supported on $\{\mathcal{C}_1, \ldots, \mathcal{C}_{T'}\}$ (which is possible by the explicit construction of a finite random code in Proposition 67). Note, that we can write

$$1 - \overline{P}_{CET}(\mathcal{C}_t, \mathcal{N}_{b_n,\sigma}) \ = \ \mathrm{tr}E_t\sigma \tag{3.74}$$

with a matrix $0 \leqslant E_t \leqslant \mathbb{1}$ for each $t \in [T']$. By (3.73) and (3.74), together with linearity of expectation, it holds

$$\inf_{\sigma \in \mathcal{H}_J^{\otimes b_n}} \mathrm{tr}\overline{E}\sigma \ = \ \mathbb{E}_{\mu_{b_n}}\left[1 - \inf_{\sigma \in \mathcal{H}_J^{\otimes b_n}} \overline{P}_{CET}(\cdot, \mathcal{N}_{n,\sigma})\right] \ \leqslant 2^{-b_n c}, \tag{3.75}$$

where we defined $\overline{E} := \mathbb{E}_{\mu_{b_n}} E_t$. By Fact 72, combined with the bound in (3.75), $\overline{E} \leqslant$

$2^{-nc}\mathbb{1}$. *Let* $X^{(1)}, \dots X^{(\tilde{M}_1)}$ *be i.i.d. random matrices, each distributed according to* μ_{b_n}. *By Proposition 73, It holds*

$$\mathbb{P}\left(\frac{1}{\tilde{M}_1}\sum_{t=1}^{\tilde{M}_1} E_t \geqslant \left(2^{-nc} + \frac{1}{n}\right)\mathbb{1}\right) \leqslant d_J^{b_n} \exp\left(-\tilde{M}_1/n^2\right). \tag{3.76}$$

By our choice of a_n, *the RHS of (3.76) is strictly smaller than one for each large enough* n. *Therefore, we find* $\mathcal{C}_1, \dots, \mathcal{C}_{\tilde{M}_1}$ *such that*

$$1 - \frac{1}{\tilde{M}_1}\sum_{t=1}^{\tilde{M}_1} \overline{P}_{CET}(\mathcal{C}_t, \mathcal{N}_{n,\sigma}) \leqslant 2^{-b_n c} + \frac{1}{n} := \gamma_n$$

holds. Let $\mathcal{C}_t = (\mathcal{P}_{t,m}^{(2)}, \mathcal{R}_{t,m}^{(2)})_{m=1}^{M_1}$. *We define an* (n, M_1, M_2) *deterministic CET code* $\mathcal{C} = (\mathcal{P}_m, \mathcal{R}_m)_{m=1}^{M_1}$ *with*

$$\mathcal{P}_m := \frac{1}{\tilde{M}_1}\sum_{t=1}^{\tilde{M}_1} \mathcal{P}_t^{(1)}(\pi_1) \otimes \mathcal{P}_{t,m}^{(2)}, \quad \textit{and} \quad \mathcal{R}_m := \operatorname{tr}_{\mathcal{H}^{\otimes a_n}} \circ \sum_{t=1}^{\tilde{M}_1} \mathcal{R}_t^{(1)} \otimes \mathcal{R}_{t,m}^{(2)}$$

To evaluate the fidelity of the above code, we notice, that for each $\sigma \in \mathcal{S}(\mathcal{H}_J^{\otimes n}), t \in [\tilde{M}_1], m \in [M_1]$

$$F(\Phi_1 \otimes \Phi_2, \mathrm{id} \otimes \circ \mathcal{R}_t^{(1)} \otimes \mathcal{R}_{t,m}^{(2)} \circ \mathcal{N}_{n,\sigma} \circ \mathcal{P}_t^{(1)} \otimes \mathcal{P}_{t,m}^{(2)}(\Phi_1 \otimes \Phi_2)) = \operatorname{tr} F_t^{(1)} \otimes F_{t,m}^{(2)} \sigma \tag{3.77}$$

holds with effects $F_t^{(1)}$, $F_{t,m}^{(2)}$. *This is advantageous, since*

$$\frac{1}{\tilde{M}_1}\sum_{t=1}^{\tilde{M}_1} \operatorname{tr} F_t^{(1)} \tau = \overline{P}_{CET}(\mathcal{C}^{(1)}, \mathcal{N}_{a_n,\tau}) \geqslant 1 - \epsilon_n,$$

and

$$\frac{1}{M_1}\frac{1}{\tilde{M}_1}\sum_{t=1}^{\tilde{M}_1}\sum_{m=1}^{M_1} \operatorname{tr} F_{t,m}^{(2)} \tau = \overline{P}_{CET}(\mathcal{C}_t^{(2)}, \mathcal{N}_{b_n,\tau}) \geqslant 1 - \gamma_n.$$

We have for each $\sigma \in \mathcal{S}(\mathcal{H}_J^{\otimes n})$

$$\begin{aligned}
&\overline{P}_{CET}(\mathcal{C}, \mathcal{N}_{n,\sigma}) \\
&= \frac{1}{M_1 \tilde{M}_1} \sum_{t,t'=1}^{\tilde{M}_1} \sum_{m=1}^{M_1} F\left(\Phi_2, \mathrm{id}_{\mathcal{H}_J^{\otimes b_n}} \otimes \mathrm{tr}_{\mathcal{H}_J^{\otimes a_n}} \circ \mathcal{R}_{t'}^{(1)} \otimes \mathcal{R}_{t',m}^{(2)} \circ \mathcal{N}_{n,\sigma} \circ \mathcal{P}_t^{(1)}(\pi_1) \otimes \mathcal{P}_{t,m}^{(2)}(\Phi_2)\right) \\
&\geqslant \frac{1}{M_1 \tilde{M}_1} \sum_{t,t'=1}^{\tilde{M}_1} \sum_{m=1}^{M_1} F\left(\Phi_1 \otimes \Phi_2, \mathrm{id}_{\mathcal{H}_J^{\otimes b_n}} \otimes \mathcal{R}_{t'}^{(1)} \otimes \mathcal{R}_{t',m}^{(2)} \circ \mathcal{N}_{n,\sigma} \circ \mathcal{P}_t^{(1)} \otimes \mathcal{P}_{t,m}^{(2)}(\Phi_1 \otimes \Phi_2)\right) \\
&\geqslant \frac{1}{M_1 \tilde{M}_1} \sum_{t=1}^{\tilde{M}_1} \sum_{m=1}^{M_1} F\left(\Phi_1 \otimes \Phi_2, \mathrm{id}_{\mathcal{H}_J^{\otimes b_n}} \otimes \mathcal{R}_{t}^{(1)} \otimes \mathcal{R}_{t,m}^{(2)} \circ \mathcal{N}_{n,\sigma} \circ \mathcal{P}_t^{(1)} \otimes \mathcal{P}_{t,m}^{(2)}(\Phi_1 \otimes \Phi_2)\right) \\
&= \frac{1}{M_1 \tilde{M}_1} \sum_{t=1}^{\tilde{M}_1} \sum_{m=1}^{M_1} \mathrm{tr} F_t^{(1)} \otimes F_{t,m}^{(2)} \sigma. \qquad (3.78)
\end{aligned}$$

The first inequality above is by monotonicity of the fidelity under CPTP maps. The last equality is from (3.77). Now, let σ_1 *be the marginal of* σ *on the first* a_n *tensor factors of* $\mathcal{H}_J^{\otimes n}$*, and* σ_2 *the marginal on the last* b_n *tensor factors.*

$$A \otimes B \ \geqslant\ \mathbb{1} \otimes \mathbb{1} - \mathbb{1} \otimes (\mathbb{1} - B) - (\mathbb{1} - A) \otimes \mathbb{1}$$

which holds for any two matrices $0 \leqslant A, B \leqslant \mathbb{1}$*. We have*

$$\mathrm{tr} F_t^{(1)} \otimes F_{t,m}^{(2)} \sigma \ \geqslant 1 - \ \mathrm{tr}(\mathbb{1} - F_t^{(1)})\sigma_1 - \mathrm{tr}(\mathbb{1} - F_{t,m}^{(2)})\sigma_2 \qquad (3.79)$$

Combining (3.78), and (3.79), we can bound

$$\overline{P}_{SET}(\mathcal{C}, \mathcal{N}_{n,\sigma}) \ \geqslant\ P(\mathcal{C}^{(1)}, \mathcal{N}_{a_n,\sigma_1}) + \frac{1}{\tilde{M}_1} \sum_{t=1}^{\tilde{M}1} P(\mathcal{C}_t^{(2)}, \mathcal{N}_{b_n,\sigma_2}) - 1$$

Minimizing over all states on $\mathcal{H}_J^{\otimes n}$*, we obtain*

$$\inf_{\sigma \in \mathcal{S}(\mathcal{H}_J^{\otimes n})} \overline{P}_{SET}(\mathcal{C}, \mathcal{N}_{n,\sigma}) \ \geqslant 1 - \epsilon_n - \gamma_n.$$

The right hand side approaches one for $n \to \infty$*. Since also* $\frac{a_n}{n} \to 0$ *and* $\frac{b_n}{n} \to 1$ *for* $n \to \infty$*, it is clear, that we achieve* (R_1, R_2) *with the codes defined.*

4. Universal superposition codes:capacity regions for quantum broadcast channel

In this chapter we derive universal codes for transmission of broadcast and confidential messages over classical-quantum-quantum and fully quantum channels. These codes are robust to channel uncertainties considered in the compound model. To construct these codes we generalize random codes for transmission of public messages, to derive a universal superposition coding for the compound quantum broadcast channel. As an application, we give a multi-letter characterization of regions corresponding to capacity of the compound quantum broadcast channel for transmitting broadcast and confidential messages simultaneously. This is done for two types of broadcast messages, one called public and the other common.

4.1. Introduction

In this chapter we consider the compound quantum broadcast channel, connecting one sender to two receivers of different permissions or priorities. The channel is used to perform an integrated task, in which a confidential message, kept secret from the third party, is communicated simultaneously with a broadcast message available to both receivers. The requirements on the broadcast message, determine two communication scenarios. In the first scenario, we consider the case where both receivers are required to decode the broadcast message. We refer to this message as the common message. In the second scenario the decoding condition is relaxed on one of the receivers. That is, the third party, namely the receiver from whom the confidential message is kept secret, may or may not decode the broadcast message, to which, in this scenario, we refer as the public message. The capacity of the channel for performing such tasks, will include trade-off regions, determining the resourcefulness of the public/common message transmission capacity, for enhancement of confidential message transmission. Information theoretic analysis of these tasks, will naturally be significant when regions beyond those achieved by simple time-sharing between the two tasks are achieved. We first consider the case where the sender is restricted to classical inputs, namely the classical-quantum-quantum (cqq) broadcast model. This model proves useful for obtaining capacity results for the fully quantum broadcast model, where this restriction is lifted.

The classical counterparts of our results were given in [Schaefer and Boche(2014b)]. Therein, the authors first derive robust codes for the bidirectional channel, in which both receivers are meant to decode the message. This common message will then piggyback a public message decoded by Bob. The privacy amplification strategies are then applied on part of the public codes to obtain information theoretic security via equivocation. We follow a similar approach in the context of quantum information theory. We obtain codes for the bidirectional channel (broadcast channel with no security requirement) by generalizing the random codes from [Mosonyi(2015)]. Our generalization of these results (see Appendix D), yields a universal superposition coding for cq channels. Our input structure allows us to use privacy amplification arguments ([Boche et al.(2014)Boche, Cai, Cai, and Deppe]) on part of the codebook to achieve the desired secrecy rates.

The quantum broadcast model in which the channel is assumed to perfectly known by communicating parties was considered in [Hsieh and Wilde(2009), Wilde and Hsieh(2011b)], with and without a pre-shared secret key respectively. Therein, the authors have established a dynamic capacity trade off region using a coding strategy that is channel-dependent. We use a different strategy in which establish universal superposition codes for the compound bidirectional channel, exploiting properties of Renyi entropies.

Another regime in which the quantum broadcast model with confidential messages has been studied, is the one-shot (single serving) model. A one-shot dynamic capacity theorem was derived for regions corresponding to tasks of common, public and private message transmission over the quantum channel in [Salek et al.(2020)Salek, Anshu, Hsieh, Jain, and Fonollosa]. It would be interesting to see if the coding strategies used therein, derived from position based decoding (see [Anshu et al.(2017)Anshu, Devabathini, and Jain, ANSHU(2018), Anshu et al.(2019a)Anshu, Jain, and Warsi]), can be used to design codes for the compound channel model.

Precise definitions of channel models, codes and rate regions along with our main results for the cqq model are given in Section 4.2. We prove the direct part of our capacity results for the cqq model in Section 4.3, that is followed by the proof of converse in Section 4.4. The security criterion that we impose on the confidential message, is the mutual information between Alice and Eve to be arbitrarily small for large numbers of channel uses. As the common or indeed the public messages are available to Eve, we require the mentioned mutual information to be conditioned on the broadcast message. Proving the existence of capacity achieving codes is done in two steps. First we consider the case where there is no security criterion placed on the messages sent to Bob and Eve. In this case, we have a bidirectional channel, where Alice, is sending a message to be decoded by Bob and potentially by Eve (weather Eve decodes this message depends on which scenario is considered, determining in turn our labeling of it as common or public). Conditioned on this message (the corresponding codewords are distributed according to a certain structure), Alice is simultaneously transmitting a second type of message, that is decoded by Bob. The random coding that makes precisely this task possible, is given by Lemma 84, which is our universal superposition coding result. Application of this lemma gives us the desired bidirectional codes in forms of Lemma 90 (where the conditioning message is common) and 96 (where the conditioning message is public). In the second step, the second type of message described above, is used for privacy amplification. We give the code definitions and capacity results for the fully quantum channel independently in Section 4.5.

4.2. Basic definitions and main results

In this section we state the main results and definitions for the compound classical-quantum-quantum (cqq) broadcast channel. The results and definitions related to the fully quantum broadcast channel are stated in Section 4.5. For finite alphabet $\mathcal{X}$ and Hilbert spaces $\mathcal{H}_B, \mathcal{H}_E$, let $\mathcal{W} := \{W_s\}_{s \in S} \subset CQ(\mathcal{X}, \mathcal{H}_B \otimes \mathcal{H}_E)$ be a set of cqq channels. The compound cqq broadcast channel generated by this set is given by family $\{W_s^{\otimes n}, s \in S\}_{n=1}^{\infty}$. In other words, using n instances of the compound channel is equivalent to using n instances of one of the channels from the uncertainty set. The users of this channel may or may not have access to the Channel State Information (CSI). In this document, we consider the case where both users only know the uncertainty set, to which the actual channel belongs. We consider two closely related communication scenarios of significance, having both appeared in the literature hitherto.

- **Broadcasting Common and Confidential messages (BCC)**, where the compound channel is used $n \in \mathbb{N}$ times by the sender Alice in control of the input of the channel, to send two types of messages (m_0, m_c) simultaneously over the channel.
  - $m_0 \in [M_{0,n}]$, called the common message, that has to be reliably decoded by receiver Bob in control of Hilbert space $\mathcal{H}_B$ and Eve in control of Hilbert space $\mathcal{H}_E$.
  - $m_c \in [M_{c,n}]$, called the confidential message, that has to be decoded reliably by Bob while Eve, the wiretapper, is kept ignorant.
- **Transmitting Public and Confidential messages (TPC)**, where along with the confidential message $m_c \in [M_{0,n}]$ and instead of the common message, Alice wishes to send a "public" message $m_1 \in [M_{1,n}]$, that is reliably decoded by Bob while it may or may not be decoded by Eve.

We consider the main concepts and results related to each task in the following. We start with the BCC scenario. The precise definition of the BCC codes is given by the following.

Definition 75 (BCC codes) *An $(n, M_{0,n}, M_{c,n})$ BCC code for $\mathcal{W}$, is a family $\mathcal{C} = (E(\cdot|\mathbf{m}), D_{B,\mathbf{m}}, D_{E,m_0})_{\mathbf{m}\in\mathbf{M}}$ with $\mathbf{M} := [M_{0,n}] \times [M_{c,n}]$, stochastic encoder $E : \mathbf{M} \to \mathcal{P}(\mathcal{X}^n)$, POVMs $(D_{B,\mathbf{m}})_{\mathbf{m}\in\mathbf{M}}$ on $\mathcal{H}_B^{\otimes n}$ and $(D_{E,m_0})_{m_0\in[M_{0,n}]}$ on $\mathcal{H}_E^{\otimes n}$.*

We define the transmission error functions, for any cqq broadcast channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H}_B \otimes \mathcal{H}_E)$ and $n \in \mathbb{N}$ by

- $\bar{e}_B(\mathcal{C}, W^{\otimes n}) := \frac{1}{|\mathbf{M}|}\sum_{\mathbf{m}\in\mathbf{M}}\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|\mathbf{m})\mathrm{tr}(D_{B,\mathbf{m}}^c W_B^{\otimes n}(\mathbf{x}))$ and
- $\bar{e}_E(\mathcal{C}, W^{\otimes n}) := \frac{1}{|\mathbf{M}|}\sum_{m\in\mathbf{M}}\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|\mathbf{m})\mathrm{tr}(D_{E,m_0}^c W_E^{\otimes n}(\mathbf{x}))$,

where, $W_\gamma, \gamma \in \{B, E\}$ are the marginal channels of W. Moreover, we use the security criterion given by

$$I(M_c; E|M_0, \sigma_{s,n}), \tag{4.1}$$

where $\sigma_{s,n}$ is the code state defined by

$$\sigma_{s,n} := \frac{1}{|\mathbf{M}|}\sum_{\mathbf{m}\in\mathbf{M}} |\mathbf{m}\rangle\langle\mathbf{m}| \otimes \sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|m)W_s^{\otimes n}(\mathbf{x}), \qquad (s \in S, n \in \mathbb{N}). \tag{4.2}$$

The conditional mutual information should be understood given (1.13) and considering ONBs $\{|m_i\rangle\}_{m_i\in[M_i]} \in \mathbb{C}^{M_i}$ for $i \in \{0, c\}$ and $|\mathbf{m}\rangle := |m_0\rangle \otimes |m_c\rangle$. Based on this, we define the following achievable rate pairs.

Definition 76 *(Achievable BCC rate pair) A pair (R_0, R_c) of non-negative numbers is called an achievable BCC rate pair for $\mathcal{W}$, if for each $\epsilon, \delta > 0$, exists an $n_0(\epsilon, \delta) \in \mathbb{N}$, such that for all $n > n_0$, we find an $(n, M_{0,n}, M_{c,n})$ BCC code $\mathcal{C} = (E(\cdot|\mathbf{m}), D_{B,\mathbf{m}}, D_{E,m_0})_{\mathbf{m}\in\mathbf{M}}$ such that*

1. $\frac{1}{n}\log M_{i,n} \geqslant R_i - \delta$ *($i \in \{0, c\}$),*
2. $\sup_{s\in S}\bar{e}_\gamma(\mathcal{C}, W_s^{\otimes n}) \leqslant \epsilon$ *($\gamma \in \{B, E\}$),*
3. $\sup_{s\in S} I(M_c; E|M_0, \sigma_{s,n}) \leqslant \epsilon$,

are simultaneously fulfilled.

We define the BCC capacity region of $\mathcal{W}$ by

$$C_{BCC}[\mathcal{W}] := \{(R_0, R_c) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : (R_0, R_c) \text{ is achievable BCC rate pair for } \mathcal{W}\}. \tag{4.3}$$

To state our theorem, we define the following regions, given finite alphabets $\mathcal{U}, \mathcal{Y}$ and probability distribution $p = p_{UYX} \in \mathcal{P}(\mathcal{U} \times \mathcal{Y} \times \mathcal{X}^n)$, with the random variables U, Y, X distributed accordingly.

$$\begin{aligned}\hat{C}^{(1)}(\mathcal{W}, p, n) := \ & (R_0, R_c) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : R_0 \leqslant \inf_{s\in S}\min\{I(U; B, \omega_s), I(U; E, \omega_s)\} \wedge \\ & R_c \leqslant \inf_{s\in S} I(Y; B|U, \omega_s) - \sup_{s\in S} I(Y; E|U, \omega_s)\big\}.\end{aligned}$$

with

$$\omega_s := \sum_{(u,y,\mathbf{x}) \in \mathcal{U} \times \mathcal{Y} \times \mathcal{X}^n} p(u,y,\mathbf{x}) \, |u\rangle\langle u| \otimes |y\rangle\langle y| \otimes W_s^{\otimes n}(\mathbf{x}). \tag{4.4}$$

We state the following theorem.

Theorem 77 *Let $\mathcal{W} := \{W_s\}_{s \in S} \subset CQ(\mathcal{X}, \mathcal{H}_B \otimes \mathcal{H}_E)$ be any compound cqq broadcast channel. It holds*

$$C_{BCC}[\mathcal{W}] = \mathrm{cl}\Bigg(\bigcup_{l=1}^{\infty} \bigcup_{p} \frac{1}{l} \hat{C}^{(1)}(\mathcal{W}, p, l) \Bigg), \tag{4.5}$$

where we have used $\frac{1}{l}A := \{(\frac{1}{l}x_1, \frac{1}{l}x_2) : (x_1, x_2) \in A\}$. The second union is taken over all $p_{UYX} \in \mathcal{P}(\mathcal{U} \times \mathcal{Y} \times \mathcal{X}^l)$ such that random variable $U - Y - X$ form a Markov chain and alphabets $\mathcal{U}$ and $\mathcal{Y}$ are finite.

Remark 78 *The set given on the right hand side of (4.5) is convex and hence we do not need further convexification here. This results from time sharing arguments applied on the entropic quantities appearing in (4.5). For a short proof of a similar statement, see [Boche et al.(2019b)Boche, Janßen, and Saeedinaeeni].*

We proceed with the TPC scenario. The precise definition of the TPC codes is given in the following.

Definition 79 (TPC codes) *An $(n, M_{1,n}, M_{c,n})$ TPC code for $\mathcal{W}$, is a family $\mathcal{C} = (E(\cdot|\mathbf{m}), D_{B,\mathbf{m}})_{\mathbf{m} \in \mathbf{M}}$ with $\mathbf{M} := [M_{1,n}] \times [M_{c,n}]$, stochastic encoder $E : \mathbf{M} \to \mathcal{P}(\mathcal{X}^n)$ and a POVM $(D_{B,\mathbf{m}})_{\mathbf{m} \in \mathbf{M}}$ on $\mathcal{H}_B^{\otimes n}$.*

We define the relevant transmission error function, for any cqq broadcast channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H}_B \otimes \mathcal{H}_E)$ and $n \in \mathbb{N}$ by

$$\bar{e}_B(\mathcal{C}, W^{\otimes n}) := \frac{1}{|\mathbf{M}|} \sum_{\mathbf{m} \in \mathbf{M}} \sum_{\mathbf{x} \in \mathcal{X}^n} E(\mathbf{x}|\mathbf{m}) \mathrm{tr}(D_{B,\mathbf{m}}^c W_B^{\otimes n}(\mathbf{x})).$$

Moreover, we use the security criterion given by

$$I(M_c; E | M_1, \sigma_{s,n}), \tag{4.6}$$

where $\sigma_{s,n}$ is the code state defined by

$$\sigma_{s,n} := \frac{1}{|\mathbf{M}|} \sum_{\mathbf{m} \in \mathbf{M}} |\mathbf{m}\rangle\langle\mathbf{m}| \otimes \sum_{\mathbf{x} \in \mathcal{X}^n} E(\mathbf{x}|\mathbf{m}) W_s^{\otimes n}(\mathbf{x}). \tag{4.7}$$

Again, we not that the conditional mutual information should be understood given (1.13) and considering ONBs $\{|m_i\rangle\}_{m_i \in [M_i]} \in \mathbb{C}^{M_i}$ for $i \in \{1, c\}$ and $|\mathbf{m}\rangle := |m_1\rangle \otimes |m_c\rangle$. Based on this, we define the following achievable rate pairs.

Definition 80 *(Achievable TPC rate pair) A pair (R_1, R_c) of non-negative numbers is* *called an achievable TPC rate pair for* $\mathcal{W}$*, if for each* $\epsilon, \delta > 0$*, exists an* $n_0(\epsilon, \delta) \in \mathbb{N}$*, such that for all* $n > n_0$*, we find an* $(n, M_{1,n}, M_{c,n})$ *TPC code* $\mathcal{C} = (E(\cdot|\mathbf{m}), D_{B,\mathbf{m}})_{\mathbf{m}\in\mathbf{M}}$ *such that*

1. $\frac{1}{n}\log M_{i,n} \geqslant R_i - \delta$ *(* $i \in \{1, c\}$*),*

2. $\sup_{s\in S} \overline{e}_B(\mathcal{C}, W_s^{\otimes n}) \leqslant \epsilon$*,*

3. $\sup_{s\in S} I(M_c; E|M_1, \sigma_{s,n}) \leqslant \epsilon$

are simultaneously fulfilled.

We define the TPC capacity region of $\mathcal{W}$ by

$$C_{TPC}[\mathcal{W}] := \{(R_1, R_c) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : (R_1, R_c) \text{ is achievable TPC rate for } \mathcal{W}\}. \tag{4.8}$$

To state our theorem, we define the following sub-regions, given finite alphabets $\mathcal{V}, \mathcal{Y}$ and probability distribution $p = p_{VYX} \in \mathcal{P}(\mathcal{V} \times \mathcal{Y} \times \mathcal{X}^n)$, with the random variables V, Y, X distributed accordingly.

$$C^{(1)}\big(\mathcal{W}, p, n\big) := \big\{(R_1, R_c) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : R_1 \leqslant \inf_{s\in S} I(V; B, \omega_s) \wedge \\ R_c \leqslant \inf_{s\in S} I(Y; B|V, \omega_s) - \sup_{s\in S} I(Y; E|V, \omega_s)\big\}.$$

with

$$\omega_s := \sum_{(v,y,\mathbf{x})\in\mathcal{V}\times\mathcal{X}\times\mathcal{X}} p(v, y, \mathbf{x})\,|v\rangle\langle v| \otimes |y\rangle\langle y| \otimes W_s^{\otimes n}(\mathbf{x}). \tag{4.9}$$

We can state the following theorem.

Theorem 81 *Let* $\mathcal{W} := \{W_s\}_{s\in S} \subset CQ(\mathcal{X}, \mathcal{H}_B \otimes \mathcal{H}_E)$ *be any compound cqq broadcast channel. It holds*

$$C_{TPC}[\mathcal{W}] = \mathrm{cl}\bigg(\bigcup_{l=1}^{\infty}\bigcup_{p}\frac{1}{l}C^{(1)}\big(\mathcal{W}, p, l\big)\bigg). \tag{4.10}$$

The second union is taken over all $p_{VYX} \in \mathcal{P}(\mathcal{V} \times \mathcal{Y} \times \mathcal{X}^l)$ *such that random variable* $V - Y - X$ *form a Markov chain and alphabets* $\mathcal{V}$ *and* $\mathcal{Y}$ *are finite.*

Again, we note Remark 78, regarding convexity of the set on the right hand side of (4.10).

4.3. Coding for broadcast channel

In this section we present coding strategies for BCC and TPC communication scenarios sufficient to achieve each point in the capacity region. We prove appropriate inner bounds on the capacity regions, namely the direct parts of the main theorems presented in the previous section. Here, we begin by some preliminary results, in the statements of which, we make use of typical sets and projections. The use of these objects are standard in classical as well as quantum information theory. The reader will find detailed explanations in [Csiszár and Körner(2011a)]. We begin this section nevertheless, by introducing these objects. Given two probability distributions $p \in \mathcal{P}(\bar{\mathcal{X}})$ and $\forall x \in \bar{\mathcal{X}}$, $t(\cdot|x) \in \mathcal{P}(\bar{\mathcal{Y}})$, $n \in \mathbb{N}$, $\delta > 0$, we define the following sets. The set of δ-typical sequences in $\bar{\mathcal{X}}^n$, is defined by

$$T^n_{p,\delta} := \{\mathbf{x} : \forall x \in \bar{\mathcal{X}}, |\frac{1}{n}N(x|\mathbf{x}) - p(x)| \leqslant \delta \ \wedge \ q(x) = 0 \iff N(x|\mathbf{x}) = 0\} \tag{4.11}$$

with $N(x|\mathbf{x})$, the number of occurrences of letter x in word $\mathbf{x}$. Also, the set of conditionally typical sequences in $\bar{\mathcal{Y}}^n$, is given by

$$\begin{aligned} T_{t,\delta}(\mathbf{x}) : = \{\mathbf{y} \in \bar{\mathcal{Y}}^n : \forall x \in \bar{\mathcal{X}}, y \in \bar{\mathcal{Y}} : |\frac{1}{n}N(x,y|\mathbf{x},\mathbf{y}) - \frac{1}{n}t(y|x)N(x|\mathbf{x})| \leqslant \delta \text{ and} \\ t(y|x) = 0 \iff N(x,y|\mathbf{x},\mathbf{y}) = 0 \text{ for } x \in \bar{\mathcal{X}}, y \in \bar{\mathcal{Y}}\}. \end{aligned}$$

The pruned distributions associated with p and $t(\cdot|x)$ are given by the following respectively.

$$p'_{n,\delta}(\mathbf{x}) := \begin{cases} \frac{p^{\otimes n}(\mathbf{x})}{p^{\otimes n}(T^n_{p,\delta})}, & \text{if } \mathbf{x} \in T^n_{p,\delta} \\ 0, & \text{otherwise,} \end{cases} \tag{4.12}$$

and

$$t'_{n,\delta}(\mathbf{y}|\mathbf{x}) := \begin{cases} \frac{t^{\otimes n}(\mathbf{y}|\mathbf{x})}{t^{\otimes n}(T_{t,\delta}(\mathbf{x})|\mathbf{x})}, & \text{if } \mathbf{y} \in T_{t,\delta}(\mathbf{x}) \\ 0, & \text{otherwise.} \end{cases} \tag{4.13}$$

For the remainder of this section, pruned distributions defined above, will be denoted by primed letters indicating the probability distribution, indexed by the number of available copies of the system. For instance the pruned probability distribution related to $r \in \mathcal{P}(\mathcal{X})$, over $T^n_{r,\delta}$ will be denoted by $r'_{n,\delta}$. In (4.11), when $\delta = 0$, we have the exact type notified by T^n_p. We also define the set of types by

$$\mathcal{T}(\bar{\mathcal{X}}, n) := \{\lambda \in \mathcal{P}(\bar{\mathcal{X}}) : T^n_\lambda \neq \varnothing\}. \tag{4.14}$$

The following lemma contains the properties typical projections, that projection operators assigned to typical sets.

Lemma 82 *Let $\lambda \in \mathcal{P}(\mathcal{A})$ with $\lambda(x) > 0$ for all $x \in \mathcal{A} \subset \mathcal{X}$, $\{\rho_x\}_{x\in\mathcal{X}} \subset \mathcal{S}(\mathcal{K}_A)$ and $\delta > 0$. For $\mathbf{x} \in T^n_{\lambda,\delta}$ with $\mathbf{x} := (x_1, \dots, x_n)$ and $\rho_{\mathbf{x}} := \bigotimes_{i=1}^n \rho_{x_i}$. Define*

$$\theta := \sum_{x\in\mathcal{X}} \lambda(x)\, |x\rangle\langle x|^X \otimes \rho_x.$$

There exist positive constants $\varUpsilon(\delta), \varGamma(\delta)$ and $\varDelta(\delta)$ depending on δ and an orthogonal projector $\varPi_{\rho_{\mathbf{x}},\delta}$ such that

1. $\mathrm{tr}(\rho_{\mathbf{x}}\varPi_{\rho_{\mathbf{x}},\delta}) \geqslant 1 - 2^{-n\varUpsilon(\delta)}$,

2. $\mathrm{tr}(\Pi_{\rho_{\mathbf{x}},\delta}) \leqslant 2^{n(S(A|X,\theta)+\Delta(\delta))}$,

3. $\Pi_{\rho_{\mathbf{x}},\delta}\rho_{\mathbf{x}}\Pi_{\rho_{\mathbf{x}},\delta} \leqslant 2^{-n(S(A|X,\theta)+\Gamma(\delta))}\Pi_{\rho_{\mathbf{x}},\delta}$,
Also, let $W : \mathcal{Y} \to \mathcal{S}(\mathcal{K}_B)$ be a cq channel and $r(\cdot|x) \in \mathcal{P}(\mathcal{Y})$, for all $x \in \mathcal{X}$. Define the state

$$\theta' := \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \lambda(x)\,|x\rangle\langle x| \otimes r(y|x)\,|y\rangle\langle y| \otimes W(y).$$

For $\mathbf{y} \in T_{r,\delta}(\mathbf{x})$, there exist positive constants $\Upsilon'(\delta), \Delta'(\delta), \Gamma'(\delta)$ and an orthogonal projector $\Pi_{W,\mathbf{x},\delta}(\mathbf{y})$, commuting with $W^{\otimes n}(\mathbf{y})$, satisfying

4. $\mathrm{tr}[W^{\otimes n}(\mathbf{y})\Pi_{W,\mathbf{x},\delta}(\mathbf{y})] \geqslant 1 - 2^{-n\Upsilon'(\delta)}$,

5. $\mathrm{tr}[\Pi_{W,\mathbf{x},\delta}(\mathbf{y})] \leqslant 2^{n(S(B|XY,\theta)+\Delta'(\delta))}$,

6. $\Pi_{W,\mathbf{x},\delta}(\mathbf{y})W^{\otimes n}(\mathbf{y})\Pi_{W,\mathbf{x},\delta}(\mathbf{y}) \leqslant 2^{-n(S(B|XY,\theta')+\Gamma'(\delta))}\Pi_{W,\mathbf{x},\delta}(\mathbf{y})$.

Finally, we have the following total conditional subspace projection. For $\rho_x = \sum_{y\in\mathcal{Y}} r(y|x)W(y)$, the projection $\Pi_{W,\mathbf{x},\delta} := \Pi_{\rho_{\mathbf{x}},\delta}$ with properties 1-3, for $\mathbf{y} \in T_{r,\delta}(\mathbf{x})$ also has the following property.

$$\mathrm{tr}(\Pi_{W,\mathbf{x},\delta}W^{\otimes n}(\mathbf{y})) \geqslant 1 - 2^{-n\Upsilon"(\delta)}, \tag{4.15}$$

for some constant $\Upsilon"(\delta) > 0$ depending on δ.

Proof 83 *Properties 1-3 result directly from Lemma 14 [Boche et al.(2019b)Boche, Janßen, and Saeedinaeeni]. Properties 4-6 and (4.15), result from applying the same concatenation arguments as in the proof of Lemma 14 [Boche et al.(2019b)Boche, Janßen, and Saeedinaeeni], on inequalities (4)-(7) from [Cai(2018)].*

A crucial ingredient for the achievablity proofs in this chapter is Lemma 84 below. It states existence of certain universal random codes for cq channels given a ”typical word”.

Lemma 84 *Let $\{W_s\}_{s\in S} \subset CQ(\mathcal{Y}, \mathcal{K}_B)$ be any set of cq channels, $q \in \mathcal{P}(\mathcal{X})$ and $r(\cdot|x) \in \mathcal{P}(\mathcal{Y})$ for each $x \in \mathcal{X}$. For $\delta > 0$, there exists $n_0 \in \mathbb{N}$, such that for $n > n_0$, for each $\mathbf{x} \in T^n_{q,\delta}$, there exists a map $y : (y_1, \dots, y_M) \mapsto (\Lambda_1(y) \dots, \Lambda_M(y))$, such that $(\Lambda_m(y))_{m\in[M]} \subset \mathcal{L}(\mathcal{K}_B^{\otimes n})$ is a POVM and for any family $Y := (Y_1, \dots, Y_M)$ of random variables, distributed i.i.d according to $r'_{n,\delta}(\cdot|\mathbf{x})$, namely the pruned distribution of $r(\cdot|x)$ (see (4.13)), we have*

$$\mathbb{E}_Y\big[\sup_{s\in S}\frac{1}{M}\sum_{m\in[M]} \mathrm{tr}(W_s^{\otimes n}(Y_m)\Lambda^c_m(Y))\big] \leqslant \epsilon_n$$

with $\epsilon_n \to 0$ exponentially and

$$\frac{1}{n}\log M \geqslant \inf_{s\in S} I(Y; B|X, \sigma_s) - c\delta,$$

with some constant $c > 0$ and

$$\sigma_s := \sum_{x\in\mathcal{X}} q(x)\, |x\rangle\langle x| \otimes \sum_{y\in\mathcal{Y}} r(y|x)\, |y\rangle\langle y| \otimes W_s(y).$$

Proof 85 *We present a full argument in Appendix D.*

The following statement is an immediate consequence of the above, for the case $|\mathcal{X}| = 1$. We include this statement for clarity of reference later on.

Lemma 86 *Let $\{W_s\}_{s\in S} \subset CQ(\mathcal{Y}, \mathcal{K}_B)$ be any set of cq channels and $r \in \mathcal{P}(\mathcal{Y})$. For $\delta > 0$, there exists n_0, such that for $n > n_0$, there exists a map $y : (y_1, \ldots, y_M) \mapsto (\Lambda_1(y), \ldots, \Lambda_M(y))$, such that $(\Lambda_m(y))_{m\in[M]}$ is a POVM and for any family $Y := (Y_1, \ldots, Y_M)$ of random variables, distributed i.i.d according to $r'_{n,\delta}$, namely the pruned distribution of r (see (4.12)), we have*

$$\mathbb{E}_Y\big[\sup_{s\in S}\frac{1}{M}\sum_{m\in[M]} \operatorname{tr}(W_s^{\otimes n}(Y_m)\Lambda_m^c(Y))\big] \leqslant \epsilon_n,$$

with $\epsilon_n \to 0$ exponentially and

$$\frac{1}{n}\log M \geqslant \inf_{s\in S} I(Y; B, \sigma_s) - c\delta$$

for some constant $c > 0$ and

$$\sigma_s := \sum_{y\in\mathcal{Y}} r(y)\, |y\rangle\langle y| \otimes W_s(y).$$

In Section 4.3.1 and Section 4.3.2, we show that the above statements give us the desired codes for transmission of public and common messages. These statements generalize the coding results from [Mosonyi(2015)] to include pruned input distributions rather than distributions of n-fold product form.

Finally, to obtain codes for transmission of confidential messages, we perform privacy amplification arguments on the public part of the codebook achieved from Lemma 84 (cf. [Boche et al.(2014)Boche, Cai, Cai, and Deppe]). To do so, we need the following inequality.

Theorem 87 ([Ahlswede and Winter(2002)], Theorem 19) *Let $\mu > 0$, $\epsilon \in (0, \frac{1}{2})$ be positive numbers and $X_1, \ldots, X_L$ an independent and identically distributed family of positive semi-definite random matrices on $\mathbb{C}^d$ such that the bounds $X \leqslant \mu\mathbb{1}_{\mathbb{C}^d}$ and $\mathbb{E}X \geqslant \epsilon\mathbb{1}_{\mathbb{C}^d}$ apply. It holds*

$$\Pr\left(\left\|\frac{1}{L}\sum_{i=1}^{L} X_i - \mathbb{E}X\right\|_1 \;>\; \epsilon\right) \;\leqslant\; 2\cdot d\cdot \exp\left(-L\frac{\epsilon^3}{2d\mu\ln 2}\right)$$

Equipped with these preliminary results, we prove the direct parts of the capacity theorems for BCC and TPC in the following two subsections.

4.3.1. BCC codes

In this section, we prove the following lemma.

Lemma 88 *Let $\mathcal{W} := \{W_s\}_{s\in S} \subset CQ(\mathcal{X}, \mathcal{H}_B \otimes \mathcal{H}_E)$ be any compound cqq broadcast channel. It holds*

$$C_{BCC}[\mathcal{W}] \supset \mathrm{cl}\left(\bigcup_{l=1}^{\infty}\bigcup_{p}\frac{1}{l}\hat{C}^{(1)}(\mathcal{W}, p, l)\right),$$

where the second union is taken over all $p_{UYX} \in \mathcal{P}(\mathcal{U} \times \mathcal{Y} \times \mathcal{X}^l)$ such that random variable $U - Y - X$ form a Markov chain and alphabets $\mathcal{U}$ and $\mathcal{Y}$ are finite.

The main step towards proving Lemma 88, is the following statement.

Lemma 89 (Broadcast channel with confidential messages) *Let $\mathcal{W} := \{W_s\}_{s\in S} \subset CQ(\mathcal{X}, \mathcal{H}_B\otimes\mathcal{H}_E)$ be any compound cqq broadcast channel. For $p_{UYX} \in \mathcal{P}(\mathcal{U}\times\mathcal{Y}\times\mathcal{X})$ where $U - Y - X$ form a Markov chain and $\delta, \epsilon > 0$, there exists $n_0 \in \mathbb{N}$, such that for $n > n_0$, we find an $(n, M_{0,n}, M_{c,n})$ BCC code $\mathcal{C} = (E(\cdot|m), D_{B,m}, D_{E,m_0})_{m=(m_0,m_c)\in[M_{0,n}]\times[M_{c,n}]}$ with*

1. $\frac{1}{n}\log M_{0,n} \geqslant \inf_{s\in S}\min\{I(U; B, \omega_s, I(U; E, \omega_s\} - c\delta$,
2. $\frac{1}{n}\log M_{c,n} \geqslant \inf_{s\in S} I(Y; B|U, \omega_s) - \sup_{s\in S} I(Y; E|U, \omega_s) - c\delta$ *with some constant $c > 0$ and ω_s defined by (4.4).*
3. $\inf_{s\in S}\frac{1}{|\mathbf{M}|}\sum_{\mathbf{m}\in\mathbf{M}}\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|\mathbf{m})\mathrm{tr}[W_{B,s}^{\otimes n}(\mathbf{x})D_{B,\mathbf{m}}] \geqslant 1 - \epsilon$
4. $\inf_{s\in S}\frac{1}{|\mathbf{M}|}\sum_{\mathbf{m}\in\mathbf{M}}\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|\mathbf{m})\mathrm{tr}[W_{E,s}^{\otimes n}(\mathbf{x})D_{E,m_0}] \geqslant 1 - \epsilon$
5. $\sup_{s\in S} I(M_c; E|M_0, \sigma_{s,n}) \leqslant \epsilon$

with state $\sigma_{s,n}$ defined by (4.2).

Applying standard double-blocking arguments on Lemma 89, will prove Lemma 88. In the same vein as the coding steps taken in [Schaefer and Boche(2014b)], we prove Lemma 89 in two steps. At first, we prove the following random coding result, that guarantees reliable decoding of common messages by Bob and Eve, and reliable decoding of public messages by Bob. Here, we do not concern ourselves with the security condition. In the next step, we apply privacy amplification arguments on the public part of the codebook, to achieve the desired confidential message transmission rate.

Lemma 90 *Let $\mathcal{W} := \{W_s\}_{s\in S} \subset CQ(\mathcal{Y}, \mathcal{H}_B \otimes \mathcal{H}_E)$ be any compound cqq broadcast channel and $\mathcal{U}$ be a finite alphabet. For any $\delta > 0$, $q \in \mathcal{P}(\mathcal{U}), r(\cdot|u) \in \mathcal{P}(\mathcal{Y})$, $u \in \mathcal{U}$ and large enough values of n, the following exist.*

- *A family $(u_m, D_{E,m})_{m\in[M_{0,n}]}$ of codes with $u_m \in T^n_{q,\delta}$ and $(D_{E,m})_{m\in[M_{0,n}]} \subset \mathcal{L}(\mathcal{H}_E^{\otimes n})$ a POVM.*

- *A map $y : (y_{ij})_{(i,j)\in[M_{0,n}]\times[M_{1,n}]} \mapsto (D_{B,ij}(y))_{(i,j)\in[M_{0,n}]\times[M_{1,n}]}$, such that $(D_{B,ij}(y))_{(i,j)\in[M_{0,n}]\times[M_{1,n}]} \in \mathcal{L}(\mathcal{H}_B^{\otimes n})$ is a POVM and for any family $Y = (Y_{ij})_{(i,j)\in[M_{0,n}]\times[M_{1,n}]}$ of random variables such that for each $m \in [M_{0,n}]$, $Y^m = (Y_{mj})_{j\in[M_{1,n}]}$ is distributed i.i.d according to $r^{'}(\cdot|u_m^n)$, namely the pruned distribution of $r(\cdot|u)$ (see (4.13)), we have*

$$\frac{1}{n}\log M_{0,n} \geqslant \inf_{s\in S} \min\{I(U;B,\omega_s), I(U;E,\omega_s)\} - c\delta,$$

$$\frac{1}{n}\log M_{1,n} \geqslant \inf_{s\in S} I(Y;B|U,\omega_s) - c\delta,$$

$$\mathbb{E}_Y\left[\inf_{s\in S}\frac{1}{M_{0,n}M_{1,n}}\sum_{(m,i)\in[M_{0,n}]\times[M_{1,n}]} \mathrm{tr}[W_{B,s}^{\otimes n}(Y_{mi})D_{B,mi}(Y)]\right] \geqslant 1-\epsilon_n,$$

$$\mathbb{E}_Y\left[\inf_{s\in S}\frac{1}{M_{0,n}M_{1,n}}\sum_{(m,i)\in[M_{0,n}]\times[M_{1,n}]} \mathrm{tr}[W_{E,s}^{\otimes n}(Y_{mi})D_{E,m}]\right] \geqslant 1-\epsilon_n$$

with $\epsilon_n \to 0$ exponentially, constant $c > 0$ and $\omega_s = \sum_{u\in\mathcal{U}} q(u)\,|u\rangle\langle u|\otimes r(y|u)\,|y\rangle\langle y|\otimes W_s(y)$.

Proof 91 *We approximate $\{W_s\}_{s\in S}$ by a finite τ_n-net $\{W_s\}_{s\in S_n} \subset \{W_s\}_{s\in S}$ with $\tau_n := 2^{-\frac{n\nu}{2}}$ with a constant positive number ν to be determined later. We choose the net small enough to have $\log|S_n| \leqslant 2\cdot|\mathcal{X}|\cdot\dim(\mathcal{H}_B\otimes\mathcal{H}_E)^2(\log 6 + n\nu/2)$ which is possible by Lemma 197. For $\gamma \in \{B, E\}$ and $s \in S_n$, consider the effective channel $\hat{W}_{\gamma,s,n} : \mathcal{U}^n \to \mathcal{S}(\mathcal{H}_\gamma^{\otimes n})$ defined by $\hat{W}_{\gamma,s}(\cdot) := \sum_{y\in\mathcal{Y}} r(y|\cdot)W_{\gamma,s}(y)$. Applying Lemma 86 on the channel set $\{\hat{W}_{\gamma,s}\}_{s\in S_n}$ and probability distribution q, yields the existence of the random $(n, M_{0,n})$ code $\mathcal{C}(U)$ with $U = (U_1, \ldots, U_{M_{0,n}})$, a sequence of i.i.d random variables distributed according to $q'_{n,\delta}$ and POVMs $(D_{\gamma,m}(U))_{m\in[M_{0,n}]} \subset \mathcal{L}(\mathcal{H}_\gamma^{\otimes n})$ such that*

$$\mathbb{E}_U\big[\min_{s\in S_n}\frac{1}{M_{0,n}}\sum_{m\in[M_{0,n}]}\mathrm{tr}(D_{\gamma,m}(U)\hat{W}_{\gamma,s}^{\otimes n}(U_m))\big] \geqslant 1-\epsilon_{0,n}. \tag{4.16}$$

with $\epsilon_{0,n} \to 0$ exponentially and

$$\frac{1}{n}\log M_{0,n} \geqslant \min_{s\in S_n} I(U;\gamma,\omega_s) - c_0\delta.$$

Hence we have

$$\frac{1}{n}\log M_{0,n} \geqslant \min_{s\in S_n}\min\{I(U;B,\omega_s), I(U;E,\omega_s)\} - c_0\delta.$$

Given (4.16), we can conclude the existence of one realization $(u_1, \ldots, u_{M_{0,n}})$ *of random variable* U*, and POVMs* $(D_{\gamma,m})_{m\in[M_{0,m}]} \in \mathcal{L}(\mathcal{H}_\gamma^{\otimes n})$*, suitable for transmission of common messages, namely*

$$\min_{s\in S_n} \frac{1}{M_{0,n}} \sum_{m\in[M_{0,n}]} \operatorname{tr}(D_{\gamma,m}\hat{W}_{\gamma,s}^{\otimes n}(u_m)) \geqslant 1 - \epsilon_{0,n}. \tag{4.17}$$

Before moving on to the private message, notice that for each $\mathbf{u} \in T_{q,\delta}^n$*, using the abbreviation* $T_\delta := r^{\otimes n}(T_{r,\delta}(\mathbf{u}))$*, we have*

$$\begin{aligned}\|\hat{W}_{\gamma,s}^{\otimes n}(\mathbf{u}) - \sum_{\mathbf{y}\in\mathcal{Y}^n} r'_n(\mathbf{y}|\mathbf{u})W_{\gamma,s}^{\otimes n}(\mathbf{y})\|_1 &\leqslant \sum_{\mathbf{y}\in T_{r,\delta}(\mathbf{u})} r^{\otimes n}(\mathbf{y}|\mathbf{u})(\frac{1}{T_\delta} - 1)\|W_{\gamma,s}^{\otimes n}(\mathbf{y})\|_1 \\ &+ \sum_{\mathbf{y}\in T_{r,\delta}^c(\mathbf{u})} r^{\otimes n}(\mathbf{y}|\mathbf{u})\|W_{\gamma,s}^{\otimes n}(\mathbf{y})\|_1 \leqslant 2(1-T_\delta) \leqslant 2\cdot 2^{-n\delta}.\end{aligned} \tag{4.18}$$

The upper bound above comes from the fact that T_δ *approaches unity exponentially with* n *(cf. [Csiszár and Körner(2011a)]). Now we pursue with the private message, namely the one Bob has to decode while Eve may or may not. For each* $u_{\hat{m}}, \hat{m} \in [M_{0,n}]$ *obtained above, apply Lemma 84 on* $\{W_s\}_{S_n}$ *and probability distribution* $r(\cdot|u)$*,* $u \in \mathcal{U}$*. on Lemma 84, we obtain the existence of a random code* $\mathcal{C}(Y^{u_{\hat{m}}})$ *with* $Y^{u_{\hat{m}}} = (Y_{\hat{m},1}, \ldots, Y_{\hat{m},M_{1,n}})$ *and decoding operation* $(\Lambda_m(Y^{u_{\hat{m}}}))_{m\in[M_{1,n}]}$*, such that* $Y^{u_{\hat{m}}}$ *is distributed according to* $r'_{n,\delta}(\cdot|u_{\hat{m}})^{\otimes M_{1,n}}$ *with*

$$\mathbb{E}_{Y^{u_{\hat{m}}}}\big[\inf_{s\in S_n} \frac{1}{M_{1,n}} \sum_{m\in[M_{1,n}]} \operatorname{tr}(\Lambda_m(Y^{u_{\hat{m}}})W_{B,s}^{\otimes n}(Y_{\hat{m},m}))\big] \geqslant 1 - \epsilon_{1,n}, \tag{4.19}$$

and

$$\frac{1}{n}\log M_{1,n} \geqslant I(Y;B|U,\omega_s) - c_1\delta.$$

We have

$$\begin{aligned}\min_{s\in S_n} \frac{1}{M_{0,n}} &\sum_{\hat{m}\in[M_{0,n}]} \mathbb{E}_{Y^{u_{\hat{m}}}}\big[\frac{1}{M_{1,n}} \sum_{m\in[M_{1,n}]} \operatorname{tr}(D_{\gamma,\hat{m}}W_{s,\gamma}^{\otimes n}(Y_{\hat{m},m}))\big] \\ &= \min_{s\in S_n} \frac{1}{M_{0,n}} \sum_{\hat{m}\in[M_{0,n}]} (\operatorname{tr}(D_{\gamma,\hat{m}} \sum_{\mathbf{y}\in\mathcal{Y}^n} r'_{n,\delta}(\mathbf{y}|u_{\hat{m}})W_{s,\gamma}^{\otimes n}(\mathbf{y}))) \\ &= \min_{s\in S_n} \frac{1}{M_{0,n}} \sum_{\hat{m}\in[M_{0,n}]} \operatorname{tr}(D_{\gamma,\hat{m}}\hat{W}_{\gamma,s}^{\otimes n}(u_{\hat{m}})) - 2\cdot 2^{-n\delta} \geqslant 1 - \epsilon_{2,n}.\end{aligned} \tag{4.20}$$

where in the first equality, we have calculated the expectation value given that for each $\hat{m} \in [M_{0,n}]$*,* $Pr(Y_{\hat{m},m} = \mathbf{y}) = r'_n(\mathbf{y}|u_{\hat{m}}), \forall m \in [M_{0,n}]$*, and in the last line, we have observed (4.18) and inserted (4.17), setting* $\epsilon_{2,n} := \epsilon_{0,n} + 2\cdot 2^{-n\delta}$*. Consider the random decoding operation* $(D_{B,\hat{m},m}(Y))_{(\hat{m},m)\in[M_{0,n}]\times[M_{1,n}]}$ *defined for each message pair by* $D_{B,\hat{m},m}(Y) :=$

$\sqrt{D_{B,\hat{m}}}\Lambda_m(Y^{u_{\hat{m}}})\sqrt{D_{B,\hat{m}}}$. We have

$$\begin{aligned}&\mathbb{E}_Y\Big[\frac{1}{|S_n|}\sum_{s\in S_n}\frac{1}{M_{0,n}}\sum_{\hat{m}\in[M_{0,n}]}\frac{1}{M_{1,n}}\sum_{m\in[M_{1,n}]}\operatorname{tr}(D_{B,\hat{m},m}(Y)W_{B,s}^{\otimes n}(Y_{\hat{m},m}))\Big]=\\&\frac{1}{|S_n|}\sum_{s\in S_n}\frac{1}{M_{0,n}}\sum_{\hat{m}\in[M_{0,n}]}\mathbb{E}_{Y^{u_{\hat{m}}}}\Big[\frac{1}{M_{1,n}}\sum_{m\in[M_{1,n}]}\operatorname{tr}(\sqrt{D_{B,\hat{m}}}\Lambda_m(Y^{u_{\hat{m}}})\sqrt{D_{B,\hat{m}}}W_{B,s}^{\otimes n}(Y_{\hat{m},m}))\Big]\geqslant\\&\frac{1}{|S_n|}\sum_{s\in S_n}\frac{1}{M_{0,n}}\sum_{\hat{m}\in[M_{0,n}]}\Big(\mathbb{E}_{Y^{u_{\hat{m}}}}\Big[\frac{1}{M_{1,n}}\sum_{m\in[M_{1,n}]}\operatorname{tr}(\Lambda_m(Y^{u_{\hat{m}}})W_{B,s}^{\otimes n}(Y_{\hat{m},m}))\Big]-\\&2\sqrt{1-\mathbb{E}\Big[\frac{1}{M_{1,n}}\sum_{m\in[M_{1,n}]}\operatorname{tr}(D_{B,\hat{m}}W_{B,s}^{\otimes n}(Y_{\hat{m},m}))\Big]}\Big)\geqslant 1-\epsilon_{1,n}-2\sqrt{\epsilon_{2,n}},\end{aligned}\tag{4.21}$$

where in the first inequality, we have used Lemma 208, and in the last line, we have inserted the lower bounds from (4.20) and (4.19) and used concavity of the square root function. Applying standard net approximation techniques used for example in proof of Lemma 84, we obtain the claim of the lemma.

At this point we can prove Lemma 89, by applying privacy amplification arguments (c.f [Boche et al.(2014)Boche, Cai, Cai, and Deppe]) on the M_1 part of the messages obtained in Lemma 90. This is done by using equidistribution when inputting part of these messages to confuse the eavesdropper. The other part of M_1 will then be secure.

Proof 92 (Proof of Lemma 89) *Let $p_{UYX}(u,y,x)=p_{UY}(u,y)p_{X|Y}(x|y)$ and $p_{UY}(u,y)=q(u)r(y|u)$ $\forall(u,y,x)\in\mathcal{U}\times\mathcal{Y}\times\mathcal{X}$. We approximate $\{W_s\}_{s\in S}$ by a finite τ_n-net $\{W_s\}_{S_n}\subset\{W_s\}_{s\in S}$ with $\tau_n:=2^{-\frac{n\nu}{2}}$ with a constant positive number ν to be determined later. We choose the net small enough to fulfill the cardinality bound $\log|S_n|\leqslant 2\cdot|\mathcal{X}|\cdot\dim(\mathcal{H}_B\otimes\mathcal{H}_E)^2(\log 6+n\nu/2)$ which is possible by Lemma 197. Let $\delta>0$, $n\in\mathbb{N}$ and pruned probability distributions $q'_{n,\delta},r'_n(\cdot|\mathbf{u})$ over $T^n_{q,\delta}$ and $T_{r,\delta}(\mathbf{u}),(\mathbf{u}\in\mathcal{U}^n)$ be given. Set*

$$M_{0,n}=\lfloor 2^{n\left(\min_{s\in S_n}\min\{I(U;B,\omega_s),I(U;E,\omega_s)\}-c\delta\right)}\rfloor,\tag{4.22}$$

$$J_n=\lfloor 2^{n\left(\min_{s\in S_n}I(Y;B|U,\omega_s)-\max_{s\in S_n}I(Y;E|U,\omega_s)-2\Delta(\delta)-c\delta\right)}\rfloor\tag{4.23}$$

and

$$L_n=\lceil 2^{n\max_{s\in S_n}I(Y;E|U,\omega_s)+n\Delta(\delta)}\rceil.\tag{4.24}$$

For the effective channel $\tilde{W}_s:\mathcal{Y}\to\mathcal{S}(\mathcal{H}_B\otimes\mathcal{H}_E)$ defined by $\tilde{W}_s(\cdot):=\sum_{x\in\mathcal{X}}p_{X|Y}(x|\cdot)W_s(x),\forall s\in S$, according to Lemma 90, there exists a family $(u_m,D_{E,m})_{m\in[M_{0,n}]}$ and a random family $\mathcal{C}(Y)=(Y_{mjl},D_{B,mjl}(Y))_{(m,j,l)\in[M_{0,n}]\times[J_n]\times[L_n]}$, such that for events

$$\mathbf{A}:=\ \max_{s\in S_n}\frac{1}{M_{0,n}J_nL_n}\sum_{m,j,l}\operatorname{tr}(\tilde{W}_{B,s}^{\otimes n}(Y_{mjl})D^c_{B,mji}(Y))\geqslant\sqrt{\epsilon_n}\}$$

and

$$\mathbf{B} := \quad \max_{s\in S_n} \frac{1}{M_{0,n}J_nL_n}\sum_{m,j,l}\mathrm{tr}(\tilde{W}_{E,s}^{\otimes n}(Y_{mjl})D_{E,m}^c) \geqslant \sqrt{\epsilon_n}\}, \tag{4.25}$$

we have

$$\Pr[\mathbf{A}\cup\mathbf{B}] \leqslant 2\sqrt{\epsilon_n}, \tag{4.26}$$

where we have used the Markov inequality to obtain the above probability from the expectation value of the same event, and applied the union bound to get the probability of the complementary events (one with respect to $\tilde{W}_{B,s}$ and the other with respect to $\tilde{W}_{E,s}$). Here, ϵ_n goes to zero exponentially, given the appropriate choice of τ_n, as evident in the proof of Lemma 84. We define the following quantities for each $s \in S_n$ and $\mathbf{u} \in T_{q,\delta}^n$.

$$Q_s^{\mathbf{u}}(\mathbf{y}) := \Pi_{\tilde{W}_{E,s},\mathbf{u},\delta}\Pi_{\tilde{W}_{E,s},\mathbf{u},\delta}(\mathbf{y})\tilde{W}_{E,s}^{\otimes n}(\mathbf{y})\Pi_{\tilde{W}_{E,s},\mathbf{u},\delta}(\mathbf{y})\Pi_{\tilde{W}_{E,s},\mathbf{u},\delta} \tag{4.27}$$

with quantities defied in Lemma 82 and

$$\Theta_s^{\mathbf{u}} := \sum_{y^n\in\mathcal{Y}^n} r'_{n,\delta}(\mathbf{y}|\mathbf{u})Q_s^{\mathbf{u}}(\mathbf{y}). \tag{4.28}$$

Given property 4 of Lemma 82, (4.15) and Lemma 203, we have $\forall\mathbf{u}\in\mathcal{U}^n, \mathbf{y}\in T_{r,\delta}(\mathbf{u})$ and $s\in S_n$

$$\| \tilde{W}_{E,s}^{\otimes n}(\mathbf{y}) - Q_s^{\mathbf{u}}(\mathbf{y}) \|_1 \leqslant \sqrt{2^{-n\Upsilon(\delta)+1} + \sqrt{2^{-n\Upsilon''(\delta)+2}}} := \epsilon_{1,n}. \tag{4.29}$$

Clearly $\epsilon_{1,n} \to 0$ exponentially. Applying Theorem 87 with $\mathbb{C}^d$ the range space of projection $\Pi_{\tilde{W}_{E,s},\mathbf{u},\delta}$, by property 2 of Lemma 82 we have

$$d \leqslant 2^{S(E|U,\omega_s^{\otimes n})+n\Delta(\delta)} \tag{4.30}$$

Furthermore, from the property 6 of the projections introduced in Lemma82, we have for all $\mathbf{u}\in T_{q,\delta}^n$

$$Q_s^{\mathbf{u}}(Y_{mjl}) \leqslant 2^{-S(E|YU,\omega_s^{\otimes n})+n\Gamma'(\delta)}\mathbb{1}_{\mathbb{C}^d}. \tag{4.31}$$

Let $n>2$. The hypotheses of Theorem 87 are therefore satisfied with $\epsilon = \epsilon_{0,n} := 2^{-n\Gamma'(\delta)/6}$ and $\mu = 2^{-S(E|YU,\omega_s^{\otimes n})+n\Gamma'(\delta)}$. Since $u_m \in T_{q,\delta}^n, \forall m\in[M_{0,n}]$, for the event

$$\mathbf{C}_{s,m,j} := \left\{ \left\| \frac{1}{L_n}\sum_{l=1}^{L_n} Q_s^{u_m}(Y_{mjl}) - \Theta_s^{u_m} \right\|_1 > \epsilon_{0,n} \right\},$$

we have

$$\Pr\big[\mathbf{C}_{s,m,j}\big] \leqslant 2^{S(E|U,\omega_s^{\otimes n})+n\Delta(\delta)} \times \exp\Big(-L_n\frac{\epsilon_{0,n}^3}{2\ln 2\cdot 2^{S(E|U,\omega_s^{\otimes n})-S(E|YU,\omega_s^{\otimes n})+n(\Delta(\delta)-\Gamma'(\delta))}}\Big)$$
$$\leqslant 2^{n(\log\dim\mathcal{H}_E+\Delta(\delta))} \times \exp\Big(-L_n\frac{\epsilon_{0,n}^3}{2\ln 2\cdot 2^{I(Y;E|U,\omega_s^{\otimes n})+n(\Delta(\delta)-\Gamma'(\delta))}}\Big).$$

Applying the union bound, for all $s \in S_n, j \in J_n, m \in [M_{0,n}]$ we have

$$\Pr\big[\mathbf{C} := \bigcup_{s,j,m}\mathbf{C}_{s,m,j}\big] \tag{4.32}$$
$$\leqslant J_n M_{0,n}|S_n| 2^{n(\log\dim\mathcal{H}_E+\Delta(\delta))} \times \exp\Big(-L_n\frac{\epsilon_{0,n}^3}{2\ln 2\cdot 2^{I(Y;E|U,\omega_s^{\otimes n})+n(\Delta(\delta)-\Gamma'(\delta))}}\Big). \tag{4.33}$$

From (4.26) and (4.32), we have

$$\Pr\big[\mathbf{C}\cup\mathbf{B}\cup\mathbf{A}\big] \leqslant 2\sqrt{\epsilon_n} + J_n M_{0,n} \times |S_n| 2^{n(\log\dim\mathcal{H}_E+\Delta(\delta))}$$
$$\times \exp\Big(-L_n\frac{\epsilon_{0,n}^3}{2\ln 2\cdot 2^{I(Y;E|U,\omega_s^{\otimes n})+n(\Delta(\delta)-\Gamma'(\delta))}}\Big). \tag{4.34}$$

Finally, given (4.24), we have

$$\exp\Big(-L_n\frac{\epsilon_{0,n}^3}{2\ln 2\cdot 2^{I(Y;E|U,\omega_s^{\otimes n})+n(\Delta(\delta)-\Gamma'(\delta))}}\Big) \leqslant \exp\Big(-\frac{\epsilon_{0,n}^3 2^{n\Gamma'(\delta)}}{2\ln 2\cdot}\Big), \tag{4.35}$$

which gives us a double exponential decay given that $\epsilon_{0,n} = 2^{-n\Gamma'(\delta)/6}$. Inserting (4.35) in (4.34), we conclude that we can find one realization $\{y_{mjl}\}_{(m,j,l)\in[M_{0,n}]\times[J_n]\times[L_n]}$ of Y, such that

$$\min_{s\in S_n}\frac{1}{M_{0,n}J_nL_n}\sum_{m,j,l}\mathrm{tr}(\tilde{W}_{B,s}^{\otimes n}(y_{mjl})D_{B,mjl}) \geqslant 1-\sqrt{\epsilon}_n, \tag{4.36}$$

$$\min_{s\in S_n}\frac{1}{M_{0,n}J_nL_n}\sum_{m,j,l}\mathrm{tr}(\tilde{W}_{E,s}^{\otimes n}(y_{mjl})D_{E,m}) \geqslant 1-\sqrt{\epsilon}_n \tag{4.37}$$

and

$$\max_{s\in S_n}\max_{m,j}\left\|\frac{1}{L_n}\sum_{l=1}^{L_n}Q_s^{u_m}(y_{mjl})-\Theta_s^{u_m}\right\|_1 \leqslant \epsilon_{0,n}. \tag{4.38}$$

Consider the stochastic encoder $E(\cdot|m,j) := \frac{1}{L_n}\sum_{l\in[L_n]}p_{X|Y}^n(\cdot|y_{mjl})$ and POVM $(D_{B,mj} :=$

$\sum_{l\in[L_n]} D_{mjl})_{m,j}$. Therefore with $M_{c,n} = J_n$, we have

$$\begin{aligned}\inf_{s\in S} \frac{1}{M_{0,n}M_{c,n}} & \sum_{m_0\in[M_0],m_c\in[M_{c,n}]} \mathrm{tr}(\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|m_0,m_c)W_{B,s}^{\otimes n}(\mathbf{x})D_{B,m_0,m_c}) \\ &= \frac{1}{M_{0,n}M_{c,n}} \sum_{m_0,m_c} \mathrm{tr}(\frac{1}{L_n}\sum_{\mathbf{x}\in\mathcal{X}^n}\sum_{l\in[L_n]} p^n_{X|Y}(\cdot|y_{m_0m_cl})W_{B,s}^{\otimes n}(\mathbf{x}) \sum_{l'\in[L_n]} D_{B,m_0m_cl'}) \\ &\geqslant 1-\sqrt{\epsilon}_n - 2n\tau_n, \end{aligned} \tag{4.39}$$

where in the last line we have inserted the bound from (4.36) and observed that the error due to $\{W_s\}_{s\in S_n}$ can only be $2n\tau_n$ less than the error due to $\mathcal{W}$. By the same line of reasoning we have

$$\begin{aligned}\inf_{s\in S} \frac{1}{M_{0,n}M_{c,n}} & \sum_{m_0\in[M_0],m_c\in[M_{c,n}]} \mathrm{tr}(\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|m_0,m_c)W_{E,s}^{\otimes n}(\mathbf{x})D_{E,m_0}) \\ &= \frac{1}{M_{0,n}M_{c,n}} \sum_{m_0,m_c} \mathrm{tr}(\frac{1}{L_n}\sum_{\mathbf{x}\in\mathcal{X}^n}\sum_{l\in[L_n]} p^n_{X|Y}(\cdot|y_{m_0m_cl})W_{E,s}^{\otimes n}(\mathbf{x}) \sum_{l'\in[L_n]} D_{E,m_0}) \\ &\geqslant 1-\sqrt{\epsilon}_n - 2n\tau_n. \end{aligned} \tag{4.40}$$

The 5th claim in the statement of the lemma related to the security criterion requires upper bounding $\sup_{s\in S} I(M_c;E|M_0,\sigma_{s,n})$ for all $s\in S_n$, that is done in the following. First we observe that for all $s\in S$

$$I(M_c;E|M_0,\sigma_{s,n}) = \frac{1}{M_{0,n}}\sum_{m_0\in[M_{0,n}]} I(M_c;E,\sigma_{s,n}^{m_0}), \tag{4.41}$$

with

$$\sigma_{s,n}^{m_0} := \frac{1}{M_{c,n}}\sum_{m_c\in[M_{c,n}]} \otimes \sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|m_0,m_c)W^{\otimes n}(\mathbf{x}).$$

We continue upper-bounding the mutual information on the right hands side of (4.41) for each $m_0\in[M_{0,n}]$. We note that for all $s\in S_n$

$$\begin{aligned} I(M_c;E,\sigma_{s,n}^{m_0}) &= S\Bigg(\frac{1}{M_{c,n}}\sum_{m_c\in[M_{c,n}]}\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|m_0,m_c)W_{E,s}^{\otimes n}(\mathbf{x})\Bigg) \\ &\quad - \frac{1}{M_{c,n}}\sum_{m_c\in[M_{c,n}]} S\Bigg(\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|m_0,m_c)W_{E,s}^{\otimes n}(\mathbf{x})\Bigg) \\ &= S\Bigg(\frac{1}{M_{c,n}L_n}\sum_{j\in[M_{c,n}],l\in[L_n]} \tilde{W}_{E,s}^{\otimes n}(y_{m_0jl})\Bigg) \\ &\quad - \frac{1}{M_{c,n}}\sum_{m_c\in[M_{c,n}]} S\Bigg(\sum_{l\in[L_n]} \tilde{W}_{E,s}^{\otimes n}(y_{m_0jl})\Bigg) \end{aligned} \tag{4.42}$$

Notice that, given (4.29) and (4.38) and the triangle inequality we have for all $s \in S_n$

$$\| \sum_{l \in [L_n]} \tilde{W}_{E,s}^{\otimes n}(y_{m_0 jl}) - \varTheta_s^{u_{m_0}} \|_1 \leqslant \epsilon_{0,n} + \epsilon_{1,n}. \tag{4.43}$$

Applying Lemma 209 with $\delta = \epsilon_{0,n} + \epsilon_{1,n}$, given (4.43) and (4.42) we obtain

$$I(M_c; E, \sigma_{s,n}^{m_0}) \leqslant 2\left(n(\epsilon_{0,n} + \epsilon_{1,n}) \log \dim(\mathcal{H}_E) + h(\epsilon_{0,n} + \epsilon_{1,n})\right). \tag{4.44}$$

Inserting this into (4.41), we obtain the same upper bound on the conditional mutual information quantity on the left hand side for all $s \in S_n$. Given properties of the τ-net (Lemma 197), applying Lemma 210 with $\delta = 2n\tau_n$ we obtain

$$\begin{aligned} \sup_{s \in S} I(M_c; E|M_0, \sigma_{s,n}) &\leqslant \max_{s \in S_n} I(M_c; E|M_0, \sigma_{s,n}) \\ &\quad + 2\left(2n^2\tau_n \log \dim(\mathcal{H}_B) + (1 + 2n\tau_n)h(2n\tau_n/1 + 2n\tau_n)\right) \\ &\leqslant 2\left(n(\epsilon_{0,n} + \epsilon_{1,n}) \log \dim(\mathcal{H}_E) + h(\epsilon_{0,n} + \epsilon_{1,n})\right) \\ &\quad + 2\left(2n^2\tau_n \log \dim(\mathcal{H}_B) + (1 + 2n\tau_n)h(2n\tau_n/1 + 2n\tau_n)\right). \end{aligned} \tag{4.45}$$

Given the upper bound on S_n, choosing $\nu = \frac{1}{8n|\mathcal{X}| \dim(\mathcal{H}_B \otimes \mathcal{H}_E)} \log \epsilon_n$, we obtain exponential decay of the right hand sides of (4.39) and (4.40). Also, with this value of τ_n and choosing large enough values of n, (4.45) gives us the 5th claim of the statement.

Proof 93 (Proof of Lemma 88) *According to Lemma 89,*

$$(R_0, R_c) \in \bigcup_p \hat{C}^{(1)}(\mathcal{J}, p, 1)$$

implies $(R_1, R_c) \in C_{BCC}(\mathcal{J})$. Using standard double-blocking and time sharing arguments, for each $l \in \mathbb{N}$,

$$(R_0, R_c) \in \mathrm{cl}\left(\bigcup_{l=1}^{\infty} \bigcup_p \frac{1}{l} \hat{C}^{(1)}(\mathcal{J}, p, l)\right),$$

implies $(R_0, R_c) \in C_{BCC}(\mathcal{J})$.

Here, in order to construct private codes for the broadcast channel, we first generated suitable random message transmission codes for the broadcast channel without imposing privacy constraints (Lemma 90). This was done by establishing suitable bounds for random universal "superposition codes". Subsequent application of a covering principle these codes where transformed to fulfill the security criterion in Lemma 89. Beside technical obstacles to construct superposition codes for cq broadcast channels which are robust regarding uncertainty of the channel state, the approach is rather traditional and even dates back to classical information theory (see e.g. [Csiszár and Körner(2011a)] for a general discussion, the classical counterpart to our considerations can be found in [Schaefer and

Boche(2014b)]).

4.3.2. TPC codes

In this section, we prove the following lemma.

Lemma 94 *Let $\mathcal{W} := \{W_s\}_{s\in S} \subset CQ(\mathcal{X}, \mathcal{H}_B \otimes \mathcal{H}_E)$ be any compound cqq broadcast channel. It holds*

$$C_{TPC}[\mathcal{W}] \supset \mathrm{cl}\left(\bigcup_{l=1}^{\infty}\bigcup_{p}\frac{1}{l}C^{(1)}(\mathcal{W}, p, l)\right),$$

where the second union is taken over all $p_{VYX} \in \mathcal{P}(\mathcal{V}\times\mathcal{Y}\times\mathcal{X}^l)$ such that random variable $V - Y - X$ form a Markov chain and alphabets $\mathcal{V}$ and $\mathcal{Y}$ are finite.

The main step towards proving Lemma 94, is the following statement.

Lemma 95 (Broadcast channel with confidential messages) *Let $\mathcal{W} := \{W_s\}_{s\in S} \subset CQ(\mathcal{X}, \mathcal{H}_B \otimes \mathcal{H}_E)$ be any compound cqq broadcast channel. For $p_{VYX} \in \mathcal{P}(\mathcal{V}\times\mathcal{Y}\times\mathcal{X})$ where $V - Y - X$ form a Markov chain and $\delta, \epsilon > 0$, there exists $n_0 \in \mathbb{N}$, such that for $n > n_0$, we find an $(n, M_{1,n}, M_{c,n})$ TPC code $\{E(\cdot|m), D_{B,m}, D_{E,m_1}\}_{m=(m_1,m_c)\in[M_{1,n}]\times[M_{c,n}]}$ with*

1. $\frac{1}{n}\log M_{1,n} \geqslant \inf_{s\in S} I(V; B, \omega_s) - c\delta$,
2. $\frac{1}{n}\log M_{c,n} \geqslant \inf_{s\in S} I(Y; B|V, \omega_s) - \sup_{s\in S} I(Y; E|V, \omega_s) - c\delta$ *with some constants $c > 0$ and ω_s defined by (4.4).*
3. $\inf_{s\in S}\frac{1}{|\mathbf{M}|}\sum_{\mathbf{m}\in\mathbf{M}}\sum_{\mathbf{x}\in\mathcal{X}^n} E(\mathbf{x}|\mathbf{m})\mathrm{tr}[W_{B,s}^{\otimes n}(\mathbf{x})D_{B,\mathbf{m}}] \geqslant 1 - \epsilon$
4. $\sup_{s\in S} I(M_c; E|M_1, \sigma_{s,n}) \leqslant \epsilon$

with state $\sigma_{s,n}$ defined by (4.2).

Applying standard double-blocking arguments on Lemma 95, will prove Lemma 94. We prove Lemma 95 in two steps. At first, we prove the following random coding result, that guarantees reliable decoding of public messages by Bob and Eve, and reliable decoding of public messages by Bob. In the next step, we apply privacy amplification arguments on the public part of the codebook, to achieve the desired confidential message transmission rate.

Lemma 96 *Let $\mathcal{W} := \{W_s\}_{s\in S} \subset CQ(\mathcal{Y}, \mathcal{H}_B \otimes \mathcal{H}_E)$ be any compound cqq broadcast channel and $\mathcal{V}$ be a finite alphabet. For any $\delta > 0$, $q \in \mathcal{P}(\mathcal{V}), r(\cdot|v) \in \mathcal{P}(\mathcal{Y})$, $v \in \mathcal{V}$ and large enough values of n, the following exist.*

- *A family $(v_m)_{m\in[M_{2,n}]}$ of words with $v_m \in T_{q,\delta}^n$.*

- *A map $y : (y_{ij})_{(i,j)\in[M_{1,n}]\times[M_{2,n}]} \mapsto (D_{B,ij}(y))_{(i,j)\in[M_{1,n}]\times[M_{2,n}]}$, such that $(D_{B,ij}(y))_{(i,j)\in[M_{1,n}]\times[M_{2,n}]}$ $\mathcal{L}(\mathcal{H}_B^{\otimes n})$ is a POVM and for any family $Y = (Y_{ij})_{(i,j)\in[M_{1,n}]\times[M_{2,n}]}$ of random variables such that for each $m \in [M_{1,n}]$, $Y^m = (Y_{mj})_{j\in[M_{2,n}]}$ is distributed i.i.d according to $r'(\cdot|v_m)$ we have*

$$\frac{1}{n}\log M_{1,n} \geqslant \inf_{s\in S} I(V;B,\omega_s) - c\delta,$$

$$\frac{1}{n}\log M_{2,n} \geqslant \inf_{s\in S} I(Y;B|V,\omega_s) - c\delta,$$

$$\mathbb{E}_Y\left[\inf_{s\in S}\frac{1}{M_{1,n}M_{2,n}}\sum_{(m,i)\in[M_{1,n}]\times[M_{2,n}]}\operatorname{tr}(W_{B,s}^{\otimes n}(Y_{mi})D_{B,mi}(Y)\right] \geqslant 1-\epsilon_n,$$

with $\epsilon_n \to 0$ exponentially, constant $c > 0$ and $\omega_s = \sum_{v\in\mathcal{V}} q(u)\,|v\rangle\langle v|\otimes r(y|v)\,|y\rangle\langle y|\otimes W_s(y)$.

Proof 97 *The proof is done by following exactly the lines in proof of Lemma90, except that here $\gamma = \{B\}$.*

Proof 98 (Proof of Lemma 95) *The proof follows by applying the privacy amplification arguments in the proof of Lemma 89, on $[M_{2,n}]$ part of the messages in Lemma 96. It is clear that here, we only consider upper bounding the probability of events corresponding to events* **A** *and* **C** *in the proof of that Lemma 89, and drop (4.25).*

Proof 99 (Proof of Lemma 94) *According to Lemma 95,*

$$(R_1, R_c) \in \bigcup_p C^{(1)}(\mathcal{J}, p, 1)$$

implies $(R_1, R_c) \in C_{TPC}(\mathcal{J})$. Using standard double-blocking and time sharing arguments, for each $l \in \mathbb{N}$,

$$(R_1, R_c) \in \operatorname{cl}\left(\bigcup_{l=1}^{\infty}\bigcup_p \frac{1}{l}C^{(1)}(\mathcal{J}, p, l)\right),$$

implies $(R_1, R_c) \in C_{TPC}(\mathcal{J})$.

4.4. Outer bounds for the capacity regions

In this section, we consider the ”converse” bounds stated in Theorem 77 and Theorem 81. The arguments of proof turn out to be fairly standard. Therefore, we restrict ourselves to providing proof details regarding the outer bound to the BCC capacity regions from Theorem 77.

Proposition 100 *Let $\mathcal{W} := \{W_s\}_{s\in S}$, $W_s : \mathcal{X} \to \mathcal{S}(\mathcal{H}_B \otimes \mathcal{H}_E)$, $(s \in S)$ be a set of cqq channels. It holds*

$$C_{BCC}[\mathcal{W}] \subset \operatorname{cl}\left(\bigcup_{l=1}^{\infty}\bigcup_p \frac{1}{l}\hat{C}^{(1)}(\mathcal{W}, p, l)\right).$$

The second union is taken over all $p_{UYX} \in \mathcal{P}(\mathcal{U} \times \mathcal{Y} \times \mathcal{X}^l)$ such that random variable $U - Y - X$ form a Markov chain and alphabets $\mathcal{U}$ and $\mathcal{Y}$ are finite.

Proof 101 *Let $(\mathcal{C}_n)_{n\in\mathbb{N}}$ be a sequence of $(n, M_{1,n}, M_{c,n})$ BCC codes for $\mathcal{W}$ such that with a sequence $e_n \to 0$, $(n \to \infty)$ for all $s \in S$ $\overline{e}_B(\mathcal{C}_n, W_s^{\otimes n})$, $\overline{e}_E(\mathcal{C}_n, W_s^{\otimes n})$ and $I(M_{c,n}; E^n|M_{0,n}, \sigma_{s,n})$ are simultaneously upper-bounded by e_n. While we fix the blocklength for a moment (and suppress the index n), we consider for each $s \in S$ the quadruple $(M_0, M_c, M_0^{(s)}, M_c^{(s)})$ of random variables, where $M_0^{(s)}, M_c^{(s)}$ belong to the common and confidential messages decoded by B after transmission with $W_s^{\otimes n}$. Note, that $\Pr((M_0, M_c) \neq (M_0^{(s)}, M_c^{(s)})) \leqslant \epsilon_n$ is true by assumption. It holds*

$$\log M_0 \;=\; H(M_0) \;=\; I(M_0; M_0^{(s)}) + H(M_0|M_0^{(s)}) \;\leqslant\; I(M_0; B^n, \sigma_{s,n}) + \epsilon_n \cdot \log M_0. \tag{4.46}$$

The second of the above equalities is the chain rule for the mutual information. The last inequality stems from application of Fano's lemma and the Holevo bound. A similar calculation for the second receiver leads us to the inequality

$$\log M_0 \;\leqslant I(M_0; E^n, \sigma_{s,n}) + \epsilon_n \cdot \log M_0. \tag{4.47}$$

Maximizing over all $s \in S$ in (4.46) and (4.47) and combining the resulting inequalities gives the bound

$$\log M_0 \;\leqslant\; \min\left\{\sup_{s\in S} I(M_0; B^n, \sigma_{s,n}),\; \sup_{s\in S} I(M_0; E^n, \sigma_{s,n})\right\} \;+\; \epsilon_n \log M_0.$$

In order to derive a bound on M_c, we notice the inequality

$$\log M_0 \cdot M_c \;\leqslant\; I(M_0 M_c; B^n, \sigma_{s,n}) + \epsilon_n \cdot \log M_0 M_c. \tag{4.48}$$

The chain rule for the quantum mutual information implies

$$I(M_0 M_c; B^n, \sigma_{s,n}) - \log M_0 \;\leqslant I(M_0 M_c; B^n, \sigma_{s,n}) - I(M_0; B^n, \sigma_{s,n}) \;=\; I(M_c; B^n|M_0, \sigma_{s,n}).$$

Combining the above inequality with (4.48) and rearranging terms give us the inequality

$$\log M_c \;\;\leqslant\; I(M_c; B^n|M_0, \sigma_{s,n}) + \epsilon_n \cdot \log M_0 M_c.$$

Maximizing both sides of the inequality and adding the nonnegative term $\epsilon_n - \sup_{s\in S} I(M_c; E^n|M_0, \sigma_{s,n})$ to the right hand side of the result, we obtain

$$\log M_c \;\leqslant\; \sup_{s\in S} I(M_c; B^n|M_0, \sigma_{s,n}) \;-\; \sup_{s\in S} I(M_c; E^n|M_0, \sigma_{s,n}) \;+\; \epsilon_n(\log M_0 \cdot M_c + 1). \tag{4.49}$$

Let $\delta > 0$ be arbitrary and n_0 large enough for $\epsilon_n(\log M_0 \cdot M_c) \leqslant \delta$ to hold. It is clear, for each $n > n_0$, $(\frac{1}{n}\log M_{0,n}, \frac{1}{n}\log M_{c,n})$ is contained in

$$\bigcup_{l>n_0} \frac{1}{n} \bigcup_{p} \hat{C}^{(1)}(\mathcal{W}, p, n)_\delta \ \subset \ \left[\bigcup_{l=1}^{\infty} \bigcup_{p} \frac{1}{l} \hat{C}^{(1)}(\mathcal{W}, p, l)\right]_\delta , \tag{4.50}$$

where A_δ is the δ-blowup of A for each $\delta > 0$ and $A \in \mathbb{R}_0^+ \times \mathbb{R}_0^+$, i.e

$$A_\delta := \{y \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ : \exists x \in A : \| x - y \| \leqslant \delta\}.$$

Since δ was an arbitrary positive number, we are done.

Proposition 102 *Let $\mathcal{W} := \{W_s\}_{s \in S}$, $W_s : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H}_B \otimes \mathcal{H}_E)$, $(s \in S)$ be a set of cqq channels. It holds*

$$C_{TPC}[\mathcal{W}] \ \subset \ \mathrm{cl}\left(\bigcup_{l=1}^{\infty} \bigcup_{p} \frac{1}{l} C^{(1)}(\mathcal{W}, p, l)\right).$$

The second union is taken over all $p_{VYX} \in \mathcal{P}(\mathcal{V} \times \mathcal{Y} \times \mathcal{X}^l)$ such that random variable $V - Y - X$ form a Markov chain and alphabets $\mathcal{V}$ and $\mathcal{Y}$ are finite.

Proof 103 *The proof can be conducted following exactly the same strategy as in the proof of Proposition 100, and therefore is left to the reader. The only modification is, that there is no need for E to decode the message M_1 (opposed to the case of M_0 in the proof of Proposition 100). This leads to the bound*

$$\log M_1 \ \leqslant \ \sup_{s \in S} I(M_0; B^n, \sigma_{s,n}) + \epsilon_n \log M_1.$$

on the number public messages in the code.

4.5. BCC and TPC capacities of compound quantum broadcast channels

In this section we extend our results to the "full quantum" setting where the receivers input quantum systems to the channels, i.e. the transition maps of the channels are c.p.t.p. maps instead of cq channels. Since the message transmission tasks we aim to perform are after all of a classical nature, the corresponding coding theorems can be proven applying the results from earlier chapters.

Explicitely we apply the results of the preceding sections to derive codes for full quantum broadcast channels. For the remainder of this section, we fix an arbitrary set $\mathcal{J}$:=

$\{\mathcal{N}_s\}_{s\in S}$, where

$$\mathcal{N}_s : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_E)$$

is a c.p.t.p. map for each $s \in S$. Traditionally, the c.p.t.p. map $\mathcal{N}_s$ is assumed to be an isometric channel, namely a Stinespring isometry to a given channel connecting A and B. This way of defining the channel is fairly justified, since it naturally equips E with the strongest abilities when attacking the confidential transmission goals of the remaining parties. However, dropping this assumption on the channel does not complicate any subsequent arguments.

In what follows, we consider the BCC scenario. Corresponding considerations regarding the TPC scenario are easily extrapolated and are hence left to the reader.

Definition 104 (BCC codes) *An* (n, M_0, M_c) BCC code *for* $\mathcal{J}$ *for channels in* $\mathcal{C}(\mathcal{H}_A, \mathcal{H}_B \otimes \mathcal{H}_E)$ *is a family* $\mathcal{C} = (V(m), D_{B,m}, D_{E,m_0})_{m\in\mathbf{M}}$ *with* $\mathbf{M} := [M_0] \times [M_c]$*, where* $(D_{B,m})_{m\in\mathbf{M}}$ *and* $(D_{E,m_0})_{m_0\in[M_0]}$ *are POVMs on* $\mathcal{H}_B^{\otimes n}$ *resp.* $\mathcal{H}_E^{\otimes n}$ *and* $V(m)$ *is a state on* $\mathcal{H}_A^{\otimes n}$ *for each* m*.*

The average transmission errors for the receivers B, and E with channel $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_E)$ and (n, M_0, M_c)-code $\mathcal{C}$ are defined by

$$\bar{e}_B(\mathcal{C}, \mathcal{N}^{\otimes n}) \ := \ \frac{1}{|\mathbf{M}|} \sum_{m\in\mathbf{M}} \mathrm{tr} D_{B,m}^c \mathcal{N}^{\otimes n}(V(m)),$$

and

$$\bar{e}_E(\mathcal{C}, \mathcal{N}^{\otimes n}) \ := \ \frac{1}{|\mathbf{M}|} \sum_{m\in\mathbf{M}} \mathrm{tr} D_{E,m_0}^c \mathcal{N}^{\otimes n}(V(m)).$$

By replacing the code and errors the definitions of achievable rate pairs can be directly guessed from Definition 76 (the notational ambiguity should cause no misunderstandings since the set $\mathcal{J}$ determines whether the classical-quantum or quantum broadcast channel scenario are considered.) We denote the corresponding *BCC capacity region* by $C_{BCC}[\mathcal{J}]$. We moreover define $\hat{C}^{(1)}(\mathcal{J}, p, l, (\rho_y)_{y\in\mathcal{Y}})$ the set of all points in $\mathbb{R}^2$ which fulfill the inequalities

$$\begin{aligned} 0 \ &\leqslant \ R_0 \leqslant \inf_{s\in S} \ \min \ \{I(U; B, \omega_s), \ I(U; E, \omega_s)\} \qquad\qquad \text{and} \\ 0 \ &\leqslant \ R_c \leqslant \inf_{s\in S} I(Y; B|U, \omega_s) \ - \ \inf_{s\in S} I(Y; E|U, \omega_s) \end{aligned}$$

where we understand the entropic quantities above as being evaluated on the ccq state

$$\omega_s \ := \ \omega(\mathcal{N}_s, p, l) \ := \sum_{u\in\mathcal{U}, y\in\mathcal{Y}} P_{UY}(u, y) \cdot |u, y\rangle\langle u, y| \otimes \mathcal{N}_s^{\otimes l}(\rho_y)$$

for each $s \in S$.

Theorem 105 *It holds*

$$C_{BCC}[\mathcal{J}] \;=\; \mathrm{cl}\left(\bigcup_{l=1}^{\infty}\bigcup_{p}\frac{1}{l}\hat{C}^{(1)}(\mathcal{J}, p, l)\right)$$

The second union is taken over all $p_{UYX} \in \mathcal{P}(\mathcal{U} \times \mathcal{Y} \times \mathcal{X}^l)$ such that random variable $U - Y - X$ distributed accordingly, form a Markov chain and alphabets $\mathcal{U}$ and $\mathcal{Y}$ are finite.

Proof 106 *The proof of achievability is easily performed by referring to the corresponding result for ccq broadcast channels. Namely, if we fix $l \in \mathbb{N}$, probability distributions P_U and $P_{Y|U}$ and a family $(\rho_y)_{y\in\mathcal{Y}}$ of quantum states on $\mathcal{H}_A^{\otimes n}$ we have*

$$\omega(\mathcal{N}_s, p, l, (\rho_y)_{y\in\mathcal{Y}}) \;=\; \sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}} P_U(u)\cdot P_{Y|U}(y|u)\,|u,y\rangle\langle u,y| \otimes \mathcal{N}_s^{\otimes l}(\rho_y) \;=\; \omega(\tilde{V}_s, p, 1)$$

with an effective cqq channel with signals $\widetilde{V}_s(y) \;:=\; \mathcal{N}_s^{\otimes l}(\rho_y)$, $(y \in \mathcal{Y})$. As a consequence, $\frac{1}{l}\hat{C}^{(1)}(\mathcal{J}, p, l, (\rho_y)_{y\in\mathcal{Y}}) = \frac{1}{l}\hat{C}^{(1)}(\{\widetilde{V}_s\}_{s\in S}, p, 1)$. We know from Theorem 77, that each point on the r.h.s. of the preceding inequality is achievable. To prove the converse, we assume, that $\mathcal{C}_n := (D_{B,m}, D_{E,m_0}, V(m))_{m\in\mathbf{M}}$ is an (n, M_0, M_c)-code with

$$\overline{e}_B(\mathcal{C}_n, \mathcal{N}_s^{\otimes n}),,\; \overline{e}_E(\mathcal{C}_n, \mathcal{N}_s^{n}), \qquad \textit{and} \qquad I(M_c; E|M_0, \sigma_{s,n})$$

are simultaneously bounded by $\epsilon_n \in (0,1)$. Note, that the mutual information quantity above is evaluated on the code state

$$\sigma_{s,n} := \frac{1}{\mathbf{M}}\sum_{m\in\mathbf{M}} |m\rangle\langle m| \otimes \mathcal{N}_s^{\otimes n}(V(m)).$$

Using the above bounds and repeating the corresponding steps from the proof of Proposition 100, we obtain the inequalities

$$\log M_0 \leqslant \min\left\{\sup_{s\in S} I(M_0; B^n, \sigma_{s,n}),\; \sup_{s\in S} I(M_0; E^n, \sigma_{s,n})\right\} \;+\; \epsilon_n \log M_0.$$

and $\log M_c \leqslant I(M_c; B^n|M_0, \sigma_{s,n}) + \epsilon_n \cdot \log M_0 M_c$ The remaining steps directly carry over from the cqq converse.

5. Computability aspects

We analyze general achievability (lower-) and converse (upper-) bounds on the ϵ-capacity function from a fundamental point of view by studying whether or not such bounds can be computed by any algorithms in principle (without putting any constraints on the computational complexity of such algorithms). For this purpose, the concept of Turing machines is used, which provides the fundamental performance limits of digital computers. To this end, computable continuous functions are studied and properties of computable sequences of such functions are identified. Subsequently, these findings are exemplary applied to the ϵ-capacity of the two-state compound channel. It is shown that there are examples for which this function (derived here under two capacity notions) is a non-computable function of its error input. As a result, it is stated that either the achievability or converse yields a non-computable bound. The crucial consequence is that the ϵ-capacity cannot be characterized by a finite-letter entropic expression. We also consider a less restrictive conditions of decidability for the derived capacity functions and obtain negative results. The channel examples that give us the general non-computability of the capacity functions, are those that prove communicating parties can have asymptotic gains by pre-shared entanglement or randomness. This gain cannot necessarily be harnessed by a digital computer due to general non-computability of the capacity functions.

5.1. Introduction

For the ϵ-capacity of compound channels [Blackwell et al.(1959)Blackwell, Breiman, and Thomasian, Wolfowitz(1960), Ahlswede(2015)], it is shown and argued that either the achievability or converse (or both) must result in a non-computable lower or upper bound, respectively. Accordingly, it is impossible that both achievability and converse are effectively computable at the same time and, as a consequence, we cannot find a finite-letter entropic characterization for the ϵ-capacity. This has important implications on the question of the existence of a strong conserve and the second order coding rate. Both questions cannot be answered algorithmically as we will demonstrate.

The asymptotic bound for error-correcting codes is a fundamental and open problem in coding theory [Tsfasman et al.(2007)Tsfasman, Vladut, and Nogin, Joyner and Kim(2011)]. Despite tremendous effort, attempts to characterize this function have failed. Except for some trivial points, not much is known about this function and its behavior. It is conjectured that this function is indeed a non-computable function. With the previous findings, this explains the difficulties as either the lower or upper bound on the asymptotic bound must be non-computable. Thus, it is impossible to derive computable lower and upper bounds that are asymptotically tight.

The underlying *computability framework* is introduced in Section 5.2. A Turing machine is a mathematical model of an abstract machine that manipulates symbols on a strip of tape according to certain given rules. It can simulate any given algorithm and therewith provides a simple but very powerful model of computation. Turing machines have no limitations on computational complexity, unlimited computing capacity and storage, and execute programs completely error-free. They are further equivalent to the von Neumann-architecture without hardware limitations cf. also [Avigad and Brattka(2014), Gödel(1930), Gödel(1934), Kleene(1952), Minsky(1961)]. Accordingly, Turing machines provide fundamental performance limits for today's digital computers. Since bounds on the capacity are usually evaluated and often plotted on digital computers, Turing machines are the ideal concept to study whether or not such upper and lower bounds can be found algorithmically in principle (without putting any constraints on the computational complexity of such an algorithm).

The underlying *computability framework* is introduced in Section 5.2. Of particular interest here are *computable continuous functions* [Pour-El and Richards(2017)] since such functions can be effectively approximated by computable polynomial sequences. To this end, Section 5.2 also studies further properties and insights of computable sequences of such computable continuous functions. In Section 5.3 we derive our capacity results for the ϵ-capacity of the compound channel with two channels present in the uncertainty set, that give us significant examples of channels that are fundamental to our computability results in later sections. In Section 5.4, the findings of Section 5.2 are applied to these examples. In Section 5.5, we use the results from previous sections to prove that either the converse (upper-bound) or achievability (lower-bound) is not algorithmically computable as a function of the error input. We refute computability already at compound channels with two channel states, that is the smallest possible uncertainty by assuming the not-so-realistic tolerated error of 1/2. As mentioned before however, practical coding strategies must be robust to compound channels of infinite cardinality, that which carries our negative results to arbitrarily small values of tolerated error (see [Ahlswede(2015)] for examples of compound channels with more than two channel states).

In Section 5.6, we consider the less restrictive conditions of *decidability* and *semi-decidability*, and demonstrate that even this conditions are not necessarily satisfied by ϵ-capacity as a function of the error input. In Section 5.7 similar statements are made on computability of the capacities of assisted scenarios where the communicating parties have access to pre-shared entanglement or common randomness. We show the existence of examples where such resources improve the ϵ-capacity of compound channels. These examples are therefore significant as they refute the generality of statements that deny the use of entanglement when it pertains to classical communication. This improvement of capacity should be looked at with skepticism however, as at least in the case of common-randomness assistance, the resulting function is in general non-computable. The gain in capacity resulting from pre-shared resources, is in other words, cannot necessarily be harnessed by a digital computer.

5.2. Introduction to Turing Machines and computability framework

The formalization of *computability* was established by [Turing(1936), Turing(1937)] and [Church(1936)] by two different approaches. In a mathematical sense, both frameworks are fully equivalent. For brevity, we restrict ourselves to Turing's method: he introduced the idea of what is known today as *Turing machine.*

A Turing machine is a hypothetical machine that manipulates strings of symbols on an infinite *work tape.* The symbols on the tape emanate from a finite *machine alphabet* $\mathcal{S} = \{s_1, s_2, \dots, s_k\} \cup \{\sqcup\}$, where $''\sqcup''$ is the distinguished symbol that marks a *blank space.* Only a finite number of symbols on the tape may differ from the blank space symbol.

Given an initial tape configuration $\mathbf{s} =: \mathbf{s}^0$, the machine sequentially manipulates one symbol on the tape at a time, creating a new tape configuration in each step. Simultaneously, the Turing machine passes trough a sequence of internal *states.* The succeeding pair of tape configuration and internal state depend exclusively on the current tape configuration and the current internal state. This way, we obtain a chain of pairs

$$(\mathbf{s}^0, q^0) \mapsto (\mathbf{s}^1, q^1) \mapsto (\mathbf{s}^2, q^2) \mapsto \dots, \tag{5.1}$$

where $q^0, q^1, q^2, \dots \in \mathcal{Q}$ denotes the current internal state.

The set $\mathcal{Q} := \{q_1, q_2, \dots, q_l\} \cup \{q_\mathrm{S}\} \cup \{q_{\mathrm{H},1}, q_{\mathrm{H},2}, \dots, q_{\mathrm{H,m}}\}$ of internal states contains a distinguished *initial state* q_S such that $q^0 = q_\mathrm{S}$ in (5.1), as well as a set of distinguished *halting states* $q_{\mathrm{H},1}, q_{\mathrm{H},2}, \dots, q_{\mathrm{H,m}}$. Whenever the Turing machine reaches one of the halting states, the computation ends. In this case, we obtain a sequence

$$(\mathbf{s}^0, q_\mathrm{S}) \mapsto (\mathbf{s}^1, q^1) \mapsto (\mathbf{s}^2, q^2) \mapsto \quad \dots \quad \mapsto (\mathbf{s}^n, q^n) \tag{5.2}$$

with $\mathbf{q}^n \in \{q_{\mathrm{H},1}, q_{\mathrm{H},2}, \dots, q_{\mathrm{H,m}}\}$ for some $n \in \mathbb{N}$. The Turing machine is said to *halt* with *output* $(\mathbf{s}^n, q^n, n)$ for *input* $\mathbf{s}^0$. On the other hand, given an input $\mathbf{s}^0$, there may not exist an $n \in \mathbb{N}$ such that $q^n \in \mathcal{Q}$, in wich case the Turing machine continues it's computation infinitely.

For a Turing machine TM, let $\mathcal{T}$ be the set of tape configurations. We denote $\mathcal{D}(TM) \subseteq \mathcal{T}$ the set of inputs for which the Turing machine halts with some output. In this sense, a Turing machine is a mapping

$$TM : \mathcal{D}(TM) \to \mathcal{T} \times \{q_{\mathrm{H},1}, q_{\mathrm{H},2}, \ldots, q_{\mathrm{H},\mathrm{m}}\} \times \mathbb{N}, \mathbf{s} \mapsto TM(\mathbf{s}), \tag{5.3}$$

where $TM(\mathbf{s}) \in \mathcal{T} \times \{q_{\mathrm{H},1}, q_{\mathrm{H},2}, \ldots, q_{\mathrm{H},\mathrm{m}}\} \times \mathbb{N}$ is the output corresponding to input $\mathbf{s}$. We denote $[TM(\mathbf{s})]_{\mathcal{T}}$ the first component, $[TM(\mathbf{s})]_{\mathcal{Q}}$ the second component and $[TM(\mathbf{s})]_{\mathbb{N}}$ the third component of the triple $TM(\mathbf{s})$. With some abuse of notation, we may never the less write $TM(\mathbf{s})$ instead of $[TM(\mathbf{s})]_{\mathcal{T}}$, $[TM(\mathbf{s})]_{\mathcal{Q}}$ or $[TM(\mathbf{s})]_{\mathbb{N}}$, if it is clear from the context which of the three we are referring.

5.2.1. Arithmetic Computations

By encoding the set of n-tuples of natural numbers into the set of tape configurations, we can perform arithmetic calculations on a Turing machine. The simplest sufficient encoding is the *unary numeral system*, with successive components of a given n-tuple $\mathbf{x} \in \mathbb{N}^n$ being separated by a blank space. Let $(g_n)_{n\in\mathbb{N}}$ with $g_n : \mathbb{N}^n \to \mathcal{T}$ for all $n \in \mathbb{N}$ be a family of suitable encodings. A function $f : \mathcal{D}(f) \to \mathbb{N}$ with $\mathcal{D}(f) \subseteq \mathbb{N}^n$ is called *computable*, if there exists a Turing machine TM that satisfies the following properties:

a) If $\mathbf{x} \in \mathcal{D}(f)$ for some $\mathbf{x} \in \mathbb{N}^n$, then $g_n(\mathbf{x}) \in \mathcal{D}(TM)$.

b) If $\mathbf{x} \in \mathbb{N}^n \backslash \mathcal{D}(f)$ for some $\mathbf{x} \in \mathbb{N}^n$, then $g_n(\mathbf{x}) \in \mathcal{T} \backslash \mathcal{D}(TM)$.

c) We have $[TM(g_n(\mathbf{x}))]_{\mathcal{Q}} = g_1(f(\mathbf{x}))$ for all $\mathbf{x} \in \mathcal{D}(f)$.

The set of computable functions, which we denote by $\mathcal{C}^*$, is a true subset of the set $\mathcal{F} := \bigcup_{n=0}^{\infty} \{f : \mathbb{N}^n \to \mathbb{N}\}$ (here, the set $\{f : \mathbb{N}^0 \to \mathbb{N}\} = \mathbb{N}$ denotes constant natural numbers). Other than by the use of Turing machines, the set $\mathcal{C}^*$ is characterized through the axioms of *μ-recursive functions*, in the following simply referred to as *recursive functions*. That is, a function $f : \mathcal{D}(f) \to \mathbb{N}$ with $\mathcal{D}(f) \subseteq \mathbb{N}^n$ is computable if and only if it is a recursive function. A recursive function $f : \mathcal{D}(f) \to \mathbb{N}$ with $\mathcal{D}(f) \subseteq \mathbb{N}^n$ is called *partial* if $\mathcal{D}(f) \neq \mathbb{N}^n$; it is called *total* if $\mathcal{D}(f) = \mathbb{N}^n$.

5.2.2. Recursively enumerable sets and the halting problem

Given a recursive function $f : \mathcal{D}(f) \to \mathbb{N}$ with $\mathcal{D}(f) \subseteq \mathbb{N}^n$, the indicator function

$$\mathbf{1}_{\mathcal{D}(f)} : \mathbb{N}^n \to \{0, 1\}, \mathbf{x} \mapsto \begin{cases} 1 & \text{if } \mathbf{x} \in \mathcal{D}(f) \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

of $\mathcal{D}(f)$ is in general not a recursive function. This insight is known as the *halting problem*, since it is equivalent to determining whether a Turing machine halts for a certain input or not. Strongly related is the concept of *recursively enumerable* and *recursive* sets [Soare(1987)], which, in many cases, is is essentail for deriving the noncomputability of certain mathematical problems.

Definition 107 *A set $A \subseteq \mathbb{N}$ is called* recursively enumerable *if there exists a recursive bijection $\varphi_A : \mathbb{N} \to A$. The mapping φ* enumerates *the set A.*

Remark 108 *A set $A \subseteq \mathbb{N}$ is recursively enumerable if and only if there exists a recursive function $f : \mathcal{D}(f) \to \mathbb{N}$ that satisfies $\mathcal{D}(f) = A$. Furthermore, A is recursively enumerable if and only if there exists a recursive function $f : \mathcal{D}(f) \to \mathbb{N}$ that satisfies $\{n : \exists x \in \mathcal{D}(f) : f(x) = n\} = A$.*

Definition 109 *A set $A \subseteq \mathbb{N}$ is called* recursive *if the indicator function $\mathbf{1}_A : \mathbb{N} \to \{0, 1\}$ of A is a recursive function.*

Remark 110 *A set $A \subseteq \mathbb{N}$ is* recursive *if and only if both of the sets A and $A^{\mathrm{c}} := \mathbb{N} \backslash A$ are recursively enumerable.*

From Remark 108 we know that the domain $\mathcal{D}(f)$ of any recursive function $f : \mathcal{D}(f) \to \mathbb{N}$ is recursively enumerable. On the other hand, we have previously stated that $\mathbf{1}_{\mathcal{D}(f)}$ is not a recursive function in general. Hence, $\mathcal{D}(f)$ may be a non-recursive set. The halting problem thus ensures the existence of sets that are recursively enumerable but not recursive.

5.2.3. Computable real numbers and functions

The basic techniques from Computable Analysis are essential to our work, and will thus be reviewed in the following.

A sequence of rational numbers $(r_n)_{n\in\mathbb{N}}$ is called a *computable sequence of rational numbers* if there exist recursive functions $a, b, s : \mathbb{N} \to \mathbb{N}$ that satisfy

$$r_n = (-1)^{s(n)} \frac{a(n)}{b(n)} \tag{5.5}$$

for all $n \in \mathbb{N}$. Note that this definition implies $b(n) \neq 0$ for all $n \in \mathbb{N}$.

A real number x is said to be *computable* if there exists a computable sequence of rational numbers $(r_n)_{n\in\mathbb{N}}$ such that

$$|x - r_n| < 2^{-n} \tag{5.6}$$

holds true for all $n \in \mathbb{N}$. If the latter is satisfied, we have $\lim_{n\to\infty} r_n = x$. We denote the set of computable real numbers by $\mathbb{R}_c$.

Remark 111 *Let $(\widetilde{r}_n)_{n\in\mathbb{N}}$ be a computable sequence of rational numbers that satisfies $\lim_{n\to\infty} \widetilde{r}_n = x$. Assume there exists a recursive function $\zeta : \mathbb{N} \to \mathbb{N}$ such that*

$$|x - \widetilde{r}_n| < 2^{-M} \tag{5.7}$$

is satisfied for all $n, M \in \mathbb{N}$ with $n \geqslant \zeta(M)$. Then, the sequence $(\widetilde{r}_n)_{n\in\mathbb{N}}$ is said to converge effectively *to x. By setting*

$$r_n := \widetilde{r}_{\zeta(n)}, \tag{5.8}$$

the computable sequence $(r_n)_{n\in\mathbb{N}}$ of rational numbers satisfies $|x - r_n| < 2^{-n}$ for all $n \in \mathbb{N}$. Consequently, a real number x is a computable number if and only if there exists a computable sequence $(\widetilde{r}_n)_{n\in\mathbb{N}}$ of rational numbers that converges effectively to x.

Remark 112 *A computable number x can be represented by a triple (a, b, s) of recursive functions such that the corresponding computable sequence $(r_n)_{n\in\mathbb{N}}$ of rational numbers satisfies $|x - r_n| < 2^{-n}$ for all $n \in \mathbb{N}$. On the other hand, the number x can be represented by a quadruple $(\widetilde{a}, \widetilde{b}, \widetilde{s}, \zeta)$, such that the corresponding computable sequence $(\widetilde{r}_n)_{n\in\mathbb{N}}$ of rational numbers satisfies $|x - \widetilde{r}_n| < 2^{-n}$ for all $n \in \mathbb{N}$.*

In practical applications, it is common to encounter sequences of real numbers, which, in general, may be irrational. For example, an information theoretic channel model may yield a recursive function $f : \mathbb{N} \to \mathbb{N}$ that specifies the number of messages $f(n)$ that can be transmitted trough n successive uses of the channel with respect to some error criterion. As done in previous chapters of this work, this number is turned into a *channel capacity* by setting $x_n := \frac{1}{n}\log_2 f(n)$ and $C := \lim_{n\to\infty} x_n$ (if the limit exists). The number x_n is not necessarily rational, and the sequence $(x_n)_{n\in\mathbb{N}}$ is not necessarily a computable sequence of computable numbers. In order to investigate such sequences with respect to their computability properties, we introduce the concept of computable sequences of computable numbers.

A sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers is called *computable sequence of computable numbers* if there exists a computable double sequence $(r_{n,m})_{n,m\in\mathbb{N}}$ of rational numbers such that

$$|x_n - r_{n,m}| < 2^{-m} \tag{5.9}$$

holds true for all $n, m \in \mathbb{N}$.

Remark 113 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers such that there exists a computable double sequence $(\widetilde{r}_{n,m})_{n,m\in\mathbb{N}}$ of rational numbers as well as a recursive function $\zeta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ that satisfy*

$$|x_n - \widetilde{r}_{n,m}| < 2^{-M} \tag{5.10}$$

for all n, m, M with $m \geqslant \zeta(n, M)$. Setting $r_{n,m} := \widetilde{r}_{n,\zeta(n,m)}$, we obtain a computable double sequence $(r_{n,m})_{n,m\in\mathbb{N}}$ of rational numbers such that $|x_n - r_{n,m}| < 2^{-m}$ holds true for all $n, m \in \mathbb{N}$. Thus, $(x_n)_{n\in\mathbb{N}}$ is a computable sequence of computable numbers that is fully specified by the pair $((r_{n,m})_{n,m\in\mathbb{N}}, \zeta)$.

The set of computable functions $\mathcal{C}^*$ is recursively enumerable. In particular, there exist recursive functions $A, B, S : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, such that for all triples $a, b, s : \mathbb{N} \to \mathbb{N}$ of computable functions, there exists $n \in \mathbb{N}$ such that

$$A(n,m) = a(m) \quad \wedge \quad B(n,m) = b(m) \quad \wedge \quad S(n,m) = s(m) \tag{5.11}$$

is satisfied for all $m \in \mathbb{N}$ (including undefined values). Consequently, given a computable enumeration (A, B, S) of all triples $a, b, s \in \mathcal{C}^*$, we can specify each computable real number x by a single integer $n \in \mathbb{N}$, leading to the concept of Borel-Turing computable functions.

Definition 114 *A function $f : \mathbb{R}_c \to \mathbb{R}_c$ is called* Borel-Turing computable function *if there exists a computable enumeration (A, B, S) of all triples $a, b, s \in \mathcal{C}^*$ as well as a computable function $g : \mathbb{N} \to \mathbb{N}$ such that for all $n \in \mathbb{N}$, $g(n)$ is a representation of $f(x)$ with respect to (A, B, S) whenever n is a representation of x with respect to (A, B, S).*

Remark 115 *In other words, a function $f : \mathbb{R}_c \to \mathbb{R}_c$ is called Borel-Turing computable function if there exists a Turing machine that transforms representations of a computable number x into representations of the computable number $f(x)$.*

The assumptions made in Definition 114 can be weakened to obtain a computability concept without the requirement of a computable enumeration.

Definition 116 *A function $f : \mathbb{R}_c \to \mathbb{R}_c$ is called* Banach-Mazur computable function *if the sequence $(f(x_n))_{n\in\mathbb{N}}$ is a computable sequence of computable numbers whenever the sequence $(x_n)_{n\in\mathbb{N}}$ is a computable sequence of computable numbers.*

Remark 117 *If $f : \mathbb{R}_c \to \mathbb{R}_c$ is Borel-Turing computable, it is also Banach-Mazur computable. The converse does not hold true in general. For an overview on the relations between different notions of computability, the reader is referred to [Avigad and Brattka(2014)] and the introductory textbook [Weihrauch(2000)].*

The concept of Banach-Mazur computability may be extended to allow the investigation of *computable continuity* properties of real-valued functions. A rectangle $\mathbb{I}_c \subset \mathbb{R}_c^d$, $d \in \mathbb{N}$, is called a *computable rectangle* if the boundary values are computable numbers.

Definition 118 *Let $\mathbb{I}_c \subset \mathbb{R}_c^d$, $d \in \mathbb{N}$, be a computable rectangle. A Banach-Mazur computable function $f : \mathbb{I}_c \to \mathbb{R}_C$ is called* computably continuous *if it is* effectively uniformly continuous, *i.e., there exists a recursive function $g : \mathbb{N} \to \mathbb{N}$ such that $|f(x) - f(y)| \leqslant \frac{1}{2^M}$ holds true for all $x, y \in \mathbb{I}_c$ and all $M \in \mathbb{N}$ that satisfy $\|x - y\| \leqslant \frac{1}{g(M)}$.*

Remark 119 *Per definition, every function that is computably continuous is also Banach-Mazur computable. However, there exist infinitely many Banach-Mazur computable functions that are not computably continuous (see [Avigad and Brattka(2014)] for a detailed discussion.) Accordingly, it is not possible to compute the local variations for functions of this kind.*

5.2.4. General results for computable sequences of numbers and functions

In the following we establish some properties of computable sequences which will be needed subsequently.

Lemma 120 *Let $\{r_n^{\triangle}\}_{n\in\mathbb{N}}$ and $\{r_n^{\triangledown}\}_{n\in\mathbb{N}}$ be computable sequences of rational numbers that satisfy*

$$r_n^{\triangle} \leqslant r_{n+1}^{\triangle} \qquad \wedge \qquad r_n^{\triangledown} \geqslant r_{n+1}^{\triangledown} \qquad \wedge \qquad \lim_{m\to\infty} r_m^{\triangle} = \lim_{m\to\infty} r_m^{\triangledown} \tag{5.12}$$

for all $n \in \mathbb{N}$. Then, $x := \lim_{m\to\infty} r_m^{\triangle} = \lim_{m\to\infty} r_m^{\triangledown}$ is a computable number.

Proof 121 *See e.g. [Pour-El and Richards(2017)].*

Lemma 122 *Let $\{\widetilde{r}_n^{\triangle}\}_{n\in\mathbb{N}}$ and $\{\widetilde{r}_n^{\triangledown}\}_{n\in\mathbb{N}}$ be computable sequences of rational numbers that satisfy*

$$\widetilde{r}_n^{\triangle} \leqslant \lim_{m\to\infty} \widetilde{r}_m^{\triangle} \qquad \wedge \qquad \widetilde{r}_n^{\triangledown} \geqslant \lim_{m\to\infty} \widetilde{r}_m^{\triangledown} \qquad \wedge \qquad \lim_{m\to\infty} \widetilde{r}_m^{\triangle} = \lim_{m\to\infty} \widetilde{r}_m^{\triangledown}$$

for all $n \in \mathbb{N}$. Then, $\widetilde{x} := \lim_{m\to\infty} \widetilde{r}_m^{\triangle} = \lim_{m\to\infty} \widetilde{r}_m^{\triangledown}$ is a computable number.

Proof 123 *Define the sequences $\{r_n^{\triangle}\}_{n\in\mathbb{N}}$ and $\{r_n^{\triangledown}\}_{n\in\mathbb{N}}$ by setting $r_n^{\triangle} := \max\{\widetilde{r}_m^{\triangle} : m \leqslant n\}$ and $r_n^{\triangledown} := \min\{\widetilde{r}_m^{\triangledown} : m \leqslant n\}$ for all $n \in \mathbb{N}$. Then, the sequences $\{r_n^{\triangle}\}_{n\in\mathbb{N}}$ and $\{r_n^{\triangledown}\}_{n\in\mathbb{N}}$ satisfy* (5.12) *with $\widetilde{x} = \lim_{m\to\infty} r_m^{\triangle} = \lim_{m\to\infty} r_m^{\triangledown}$. Furthermore, since minimization and maximization are recursive operations, the sequences $\{r_n^{\triangle}\}_{n\in\mathbb{N}}$ and $\{r_n^{\triangledown}\}_{n\in\mathbb{N}}$ are computable sequences of rational numbers. Thus, by Lemma 120, we have $\widetilde{x} \in \mathbb{R}_c$.*

Theorem 124 *Let $\{x_n^{\triangle}\}_{n\in\mathbb{N}}$ and $\{x_n^{\triangledown}\}_{n\in\mathbb{N}}$ be computable sequences of computable numbers that satisfy*

$$x_n^{\triangle} \leqslant \lim_{m\to\infty} x_m^{\triangle} \qquad \wedge \qquad x_n^{\triangledown} \geqslant \lim_{m\to\infty} x_m^{\triangledown} \qquad \wedge \qquad \lim_{m\to\infty} x_m^{\triangle} = \lim_{m\to\infty} x_m^{\triangledown}$$

for all $n \in \mathbb{N}$. Then, $x_* := \lim_{m\to\infty} x_m^{\triangle} = \lim_{m\to\infty} x_m^{\triangledown}$ is a computable number.

Proof 125 *By assumption, there exists a computable double sequence $(\widetilde{r}_{n,m}^{\triangle})_{n,m\in\mathbb{N}}$ of rational numbers such that $|x_n^{\triangle} - \widetilde{r}_{n,m}^{\triangle}| < 2^{-m}$ holds true for all $n, m \in \mathbb{N}$. Define the sequence $(\widetilde{r}_n^{\triangle})_{n\in\mathbb{N}}$ by setting $\widetilde{r}_n^{\triangle} := \widetilde{r}_{n,n}^{\triangle} - 2^{-n}$ for all $n \in \mathbb{N}$. By construction, $(\widetilde{r}_n^{\triangle})_{n\in\mathbb{N}}$ is a computable sequence of rational numbers which satisfies $\lim_{n\to\infty} \widetilde{r}_n^{\triangle} = x_*$ as well as $\widetilde{r}_n^{\triangle} \leqslant x_*$ for all $n \in \mathbb{N}$. Likewise, we can find a computable sequence $(\widetilde{r}_n^{\triangledown})_{n\in\mathbb{N}}$ of rational numbers that satisfies $\lim_{n\to\infty} \widetilde{r}_n^{\triangledown} = x_*$ as well as $\widetilde{r}_n^{\triangledown} \geqslant x_*$ for all $n \in \mathbb{N}$. Thus, by* Lem. *122, x_* is a computable number.*

Lemma 120, 122 and Theorem 124 are based on the representation of computable numbers through *interval arithmetics.* The same concept can be used to prove an effectivity result for monotonic sequences of computable numbers.

Theorem 126 *Let $(x_n)_{n\in\mathbb{N}}$ be a monotonically increasing computable sequence of computable numbers that satisfies $\lim_{n\to\infty} x_n = x_*$ for some (computable) real number $x_* \in \mathbb{R}_c$. Then, there exists a recursive function $\zeta : \mathbb{N} \to \mathbb{N}$ such that*

$$\left|x_* - x_n\right| < \frac{1}{2^N}.$$

holds true for all $n, N \in \mathbb{N}$ that satisfy $n \geqslant \zeta(N)$. That is, the sequence $(x_n)_{n\in\mathbb{N}}$ converges effectively to x_*.

Proof 127 *The requirement of x_* being a computable number ensures the existence of a computable sequence $(r_n)_{n\in\mathbb{N}}$ of rational numbers that satisfies $|x_* - r_n| < 2^{-n}$ for all $n \in \mathbb{N}$. By setting*

$$r_n^{\triangledown} := \min\{r_m + 2^{-m} : m \leqslant n\} \tag{5.13}$$

for all $n \in \mathbb{N}$, we obtain a monotonically decreasing computable sequence $(r_n^{\triangledown})_{n\in\mathbb{N}}$ of rational numbers with $\lim_{n\to\infty} r_n^{\triangledown} = x_*$. On the other hand, we can find a representation of the sequence $(x_n)_{n\in\mathbb{N}}$ in terms of a computable double sequence $(q_{n,m})_{n,m\in\mathbb{N}}$ of rational numbers that satisfies $|x_n - q_{n,m}| < 2^{-m}$ for all $n, m \in \mathbb{N}$. We have $q_{n,n} - 2^{-n} < x_n \leqslant x_m$ for all $n, m \in \mathbb{N}$ that satisfy $n \leqslant m$. Consequently, the computable sequence $(r_n^{\triangle})_{n\in\mathbb{N}}$ of rational numbers defined by setting

$$r_n^{\triangle} := \max\{q_{m,m} - 2^{-m} : m \leqslant n\} \tag{5.14}$$

for all $n \in \mathbb{N}$ is monotonically increasing and satisfies $r_n^{\triangle} < x_m$ for all $n, m \in \mathbb{N}$ with $n \leqslant m$ as well as $\lim_{n\to\infty} r_n^{\triangle} = x_*$. We arrive at the inequality $r_n^{\triangle} < x_m \leqslant x_* < r_n^{\triangledown}$, which holds true for all $n, m \in \mathbb{N}$ that satisfy $n \leqslant m$. Therefore, we have $|x_* - x_m| < r_n^{\triangledown} - r_n^{\triangle}$ for all $n, m \in \mathbb{N}$ that satisfy $n \leqslant m$. Following the previously established equality $\lim_{n\to\infty} r_n^{\triangledown} = \lim_{n\to\infty} r_n^{\triangle} = x_*$, we also have $\lim_{n\to\infty}(r_n^{\triangledown} - r_n^{\triangle}) = 0$. Setting

$$\zeta(N) := \min\{n : r_n^{\triangledown} - r_n^{\triangle} \leqslant 2^{-N}\} \tag{5.15}$$

yields the required recursive function.

Remark 128 *Given $x_* \in \mathbb{R}_c$, Theorem 126 proves the effective convergence of any monotonically increasing computable sequence $(x_n)_{n\in\mathbb{N}}$ of computable numbers that satisfies $\lim_{n\to\infty} x_n = x_*$. The monotonicity of $(x_n)_{n\in\mathbb{N}}$ is a necessary requirement in this context. That is, there exist computable sequences of computable numbers that converge to a computable number, but the convergence is non-effective. Consider any recursively enumerable but non-recursive set $\mathcal{A} \subset \mathbb{N}$ with recursive bijection $\varphi : \mathbb{N} \to \mathcal{A}$ and define the computable sequence of rational numbers $(r_n)_{n\in\mathbb{N}}$ by settin $r_n := 2^{-\varphi(n)}$. Then, $\lim_{n\to\infty} r_n = 0$, which is a computable number. On the other hand, $(r_n)_{n\in\mathbb{N}}$ does not converge effectively to 0, since this would contradict the non-recursivity of $\mathcal{A}$.*

Remark 129 *Note that it is possible to find a computable sequence $\{x_n\}_{n\in\mathbb{N}}$ of rational numbers that converges to a computable real number $x_* \in \mathbb{R}_c$ (which can further be rational), i.e.,*

$$\lim_{n\to\infty} \left|x_* - x_n\right| = 0,$$

but the convergence is not effective. Then this sequence is not monotonically increasing or decreasing.

Next, we establish similar results for computable sequences of computable continuous functions.

Theorem 130 *Let $F : [0,1] \to \mathbb{R}$ be a computable continuous function and $\{F_N\}_{N\in\mathbb{N}}$ be a computable sequence thereof with $F_N(x) \leqslant F_{N+1}(x)$, $x \in [0,1]$, and*

$$\lim_{N\to\infty} F_N(x) = F(x).$$

Then there exists a recursive function $\varphi : \mathbb{N} \to \mathbb{N}$ such that for all $M \in \mathbb{N}$ we have for all $N \geqslant \varphi(M)$

$$\left|F(x) - F_N(x)\right| < \frac{1}{2^M}.$$

Proof 131 *Let $Q_N(x) = F(x) - F_N(x)$, $x \in [0,1]$. We have $0 \leqslant Q_{N+1}(x) \leqslant Q_N(x)$ and $\lim_{N\to\infty} Q_N(x) = 0, \quad x \in [0,1]$. Let $M \in \mathbb{N}$ be arbitrary. There exists an $N_0 = N_0(M,x)$ with*

$$Q_N(x) < \frac{1}{2^M} \quad \text{for all } N \geqslant N_0(M,x).$$

We define the set

$$\mathcal{S}_{N,M} = \left\{x \in [0,1] : Q_N(x) < \frac{1}{2^M}\right\}$$

and observe that $\mathcal{S}_{N,M} \subset \mathcal{S}_{N+1,M}$. Now, $\{\mathcal{S}_{N,M}\}$ is a family of open sets with $[0,1] \subset \bigcup_{N=1}^{\infty} \mathcal{S}_{N,M}$. Since $[0,1]$ is a compact set [Rudin(1987)], there exists an $N_0(M)$ with $[0,1] \subset \mathcal{S}_{N_0,M}$ and therewith $Q_{N_0}(x) < \frac{1}{2^M}$ for N_0 and also all $N \geqslant N_0$. Let

$$\max_{x\in[0,1]} Q_N(x) = C_N.$$

Since Q_N is a computable continuous function, we always have $C_N \in \mathbb{R}_c$. Further, since $\{Q_N\}_{N\in\mathbb{N}}$ is a computable sequence of computable real numbers, the sequence $\{C_N\}_{N\in\mathbb{N}}$ is also a computable sequence of computable real numbers. For all $N \in \mathbb{N}$ it holds that $C_N \geqslant C_{N+1}$ and

$$\lim_{N\to\infty} C_N = 0.$$

Accordingly, there exists a recursive function $\varphi : \mathbb{N} \to \mathbb{N}$ such that for all $M \in \mathbb{N}$ we have for all $N \geqslant \varphi(M)$

$$\left|F(x) - F_N(x)\right| = \left|Q_N(x)\right| < \frac{1}{2^M}$$

which proves the desired result.

Some remarks are in order:

1. The result extends to functions on compact spaces.
2. The result remains true for monotonically decreasing functions.
3. It is important that F is a computable continuous function. Already for computable sequences of rational numbers with $x_n \leqslant x_{n+1}$ that converge to a $x_* \notin \mathbb{R}_c$, we do not have effective convergence, see e.g. [Specker(1949)].
4. A part of the proof is not effective as we required compactness which is needed to show uniform convergence. This is subsequently used to show the effective convergence of the computable continuous function F.

We can use Theorem 130 to show the following result.

Corollary 132 *Let $\{F_N\}_{N\in\mathbb{N}}$ and $\{G_N\}_{N\in\mathbb{N}}$ be computable sequences of computable continuous functions on $[0,1]$ with*

$$F_N(x) \leqslant F_{N+1}(x) \leqslant G_{N+1}(x) \leqslant G_N(x)$$

and

$$\lim_{N\to\infty} F_N(x) = \lim_{N\to\infty} G_N(x) =: \Phi(x), \quad x \in [0,1].$$

Then $\Phi : [0,1] \to \mathbb{R}$ is also a computable continuous function and $\{F_N\}_{N\in\mathbb{N}}$ and $\{G_N\}_{N\in\mathbb{N}}$ converge effectively to Φ.

Proof 133 *We set*

$$Q_N(x) = G_N(x) - F_N(x), \quad x \in [0,1],$$

and $\{Q_N\}_{N\in\mathbb{N}}$ is a computable sequence of computable continuous functions. For $x \in [0,1]$ we have

$$\begin{aligned} Q_N(x) &\geqslant G_{N+1}(x) - F_N(x) \\ &\geqslant G_{N+1}(x) - F_{N+1}(x) = Q_{N_M}(x) \end{aligned}$$

and

$$\lim_{N\to\infty} Q_N(x) = 0, \quad x \in [0,1].$$

Now, from Theorem 130 follows that the computable sequence $\{Q_N\}_{N\in\mathbb{N}}$ of computable continuous functions converges effectively to 0 proving the desired result.

Similar results for computable sequences of Banach-Mazur computable functions can be derived. We will use the following theorem in analyzing computability of upper and lower bounds for capacity functions introduced in the next section. We consider sequences of functions over computable real numbers as well as those over computable compound channels.

Theorem 134 *Let $\{F_N\}_{n\in\mathbb{N}}$ and $\{G_N\}_{n\in\mathbb{N}}$ be computable sequences of functions $F_N : [0,1] \cap \mathbb{R}_c \to \mathbb{R}_c$ and $G_N : [0,1] \cap \mathbb{R}_c \to \mathbb{R}_c$, $N \in \mathbb{N}$, with*

$$F_N(x) \leqslant F_{N+1}(x),\ x \in [0,1] \cap \mathbb{R}_c,$$

$$G_N(x) \geqslant G_{N+1}(x),\ x \in [0,1] \cap \mathbb{R}_c,$$

and

$$\lim_{N\to\infty} F_N(x) = \lim_{N\to\infty} G_N(x) =: \varPhi(x),\ x \in [0,1] \cap \mathbb{R}_c.$$

Then $\varPhi : [0,1] \cap \mathbb{R}_c \to \mathbb{R}$ is also a Banach-Mazur computable function.

This result allows us to consider the computability of lower and upper bounds on the ϵ-capacity as a function of the error input, i.e. ϵ.

5.3. Basic concepts and capacity results

We first introduce the concept of *classical* compound channels, and apply the findings thus far to its ϵ-capacity afterwards.

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite input and output alphabets and $\mathcal{S}$ be a finite state (uncertainty) set. Then for a fixed channel state $s \in \mathcal{S}$, the channel is given by a stochastic matrix $W_s : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ which we interchangeably also write as $W_s \in \mathcal{CH}(\mathcal{X};\mathcal{Y})$, where the latter denotes the set of all channels from $\mathcal{X}$ to $\mathcal{Y}$. The channel state $s \in \mathcal{S}$ is assumed to remain constant throughout the whole transmission so that the discrete memoryless channel is given by $W_s(y^n|x^n) := \prod_{i=1}^n W_s(y_i|x_i)$ for all $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$.

Definition 135 *The* compound channel *generated by uncertainty set $\mathcal{W} := \{W_s \in \mathcal{CH}(\mathcal{X};\mathcal{Y}) : s \in \mathcal{S}\}$ is given by the sequence of channels $\{W_s^n, W_s \in \mathcal{W}\}_{n=1}^{\infty}$. The set of all such compound channels is denoted by $\mathcal{CC}(\mathcal{X},\mathcal{S};\mathcal{Y})$.*

Further, let $\mathcal{CH}_c(\mathcal{X};\mathcal{Y})$ be the set of all computable channels, i.e. for a channel $W \in \mathcal{CH}_c(\mathcal{X};\mathcal{Y})$ we have $W(\cdot|x) \in \mathcal{P}_c(\mathcal{Y})$ for every $x \in \mathcal{X}$. Finally, computable compound channels are defined as follows.

Definition 136 *A compound channel generated $\mathcal{W} = \{W_s \in \mathcal{CH}_c(\mathcal{X};\mathcal{Y}) : s \in \mathcal{S}\}$ is said to be computable if there is a recursive function $\varphi : \mathcal{S} \to \mathcal{CH}_c(\mathcal{X},\mathcal{S};\mathcal{Y})$ with $\varphi(s) = W_s$ for all $s \in \mathcal{S}$. The set of all computable compound channels is denoted by $\mathcal{CC}_c(\mathcal{X},\mathcal{S};\mathcal{Y})$.*

We require namely, that the compound set $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X},\mathcal{S};\mathcal{Y})$ is algorithmically constructable i.e., for every state $s \in \mathcal{S}$ the channel W_s can be constructed by an algorithm (or Turing machine) with input s. We further need a concept of distance. For two channels $W_1, W_2 \in \mathcal{CH}(\mathcal{X};\mathcal{Y})$ we define the d-distance between W_1 and W_2 based on the total variation distance as

$$d(W_1, W_2) = \max_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \big|W_1(y|x) - W_2(y|x)\big|.$$

To extend this concept to compound channels, we consider the worst case distance between $\mathcal{W}_1 \in \mathcal{CC}(\mathcal{X}, \mathcal{S}_1; \mathcal{Y})$ and $\mathcal{W}_2 \in \mathcal{CC}(\mathcal{X}, \mathcal{S}_2; \mathcal{Y})$ as

$$\begin{aligned} D(\mathcal{W}_1, \mathcal{W}_2) = \max \quad & \max_{s_1\in\mathcal{S}_1} \min_{s_2\in\mathcal{S}_2} d(W_{s_1}, W_{s_2}), \\ & \quad \max_{s_2\in\mathcal{S}_2} \min_{s_1\in\mathcal{S}_1} d(W_{s_2}, W_{s_1})\big\}. \end{aligned} \tag{5.16}$$

Further, on the interval $\mathbb{I} = [0, 1]$ we define the distance $D_{\mathbb{I}}(\epsilon_1, \epsilon_2) = |\epsilon_1 - \epsilon_2|$.

We define the set $\mathfrak{W} = \mathcal{CC}(\mathcal{X}, \mathcal{S}; \mathcal{Y}) \times \mathbb{I}$ and the distance

$$D_{\mathfrak{W}}((\mathcal{W}_1, \epsilon_1), (\mathcal{W}_2, \epsilon_2)) = \max \quad D(\mathcal{W}_1, \mathcal{W}_2), D_{\mathbb{I}}(\epsilon_1, \epsilon_2)\big\}$$

for $(\mathcal{W}_i, \epsilon_i) \in \mathfrak{W}$, $i = 1, 2$. Then, $(\mathfrak{W}, D_{\mathfrak{W}})$ is a compact Hausdorff space [Rudin(1987)].

We further set

$$\mathfrak{W}_c = \quad (\mathcal{W}, \epsilon) : \mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y}), \epsilon \in \mathbb{I}_c\big\}$$

with $\mathbb{I}_c = \mathbb{I} \cap \mathbb{R}_c$ the computable interval. We have the following properties:

1. $D(\cdot, \cdot)$ is a computable continuous function on $\mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ which follows from its definition.
2. $D_{\mathbb{I}}(\cdot, \cdot)$ is a computable continuous function on $\mathbb{I}_c$ which follows similarly.
3. $D_{\mathfrak{W}}(\cdot, \cdot)$ is a computable continuous function on $\mathfrak{W}_c$, since it is the maximum of two computable continuous functions.

Since the actual channel state is unknown to transmitter and receiver, universal encoder and decoder are needed that are independent of the channel state.

Definition 137 *An* (n, M_n)-code *is a set of doublets* $\{(x_m, D_m), m \in [M_n]\}$ *with*

- $x_m \in \mathcal{X}^n, m \in [M_n]$ *and*
- $D_m \subset \mathcal{Y}^n, m \in [M_n]$ *such that* $D_m \bigcap D_{m'} = \varnothing$ *for* $m \neq m'$ *and* $\bigcup_{m\in[M_n]} D_m = \mathcal{Y}^n$.

As the receiver needs to decode the transmitted message for all possible channel realizations, we define the *average probability of error* for the compound channel $\mathcal{W}$ as

$$\bar{e}_n(\mathcal{W}) = \max_{s\in\mathcal{S}} \frac{1}{M_n} \sum_{m\in[M_n]} e_{m,s,n}$$

with

$$e_{m,s,n} := \sum_{y^n\in D_m^c} W_s^n(y^n|x_m). \tag{5.17}$$

This leads to two definitions for achievable rate of communication, when some $\epsilon \geqslant 0$ amount of error is allowed. We refer to the first one as the *traditional definition* (c.f [Ahlswede(2015)]) and the second one as the *alternative (optimistic) definition* (c.f [Yagi and Nomura(2014)]). In the following, we consider these two definitions and derive capacity results for each. Before stating the capacity results, we need the following notation. The *mutual information* $I(X;Y)$, between two random variables (X,Y), is defined by

$$I(X;Y) := H(X) - H(Y|X). \tag{5.18}$$

Also, given $W \in \mathcal{CH}(\mathcal{X},\mathcal{Y})$ and random variables (X,Y) distributed according to $P_X = P$ and $P_{Y|X}(\cdot|\cdot) = W(\cdot|\cdot)$ on $\mathcal{X}$ and $\mathcal{Y}$ respectively, we define

$$I(P,W) := I(X;Y). \tag{5.19}$$

This quantity is called *mutual information of the channel W*. For properties of this quantity see [Csiszár and Körner(1981)]. To state our results related to the *zero-error capacity* of the channel, we need the concept of a simple graph $G = (V(G), E(G))$, characterized by the set of vertices $V(G)$ and the set of edges $E(G)$. Again given $W \in \mathcal{CH}(\mathcal{X},\mathcal{Y})$ and any $x \in \mathcal{X}$, define sets $\mathcal{Y}_x := \{y \in \mathcal{Y} : W(y|x) > 0\}$ and the graph $G(W) = (\mathcal{X}, E_W(G))$ with

$$E_W(G) := \{(x,x') : \mathcal{Y}_x \bigcap \mathcal{Y}_{x'} = \varnothing\}. \tag{5.20}$$

Also, for $P \in \mathcal{P}(\mathcal{X}), \delta > 0$, let $T^n_{P,\delta}$ be the set of all δ-typical sequences in $\mathcal{X}^n$ (see Appendix A for properties of δ-typical sequences). Let $G^n[P,\delta]$ be the graph induced by $G(W^n)$ on the set $T^n_{P,\delta}$. We define

$$C_0(W,P) := \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \omega(G^n[P,\delta]),$$

where $\omega(G^n[P,\delta])$ is the *clique number* of the graph $G^n[P,\delta]$, namely

$$\omega(G^n[P,\delta]) := \max\{|\Omega| : \Omega \subset V(G^n[P,\delta]) : x,x' \in \Omega \to \mathcal{Y}_x \bigcap \mathcal{Y}_{x'} = \varnothing\}. \tag{5.21}$$

5.3.1. Traditional definition of ϵ-capacity

We define the following numbers for $n \in \mathbb{N}$ and $0 \leqslant \epsilon < 1$, one corresponding to the average error criterion and the next to the maximum error criterion. The following definitions lead to the traditional definitions of ϵ-capacity of the compound channel.

1. $N(\mathcal{W},\epsilon,n) := \max\{N \in \mathbb{N} : \exists (n,N) - \textit{code}$ for $\mathcal{W}$ with $\overline{e}_n(\mathcal{W}) \leqslant \epsilon\}$,
2. $N^{\max}(\mathcal{W},\epsilon,n) := \max\{N \in \mathbb{N} : \exists (n,N) - \textit{code}$ for $\mathcal{W}$ with $\max_{s\in\mathcal{S}} \max_{m\in[N]} e_{s,n}(m) \leqslant \epsilon\}$.

We consider two notations of capacity corresponding to these numbers that are defined in the following.

Definition 138

Let $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ and $0 \leqslant \epsilon < 1$. Then

- $\overline{C(\mathcal{W}, \epsilon)} := \limsup_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, n)$ *and* $\overline{C^{max}(\mathcal{W}, \epsilon)} := \limsup_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W}, \epsilon, n)$ *are the optimistic ϵ-capacities of $\mathcal{W}$ under average and maximal error criteria respectively.*

- *Also* $\underline{C(\mathcal{W}, \epsilon)} := \liminf_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, n)$ *and* $\underline{C^{max}(\mathcal{W}, \epsilon)} := \liminf_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W}, \epsilon, n)$ *are the pessimistic ϵ-capacities of $\mathcal{W}$ under average and maximal error criteria respectively.*

- *Finally, $C(\mathcal{W}, \epsilon)$ and $C^{\max}(\mathcal{W}, \epsilon)$ are the ϵ-capacities of $\mathcal{W}$ under average and maximal error criteria, if*
$\limsup_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, n) = \liminf_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, n)$ *and* $\limsup_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W}, \epsilon, n)$
$=\liminf_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W}, \epsilon, n)$ *respectively.*

The asymptotic behavior of $N^{\max}(\mathcal{W}, \epsilon, n), \epsilon \in [0, 1)$ of the compound channel has already been established in the literature and is stated in the following.

Theorem 139 ([Blackwell et al.(1959)Blackwell, Breiman, and Thomasian, Wolfowitz(1960)]) *For $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ it holds for $\epsilon \in (0, 1)$*

$$C^{\max}(\mathcal{W}, \epsilon) = \max_{P_X \in \mathcal{P}(\mathcal{X})} \min_{s \in \mathcal{S}} I(P_X, W_s) =: C(\mathcal{W}), \tag{5.22}$$

and for $\epsilon = 0$,

$$C^{\max}(\mathcal{W}, \epsilon) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{s \in \mathcal{S}} C_0(W_s, P) =: C_0(\mathcal{W}). \tag{5.23}$$

Proof 140 *See [Blackwell et al.(1959)Blackwell, Breiman, and Thomasian, Wolfowitz(1960)] for proof of (5.22) and [Csiszár and Körner(1981)] for (5.23).*

We refer to $C(\mathcal{W})$ and $C_0(\mathcal{W})$ as the *compound capacity* and *zero-error compound capacity* of $\mathcal{W}$.

This result is in fact a generalization of similar results for the case of perfectly known channel state ($|\mathcal{S}| = 1$). In other words, defining $N(W, \epsilon, n) := N(\mathcal{W}, \epsilon, n)$ for $\mathcal{W} = \{W\}$, the following holds.

Theorem 141 *For $W \in \mathcal{CH}(\mathcal{X}; \mathcal{Y})$ it holds for $\epsilon \in (0, 1)$*

$$\lim_{n\to\infty} \frac{1}{n} \log N^{\max}(W, \epsilon, n) = \max_{P_X \in \mathcal{P}(\mathcal{X})} I(P_X, W_s) =: C(W), \tag{5.24}$$

and for $\epsilon = 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log N(W, \epsilon, n) = \lim_{n\to\infty} \frac{1}{n} \log N^{\max}(W, \epsilon, n) = \max_{P \in \mathcal{P}(\mathcal{X})} C_0(W, P) =: C_0(W). \tag{5.25}$$

This gives a complete characterization of the asymptotic behavior of $N^{\max}(\mathcal{W},\epsilon,n),\epsilon\in[0,1)$. Such a characterization is missing for $N(\mathcal{W},\epsilon,n),\epsilon\in(0,1)$, because there is in general no strong converse for this number (see [Ahlswede(1967b)]). It is of course clear that for $\epsilon=0$, $N(\mathcal{W},\epsilon,n)=N^{\max}(\mathcal{W},\epsilon,n)$. This characterization does not exist even for compound channels with two channel states ($|\mathcal{S}|=2$). In what follows we state two results, one already existing from [Ahlswede(2015)] and another, the achievablity for $\epsilon=\frac{1}{2}$. A converse for this point is still missing.

Lemma 142 ([Ahlswede(2015)]) *For $\mathcal{W}:=\{W_1,W_2\}\subset\mathcal{CC}(\mathcal{X},\{1,2\};\mathcal{Y})$, it holds*

$$C(\mathcal{W},\epsilon)=\begin{cases}C(\mathcal{W}) & \text{for } 0<\epsilon<1/2\\ \min_{s=1,2}C(W_s) & \text{for } 1/2<\epsilon<1\end{cases}. \tag{5.26}$$

The following is the achievability statement for $\epsilon=1/2$.

Lemma 143 *For $\mathcal{W}\subset\mathcal{CC}(\mathcal{X},\{1,2\};\mathcal{Y})$ with $C_0(\mathcal{W})>0$, it holds*

- $\liminf_{n\to\infty}\frac{1}{n}\log N(\mathcal{W},\frac{1}{2},n)\geqslant\max\{C(\mathcal{W}),\min_{s=1,2}C_0(W_s)\}$.

Proof 144 *If $C(\mathcal{W})\geqslant\min_{s=1,2}C_0(W_s)$, the assertion is clear, as $C(\mathcal{W})\leqslant C(\mathcal{W},\frac{1}{2})$. Assume otherwise. Since by assumption of the lemma $C_0(\mathcal{W})>0$, from Theorem 141, there exists a $k\in\mathbb{N}$, for which there exists a $(2,k)$-code consisting at the encoder of $x_1,x_2\in\mathcal{X}^k$ and at the decoder of $D_1,D_2\subset\mathcal{Y}^k$, with $D_1\bigcap D_2=\varnothing,D_1\bigcup D_2=\mathcal{Y}^n$, such that $\min_{s=1,2}\min_{i=1,2}\sum_{y^k\in D_i}W_s(y^k|x_i)=1$. By Theorem 141, for $\delta>0$, there exists $l_0\in\mathbb{N}$ such that for $l>l_0$, there exist (l,M_l)-codes $\{(u_j^{(s)},\Lambda_j^{(s)}),i\in[M_n]\}$ for $s=1,2$, with $\frac{1}{l}\log M_l\geqslant\min_{s=1,2}C_0(W_s)-\delta$ and $\max_{s=1,2}\max_{m\in[M_l]}e_{l,s}(m)=0$. Construct the $(k+l,2M_l)$-code $\{(\mathring{u}_m,\mathring{D}_m),m\in[2M_l]$ as*

- $\mathring{u}_m:=x_1\oplus u_m^{(1)}$*, for* $m\in[M_l]$*,*
- $\mathring{u}_m:=x_2\oplus u_{m-M_l}^{(2)}$*, for* $m\in\{M_l+1,\dots,2M_l\}$*.*

and decoding operations defined by

- $\mathring{D}_m:=D_1\times\Lambda_m^{(1)}$*, for* $m\in[M_l]$*,*
- $\mathring{D}_m:=D_2\times\Lambda_{m-M_l}^{(2)}$*, for* $m\in\{M_l+1,\dots,2M_l\}$*.*

We calculate the error due to this code. We have thee following average probability of success for W_1:

$$\begin{aligned}\frac{1}{2M_l}\Big(\sum_{m\in[2M_l]} 1-e_{k+l,1}(m)\Big) &= \frac{1}{2M_l}\sum_{y^{k+l}\in D_1\times\Lambda_m^{(1)}} W_1^{k+l}(y^{k+l}|x_1\oplus u_m^{(1)})\\ &+\frac{1}{2M_l}\sum_{y^{k+l}\in D_2\times\Lambda_m^{(2)}} W_1^{k+l}(y^{k+l}|x_2\oplus u_m^{(2)})\\ &\geqslant \frac{1}{2M_l}\sum_{y^{k+l}\in D_1\times\Lambda_m^{(1)}} W_1^{k+l}(y^{k+l}|x_1\oplus u_m^{(1)})\\ &= \frac{1}{2M_l}\sum_{y^k\in D_1} W_1^k(y^k|x_1)\sum_{y^l\in\Lambda_m^{(1)}} W_1^l(y^l|u_m^{(1)}) = 1/2. \qquad (5.27)\end{aligned}$$

Similar calculation yields the same lower bound on $\frac{1}{2M_l}\big(\sum_{m\in[2M_l]}1-e_{k+l,2}(m)\big)$ *and hence we conclude* $\overline{e}(\mathcal{W})\leqslant\frac{1}{2}$. *Hence, for* $n:=k+l$ *we have*

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}N(\mathcal{W},\frac{1}{2},n) &\geqslant \liminf_{l\to\infty}\frac{1}{l+k}\log 2M_l\\ &\geqslant \liminf_{l\to\infty}\frac{l}{l+k}\frac{1}{l}\log 2M_l\\ &\geqslant \min_{s=1,2}C_0(V_s)-\delta. \qquad (5.28)\end{aligned}$$

As $\delta>0$ *was arbitrary, we are done.*

5.3.2. Alternative definition of ϵ-capacity

Alternatively, we can consider a definition of ϵ-capacity with a more relaxed requirement on error. Much like the previous case, we start by defining the following numbers for $n\in\mathbb{N}$ and $0\leqslant\epsilon<1$, one corresponding to the average error criterion and the next to the maximum error criterion.

1. $N_{Alt}(\mathcal{W},\epsilon,n):=\max\{N_n\in\mathbb{N}:\exists$ a sequence of $(k,N_k)-$ *codes* for $\mathcal{W}$ with $\limsup_{k\to\infty}\overline{e}_k(\mathcal{W})\leqslant\epsilon\}$,

2. $N_{Alt}^{\max}(\mathcal{W},\epsilon,n):=\max\{N_n\in\mathbb{N}:\exists$ a sequence of $(k,N_k)-$ *codes* for $\mathcal{W}$ with
$\limsup_{k\to\infty}\max_{s\in\mathcal{S}}\max_{m\in[N]}e_{s,k}(m)\leqslant\epsilon\}$.
Again, given these numbers we can define two notions of capacity.

Definition 145 *Let* $\mathcal{W}\in\mathcal{CC}(\mathcal{X},\mathcal{S};\mathcal{Y})$ *and* $0\leqslant\epsilon<1$. *Then*

- $\overline{C_{Alt}(\mathcal{W},\epsilon)}:=\limsup_{n\to\infty}\frac{1}{n}\log N_{Alt}(\mathcal{W},\epsilon,n)$ *and* $\overline{C_{Alt}^{max}(\mathcal{W},\epsilon)}:=\limsup_{n\to\infty}\frac{1}{n}\log N_{Alt}^{\max}(\mathcal{W},\epsilon,n)$

  are the optimistic ϵ*-capacities of* $\mathcal{W}$ *under average and maximal error criteria respectively.*

- *Also* $\underline{C_{Alt}(\mathcal{W},\epsilon)} := \liminf_{n\to\infty} \frac{1}{n}\log N_{Alt}(\mathcal{W},\epsilon,n)$ *and* $\underline{C_{Alt}^{max}(\mathcal{W},\epsilon)} := \liminf_{n\to\infty} \frac{1}{n}$ $\log N_{Alt}^{\max}(\mathcal{W},\epsilon,n)$ *are the pessimistic* ϵ*-capacities of* $\mathcal{W}$ *under average and maximal error criteria respectively.*
- *Finally,* $C_{Alt}(\mathcal{W},\epsilon)$ *and* $C_{Alt}^{\max}(\mathcal{W},\epsilon)$ *are the* ϵ*-capacities of* $\mathcal{W}$ *under average and maximal error criteria, if* $\underline{C_{Alt}(\mathcal{W},\epsilon)} = \overline{C_{Alt}(\mathcal{W},\epsilon)}$ *and* $\underline{C_{Alt}^{max}(\mathcal{W},\epsilon)} = \overline{C_{Alt}^{max}(\mathcal{W},\epsilon)}$ *respectively.*

To a large part the alternative definition coincides asymptotically with the previous definition. For instance, Theorem 139, can be stated as follows (see [Blackwell et al.(1959)Blackwell, Breiman, and Thomasian, Wolfowitz(1960)]).

Theorem 146 *For* $\mathcal{W} \in \mathcal{C}(\mathcal{X},\mathcal{S};\mathcal{Y})$ *it holds for* $\epsilon \in [0,1)$

$$\lim_{n\to\infty} C_{Alt}^{\max}(\mathcal{W},\epsilon) = C(\mathcal{W}). \tag{5.29}$$

Notice that here, $\epsilon = 0$ does not correspond to the zero-error capacity of the channel defined previously. We give a characterization of $N_{Alt}(\mathcal{W},\epsilon,n)$, for $0 \leqslant \epsilon < 1$ and $\mathcal{W} \in \mathcal{C}(\mathcal{X},\{1,2\};\mathcal{Y})$. We also state similar results for the case of perfectly known channel state ($|\mathcal{S}| = 1$).

Theorem 147 *For* $W \in \mathcal{CH}(\mathcal{X};\mathcal{Y})$ *it holds for* $\epsilon \in [0,1)$

$$\lim_{n\to\infty} \frac{1}{n}\log N_{Alt}^{\max}(W,\epsilon,n) = C(W). \tag{5.30}$$

We prove the following.

Theorem 148 *For* $\mathcal{W} \in \mathcal{C}(\mathcal{X},\{1,2\};\mathcal{Y})$ *it holds,*

$$C_{Alt}(\mathcal{W},\epsilon) = \begin{cases} C(\mathcal{W}) & \text{for } 0 \leqslant \epsilon < 1/2 \\ \min_{s=1,2} C(W_s) & \text{for } 1/2 \leqslant \epsilon < 1 \end{cases}. \tag{5.31}$$

We prove this theorem in two steps. The first step is proof of achievability that is formulated in the following lemma.

Lemma 149 *For* $\mathcal{W} \in \mathcal{CC}(\mathcal{X},\{1,2\};\mathcal{Y})$ *it holds*

$$\liminf_{n\to\infty} \frac{1}{n}\log N_{Alt}(\mathcal{W},\frac{1}{2},n) \geqslant \min_{s=1,2} C(W_s).$$

To prove the statement of achievability, we need the following result and in particular, a corollary of it that is stated afterwards.

Lemma 150 *For* $\mathcal{W} \in \mathcal{CC}(\mathcal{X},\{1,2\};\mathcal{Y})$ *with* $C(\mathcal{W}) = 0$ *it holds* $\limsup_{n\to\infty} \frac{1}{n}\log N_{Alt}(\mathcal{W},\epsilon,n) = 0$ *for* $0 < \epsilon < 1$.

Corollary 151 *For* $\mathcal{W} \in \mathcal{CC}(\mathcal{X},\{1,2\};\mathcal{Y})$ *with* $C(\mathcal{W}) = 0$, *it holds,* $\min_{s=1,2} C(W_s) = 0$.

Proof 152 *From Lemma 142 we have* $\limsup_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, n) = \min_{s=1,2} C(W_s)$ *for* $\epsilon \in (1/2, 1)$. *Combining this with Lemma 150 we obtain the result.*

Proof 153 (Proof of Lemma 150) *For* $\epsilon \in (0,1), n \in \mathbb{N}$ *let* $\mathcal{C}_{ID} := \{(u_m, D_m^s)_{m=1}^M : s = 1,2\}$ *be an* (n, M)*-code with informed decoder (see Section 5.3.3 for precise definition), namely*

$$\max_{s=1,2} \frac{1}{M} \sum_{m\in[M]} \sum_{y^n \in D_m^s} W_s^n(y^n|u_m) \leqslant \epsilon. \tag{5.32}$$

Let p^* *be an equidistribution on the message set. Consider the pair of random variables* (X_s, X_s') *with joint distribution*

$$\mathbb{P}(X_s = m, X_s' = m') = p^*(m) \sum_{y^n \in D_{m'}^s} W_s^n(y^n|u_m). \tag{5.33}$$

Therefore from (5.32) we have

$$\mathbb{P}(X_s \neq X_s') \leqslant \epsilon. \tag{5.34}$$

We have

$$\log M = H(p^*) = I(X_s; X_s') + H(X_s|X_s') \leqslant I(X_s; X_s') + \epsilon \log M + 1, \tag{5.35}$$

where $I(X_s; X_s')$ *is the mutual information of random variables* X_s, X_s' *and the first inequality comes from (5.34) and Fano's inequality. Rearranging the above inequality and observing that it holds for* $s = 1,2$ *yields*

$$(1-\epsilon)\log M \leqslant \min_{s=1,2} I(X_s, X_s') + 1 \leqslant \max_{P_X \in \mathcal{P}(\mathcal{X}^n)} \min_{s=1,2} I(P_X, W_s^n) + 1. \tag{5.36}$$

Notice that by definition of $N_{Alt}(\mathcal{W}, \epsilon, n)$, *the error is upper-bounded by a fixed number independent of* n. *Therefore we have* $(1-\epsilon)\limsup_{n\to\infty} \frac{1}{n}\log M \leqslant C(\mathcal{W}) = 0$. *We are done.*

Proof 154 (Proof of Theorem 149) *Let* $C(\mathcal{W}) > 0$, *otherwise from Lemma 150 we conclude* $\min_{s=1,2} C(W_s) = 0$ *and there is nothing to prove. Therefore, according to Theorem 147, there exists* $k_0 \in \mathbb{N}$ *such that for* $k > k_0$, *we have two doublets* $(u_i, D_i), i = 1,2$ *with* $u_i \in \mathcal{X}^k$ *and* $D_1 \bigcap D_2 = \varnothing, D_1 \bigcup D_2 = \mathcal{Y}^k$ *such that*

$$\min_{s=1,2} \min_{i=1,2} \sum_{y^k \in D_i} W_s(y^k|u_i) \geqslant 1 - \epsilon_k. \tag{5.37}$$

with $\epsilon_k \to 0$ *as* $k \to \infty$. *These doublets are what we use for channel state estimation. By Theorem 147, for* $\delta_0 > 0$, *there exists* l_0, *such that for* $l > l_0$ *we find* (l, M_l)*-codes*

$\{(v_m^{(s)}, \Lambda_m^{(s)}) : m \in [M_l]\}$ for $s = 1, 2$ with

$$\frac{1}{l}\log M_l \geqslant \min_{s=1,2} C(W_s) - \delta_0 \tag{5.38}$$

such that

$$\min_{s=1,2} \frac{1}{M_l} \sum_{m\in[M_l]} \sum_{y^l \in \Lambda_m^{(s)}} W_s^l(y^l|v_m^{(s)}) \geqslant 1 - \hat{\epsilon}_l, \tag{5.39}$$

with $\hat{\epsilon}_l \to 0$ as $l \to \infty$. Define the $(2M_l, k+l)$-code $\{(\mathring{u}_m, \mathring{D}_m) : m \in 2M_l\}$ with encoding sequences defined by:

- *$\mathring{u}_m := u_1 \oplus v_m^{(1)}$, for $m \in [M_l]$,*
- *$\mathring{u}_m := u_2 \oplus v_{m-M_l}^{(2)}$, for $m \in \{M_l + 1, \dots, 2M_l\}$,*

and decoding operations defined by sets

- *$\mathring{D}_m := D_1 \times \Lambda_m^{(s)}$, for $m \in [M_l]$,*
- *$\mathring{D}_m := D_2 \times \Lambda_{m-M_l}^{(2)}$, for $m \in \{M_l + 1, \dots, 2M_l\}$.*

Calculating the success probability due to this code, we obtain for W_1:

$$\begin{aligned}\frac{1}{2M_l}\Big(\sum_{m\in[2M_l]} 1 - e_{k+l,1}(m)\Big) &= \frac{1}{2M_l}\sum_{m=1}^{m=M_l} \sum_{y^{k+l}\in D_1\times\Lambda_m^{(1)}} W_1^{k+l}(y^{k+l}|u_1 \oplus v_m^{(1)}) \\ &+ \frac{1}{2M_l}\sum_{m=1}^{m=M_l} \sum_{y^{k+l}\in D_2\times\Lambda_m^{(2)}} W_1^{k+l}(y^{k+l}|u_2 \oplus v_m^{(2)}) \\ &\geqslant \frac{1}{2M_l}\sum_{m=1}^{m=M_l} \sum_{y^{k+l}\in D_1\times\Lambda_m^{(1)}} W_1^{k+l}(y^{k+l}|u_1 \oplus v_m^{(1)}) \\ &= \frac{1}{2M_l}\sum_{m=1}^{m=M_l} \sum_{y^k\in D_1} W_1^k(y^k|u_1) \sum_{y^l\in\Lambda_m^{(1)}} W_1^l(y^l|v_m^{(1)}) \\ &\geqslant \frac{1}{2}(1-\epsilon_k)(1-\hat{\epsilon}_l).\end{aligned} \tag{5.40}$$

Similar calculation yields the same lower bound on $\frac{1}{2M_l}\big(\sum_{m\in[2M_l]} 1 - e_{k+l,2}(m)\big)$ and hence we conclude $\overline{e}(\mathcal{W}) \leqslant 1 - \frac{1}{2}(1-\epsilon_k)(1-\hat{\epsilon}_l)$. Set $k = \sqrt{l}$ and $n := l + \sqrt{l}$. We therefore have

$$\begin{aligned}&\liminf_{n\to\infty} \frac{1}{n}\log N_{Alt}(\mathcal{W}, \frac{1}{2}, n) \\ &\qquad \geqslant \liminf_{l\to\infty} \frac{1}{l+\sqrt{l}} \log 2M_l \\ &\qquad = \liminf_{l\to\infty} \frac{l}{l+\sqrt{l}}\frac{1}{l}(\log 2 + \log \hat{M}_l) \geqslant \min_{s=1,2} C(W_s) - \delta_0.\end{aligned} \tag{5.41}$$

Since δ_0 was arbitrary, we are done.

To prove Theorem 148, it remains to show the following statement of converse.

Lemma 155 *For $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, \{1,2\}; \mathcal{Y})$ it holds*

$$\limsup_{n\to\infty} \frac{1}{n} \log N_{Alt}(\mathcal{W}, \frac{1}{2}, n) \leqslant \min_{s=1,2} C(W_s).$$

Proof 156 *Let $\mathcal{C}_n$ be a sequence of (n, M_n)-codes for $\mathcal{W}$ with $\limsup_{n\to\infty} \bar{e}_n(\mathcal{W}) = \frac{1}{2}$. In other words we have $\bar{e}_n(\mathcal{W}) = \frac{1}{2} + \delta_n$ with $\delta_n \to 0$ as $n \to \infty$. Choosing n large enough such that $\delta_n < \frac{1}{2}$, we obtain $\bar{e}_n(\mathcal{W}) < 1$. Hence for large enough values of n we have $N_{Alt}(\mathcal{W}, \frac{1}{2}, n) \leqslant N(\mathcal{W}, \epsilon, n)$ with $\epsilon < 1$. The proof then follows from Lemma 142.*

Proof 157 (Proof of Theorem 148) *For $\epsilon = 1/2$, given Lemma 149 and Lemma 155 we have*
$\lim_{n\to\infty} \frac{1}{n} \log N_{Alt}(\mathcal{W}, \frac{1}{2}, n) = \min_{s=1,2} C(W_s)$. For $\epsilon \in (0, \frac{1}{2}) \bigcup (\frac{1}{2}, 1)$, the operational inequality $N_{Alt}(\mathcal{W}, \epsilon, n) \geqslant N(\mathcal{W}, \epsilon, n)$ for $n \in \mathbb{N}$ is clear. We prove the inequality
$N_{Alt}(\mathcal{W}, \epsilon, n) \leqslant N(\mathcal{W}, \epsilon, n)$. Let there be a sequence of (n, M_n)-codes with $\limsup_{n\to\infty} \overline{e}_n(\mathcal{W}) \leqslant \epsilon$. Hence we have

$$\overline{e}_n(\mathcal{W}) \leqslant \epsilon + \delta_n, \tag{5.42}$$

with $\delta_n \to 0$ as $n \to \infty$.

- *For $\epsilon \in (0, \frac{1}{2})$, let $\epsilon = \frac{1}{2} - \delta$ for some $\delta > 0$. This implies*

  $$\overline{e}_n(\mathcal{W}) \leqslant \frac{1}{2} - \delta + \delta_n. \tag{5.43}$$

  Choosing n large enough such that $\delta_n < \delta$, we conclude $\overline{e}_n(\mathcal{W}) < \frac{1}{2}$ and hence $N_{Alt}(\mathcal{W}, \epsilon, n) \leqslant N(\mathcal{W}, \epsilon, n)$.

- *For $\epsilon \in (\frac{1}{2}, 1)$, let $\epsilon = 1 - \delta'$ for some $\delta' > 0$. Here (5.42) implies*

  $$\overline{e}_n(\mathcal{W}) \leqslant 1 - \delta + \delta_n. \tag{5.44}$$

  Choosing n large enough such that $\delta_n < \delta$, we conclude $\overline{e}_n(\mathcal{W}) < 1$ and hence $N_{Alt}(\mathcal{W}, \epsilon, n) \leqslant N(\mathcal{W}, \epsilon, n)$.

We are done.

5.3.3. The case with informed decoder

In this section, a variation of our results are derived for the case where the decoder knows the state $s \in \mathcal{S}$ of the channel in use. As such, the decoding sets depend on $s \in \mathcal{S}$, yielding $D_{m,s} \subset \mathcal{Y}^n$, $m \in [M_n]$ such that $D_{m,s} \bigcap D_{m',s} = \varnothing$ for $m \neq m'$ and $\bigcup_{m\in[M_n]} D_{m,s} = \mathcal{Y}^n$ for $s \in \mathcal{S}$. In the following definitions, ID stands for *informed decoder*. We show that in this case, we do not need to require that the compound zero-error capacity $C_0(\mathcal{W})$ is strictly positive for our coding. After all this assumption was required to communicate channel state information to the decoder. We define the average probability of error with informed decoder for the compound channel $\mathcal{W}$ as

$$\bar{e}_n(\mathcal{W}, ID) = \max_{s\in\mathcal{S}} \frac{1}{M_n} \sum_{m\in[M_n]} e^{ID}_{m,s,n}$$

with

$$e^{ID}_{m,s,n} := \sum_{y^n\in D^c_{m,s}} W^n_s(y^n|x_m), s = 1, 2. \tag{5.45}$$

We define the following numbers for $n \in \mathbb{N}$ and $0 \leqslant \epsilon < 1$, one corresponding to the average error criterion and the next to the maximum error criterion. The following definitions lead to the traditional definitions of ϵ-capacity of the compound channel with informed decoder.

1. $N(\mathcal{W}, \epsilon, n, ID) := \max\{N \in \mathbb{N} : \exists(n, N) - \textit{code}$ for $\mathcal{W}$ with $\bar{e}_n(\mathcal{W}, ID) \leqslant \epsilon\}$,
2. $N^{\max}(\mathcal{W}, \epsilon, n, ID) := \max\{N \in \mathbb{N} : \exists(n, N) - \textit{code}$ for $\mathcal{W}$ with $\max_{s\in\mathcal{S}} \max_{m\in[N]} e^{ID}_{s,n}(m) \leqslant \epsilon\}$.

The capacity functions $C(\mathcal{W}, \epsilon, ID), C^{max}(\mathcal{W}, \epsilon, ID)$ are defined accordingly (see Definition 138). The following three statements are essential to our computability analysis of compound broadcast channel.

Lemma 158 ([Ahlswede(2015)]) *For $\mathcal{W} := \{W_1, W_2\} \subset \mathcal{CC}(\mathcal{X}, \{1, 2\}; \mathcal{Y})$, it holds*

$$C(\mathcal{W}, \epsilon, ID) = \begin{cases} C(\mathcal{W}) & \textit{for } 0 < \epsilon < 1/2 \\ \min_{s=1,2} C(W_s) & \textit{for } 1/2 < \epsilon < 1 \end{cases}. \tag{5.46}$$

We prove an achievability statement for $\epsilon = 1/2$. Here, given the fact that the decoder is informed, we do not need the assumption of $C_0(\mathcal{W}) > 0$.

Lemma 159 *For $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, \{1, 2\}; \mathcal{Y})$, it holds*

$$\liminf_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \frac{1}{2}, n, ID) \geqslant \max\{C(\mathcal{W}), \min_{s=1,2} C_0(W_s)\}. \tag{5.47}$$

Proof 160 *If $C(\mathcal{W}) \geqslant \min_{s=1,2} C_0(W_s)$, the assertion is clear, as by [Ahlswede(2015)], $C(\mathcal{W}) = C(\mathcal{W}, \epsilon < \frac{1}{2}) \leqslant C(\mathcal{W}, \frac{1}{2}, ID)$. Assume otherwise. By Theorem 141, for $\delta > 0$, there exists $l_0 \in \mathbb{N}$ such that for $l > l_0$, there exist (l, M_l)-codes $\{(u^{(s)}_j, \mathit{\Lambda}^{(s)}_j), i \in [M_l]\}$ for $s = 1, 2$, with $\frac{1}{l} \log M_l \geqslant \min_{s=1,2} C_0(W_s) - \delta$ and $\max_{s=1,2} \max_{m\in[M_l]} e_{l,s}(m) = 0$. Choose M_l to be an even number. Construct the (l, M_l)-code with informed decoder $\{(\mathring{u}_m, \mathit{\Lambda}^{(s)}_m), m \in [M_l], s = 1, 2\}$ as*

- *$\mathring{u}_m := u^{(1)}_m$, for $m \in \{1, \ldots, \frac{M_l}{2}\}$,*

- $\mathring{u}_m := u_m^{(2)}$, *for* $m \in \{\frac{M_l}{2}+1, \ldots, M_l\}$.

We calculate the error due to this code. We have the following average probability of success for W_1*:*

$$\begin{aligned}\frac{1}{M_l}\Big(\sum_{m\in[M_l]} 1 - e_{l,1}^{ID}(m)\Big) &= \frac{1}{M_l}\sum_{m=1}^{\frac{M_l}{2}}\sum_{y^l\in\mathcal{A}_m^{(1)}} W_1^l(y^l|u_m^{(1)}) \\ &+ \frac{1}{M_1}\sum_{m=\frac{M_l}{2}+1}^{M_l}\sum_{y^l\in\mathcal{A}_m^{(1)}} W_1^l(y^l|u_m^{(2)}) \\ &\geqslant \frac{1}{M_l}\sum_{m=1}^{\frac{M_l}{2}}\sum_{y^l\in\mathcal{A}_m^{(1)}} W_1^l(y^l|u_m^{(1)}) = 1/2. \end{aligned} \tag{5.48}$$

Similar calculation yields the same lower bound on $\frac{1}{M_l}\big(\sum_{m\in[M_l]} 1 - e_{l,2}^{ID}(m)\big)$ *and hence we conclude* $\bar{e}(\mathcal{W}, ID) \leqslant \frac{1}{2}$*. Hence, we have*

$$\begin{aligned}\liminf_{l\to\infty}\frac{1}{l}\log N(\mathcal{W},\frac{1}{2},l,ID) &\geqslant \liminf_{l\to\infty}\frac{1}{l}\log M_l \\ &\geqslant \min_{s=1,2} C_0(W_s) - \delta. \end{aligned} \tag{5.49}$$

As $\delta > 0$ *was arbitrary, we are done.*

The case with informed encoder can also be considered when the decoder has channel state information. We derive similar results.

Theorem 161 *For* $\mathcal{W} \in \mathcal{C}(\mathcal{X}, \{1,2\}; \mathcal{Y})$ *it holds,*

$$C_{Alt}(\mathcal{W},\epsilon,ID) = \begin{cases} C(\mathcal{W}) & \text{for } 0 \leqslant \epsilon < 1/2 \\ \min_{s=1,2} C(W_s) & \text{for } 1/2 \leqslant \epsilon < 1 \end{cases}. \tag{5.50}$$

Proof 162 *First we show the statement for* $\epsilon = 1/2$*. In this case, from Lemma 149 we have*
$\liminf_{n\to\infty}\frac{1}{n}\log N_{Alt}(\mathcal{W},\frac{1}{2},n,ID) \geqslant \liminf_{n\to\infty}\frac{1}{n}\log N_{Alt}(\mathcal{W},\frac{1}{2},n) \geqslant \min_{s=1,2} C(W_s)$*, where the last inequality is from Lemma 149. To see the converse, let* $\mathcal{C}_{n,ID}$ *be a sequence of* (n, M_n)*-codes with informed decoder for* $\mathcal{W}$ *with* $\limsup_{n\to\infty}\bar{e}_n(\mathcal{W}, ID) = \frac{1}{2}$*. In other words we have* $\bar{e}_n(\mathcal{W}, ID) = \frac{1}{2} + \delta_n$ *with* $\delta_n \to 0$ *as* $n \to \infty$*. Choosing* n *large enough such that* $\delta_n < \frac{1}{2}$*, we obtain* $\bar{e}_n(\mathcal{W}, ID) < 1$*. Hence for large enough values of* n *we have* $N_{Alt}(\mathcal{W},\frac{1}{2},n,ID) \leqslant N(\mathcal{W},\epsilon,n,ID)$ *with* $\epsilon < 1$*. The proof then follows from Lemma 5.46. Therefore we have*
$\lim_{n\to\infty}\frac{1}{n}\log N_{Alt}(\mathcal{W},\frac{1}{2},n,ID) = \min_{s=1,2} C(W_s)$.
For $\epsilon \in (0,\frac{1}{2})\bigcup(\frac{1}{2},1)$*, the operational inequality* $N_{Alt}(\mathcal{W},\epsilon,n,ID) \geqslant N(\mathcal{W},\epsilon,n,ID)$ *for*

$n \in \mathbb{N}$ is clear. We prove the inequality
$N_{Alt}(\mathcal{W}, \epsilon, n, ID) \leqslant N(\mathcal{W}, \epsilon, n, ID)$.
Let there be a sequence of (n, M_n)-codes with $\limsup_{n\to\infty} \bar{e}_n(\mathcal{W}, ID) \leqslant \epsilon$. *Hence we have*

$$\bar{e_n}(\mathcal{W}, ID) \leqslant \epsilon + \delta_n, \tag{5.51}$$

with $\delta_n \to 0$ as $n \to \infty$.

- *For $\epsilon \in (0, \frac{1}{2})$, let $\epsilon = \frac{1}{2} - \delta$ for some $\delta > 0$. This implies*

  $$\bar{e}_n(\mathcal{W}, ID) \leqslant \frac{1}{2} - \delta + \delta_n. \tag{5.52}$$

  Choosing n large enough such that $\delta_n < \delta$, we conclude $\bar{e}_n(\mathcal{W}, ID) < \frac{1}{2}$ and hence $N_{Alt}(\mathcal{W}, \epsilon, n, ID) \leqslant N(\mathcal{W}, \epsilon, n, ID)$.

- *For $\epsilon \in (\frac{1}{2}, 1)$, let $\epsilon = 1 - \delta'$ for some $\delta' > 0$. Here (5.51) implies*

  $$\bar{e}_n(\mathcal{W}, ID) \leqslant 1 - \delta + \delta_n. \tag{5.53}$$

  Choosing n large enough such that $\delta_n < \delta$, we conclude $\bar{e}_n(\mathcal{W}, ID) < 1$ and hence $N_{Alt}(\mathcal{W}, \epsilon, n, ID) \leqslant N(\mathcal{W}, \epsilon, n, ID)$.

The proof follows from Lemma 5.46.

Example 163 ***An example of two-state compound channel ($|\mathbf{S}| = 2$), with two dimensional input-output alphabets ($|\mathcal{X}| = |\mathcal{Y}| = 2$), where $\mathbf{C}(\mathcal{W}) < \min_{\mathbf{s}=1,2} \mathbf{C}(\mathbf{W_s})$***
Consider the compound channel generated by $\mathcal{W} := \{W_1, W_2\} \subset \mathcal{CH}(\mathcal{X}, \mathcal{S}; \mathcal{Y}), \mathcal{X} = \mathcal{Y} = \{1, 2\}$, defined by the following stochastic matrices:

$$W_1 := \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix} \quad W_2 := \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}. \tag{5.54}$$

We have

$$\max_{p\in\mathcal{P}(\mathcal{X})} I(p, W_s) = \log(5/4), s = 1, 2, \tag{5.55}$$

with maximum taking place at $p = (3/5, 2/5)$ and $p' = (2/5, 3/5)$ for W_1 and W_2 respectively.Therefore we have
$\max_{p\in\mathcal{P}(\mathcal{X})} \min_{s=1,2} I(p, W_s) < \min_{s=1,2} \max_{p\in\mathcal{P}(\mathcal{X})} I(p, W_s)$. *We also note that $C_0(W_s) = 0$, $s = 1, 2$. From the capacity results of this section, we conclude for this example,*

$$C_{Alt}(\mathcal{W}, \epsilon < 1/2) = C_{Alt}(\mathcal{W}, \epsilon < 1/2, ID) < C_{Alt}(\mathcal{W}, \epsilon \geqslant 1/2). \tag{5.56}$$

The following example is from [Ahlswede(2015)]. We state this example here for later references.

Example 164 ***An example where*** $\mathbf{C}(\mathcal{W}, \in < \frac{1}{2}) < \mathbf{C}(\mathcal{W}, \in \geqslant \mathbf{1/2})$
Consider the compound channel generated by $\mathcal{W} := \{W_1, W_2\} \subset \mathcal{CH}(\mathcal{X}, \{1,2\}; \mathcal{Y}), \mathcal{X} = \mathcal{Y} = \{1, \ldots, 5\}$, *defined by the following stochastic matrices:*

$$W_1 := \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad W_{2,\delta} := \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.57}$$

It was show in [Ahlswede(2015)] that for this example, we have

$$C(\mathcal{W}, \epsilon < 1/2) < \min_{s=1,2} C(W_s) = C(\mathcal{W}, \epsilon \geqslant 1/2) = \log 3. \tag{5.58}$$

5.4. Banach-Mazur computability of ϵ-capacity

As explained in the previous section, in general the strong converse for compound channels under the average error criterion does not hold, i.e. there exist channel $\mathcal{W}_*$ and error $\epsilon_* \in (0,1)$ with

$$C(\mathcal{W}_*, \epsilon_*) > C(\mathcal{W}_*).$$

This is also true about $C_{Alt}(\mathcal{W}, \epsilon)$. This can be contrasted with the complete characterization that exists for $C^{\max}(\mathcal{W}, \epsilon)$ and $C^{\max}_{Alt}(\mathcal{W}, \epsilon)$ given by Theorems 139 and 146. Trying to address this problem, Ahlswede raised the following question [Ahlswede(2015), Secion 3]:

Ahlswede's Question: Does a (simple) recursive formula exist for the ϵ-capacity $C(\mathcal{W}, \epsilon)$ of the compound channel $\mathcal{W}$?

From a practical point of view, this is an important question. It is relevant to compute the ϵ-capacity of compound channels, since practical systems will always be designed such that they tolerate a certain fixed decoding error ϵ. However, in what follows, we will obtain a negative answer to Ahlswede's question. As a consequence, there exists no formula for the ϵ-capacity $C(\mathcal{W}, \epsilon)$ which is in contrast to the maximal error criterion capacity $C^{\max}(\mathcal{W}, \epsilon)$. The negative answer is provided to computability of $C_{Alt}(\mathcal{W}, \cdot)$, namely, the function is not computable in its error input. This is the subject of the following theorem.

Theorem 165 *The following statements hold.*

- *Given* $\mathcal{W} \in \mathcal{C}(\mathcal{X}, \mathcal{S}; \mathcal{Y})$ *and* $\hat{\epsilon} > 0$, *the capacity functions* $C(\mathcal{W}, \cdot) : (\hat{\epsilon}, 1) \cap \mathbb{R}_c \to \mathbb{R}$ *and* $C_{Alt}(\mathcal{W}, \cdot) : (\hat{\epsilon}, 1) \cap \mathbb{R}_c \to \mathbb{R}$, *are not in general Banach-Mazur computable.*

Proof 166 *We first prove the statement for C_{Alt}, for all $|\mathcal{X}| \geqslant 2, |\mathcal{Y}| \geqslant 2$ and $|\mathcal{S}| \geqslant 2$. Consider the compound channel $\mathcal{W}$ given by Example 163, given by (5.54). First consider the error $\epsilon_* = 1/2$. As shown in the example, we have $C(\mathcal{W}, \epsilon_*) = \log(5/4)$. Also, for $k \in \mathbb{N}$, consider $\epsilon_k := \frac{1}{2} - \frac{1}{2^k}$, $k \geqslant 2$. As such we have*

$$|\epsilon_* - \epsilon_k| = \frac{1}{2^k}, \quad \lim_{k \to \infty} |\epsilon_* - \epsilon_k| = 0.$$

Given the calculation done for (5.54) , we have $\forall k \in \mathbb{N}, k \geqslant 2$, $C(\mathcal{W}, \epsilon_k) = \log(5/4) - b$, for some constant $b > 0$. The remainder of the proof follows by contradiction. For this purpose, assume $C(\mathcal{W}, \cdot)$ is Banach-Mazur computable. This implies that the computable sequence $\{\epsilon_k\}_{k \in \mathbb{N}} \subset \mathbb{I}_c^k$ is mapped into a computable sequence $\{C(\mathcal{W}, \epsilon_k)\}_{k \in \mathbb{N}}$ of real numbers.

Let $\mathcal{A} \subset \mathbb{N}$ be an arbitrary recursively enumerable set such that $\mathcal{A}$ is not recursive, i.e. $\mathcal{A}^c$ is not recursively enumerable. This means (see Definition 107), we can construct a total function g, i.e. domain$(g) = \mathbb{N}$, such that $g([\mathbb{N}]) = \mathcal{A}$ and g is recursive and therewith a computable function. Furthermore, without loss of generality, we can require that $g : \mathbb{N} \to \mathcal{A}$ is a one-to-one mapping from $\mathbb{N}$ to $\mathcal{A}$.

To show that the ϵ-capacity is not Banach-Mazur computable, we present the following construction, originally conceived in [Pour-El and Richards(2017)]. For every $(k, l) \in \mathbb{N} \times \mathbb{N}$ we define the computable function $q : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ as

$$q(k,l) := \begin{cases} 2^{l+2} & k \notin \{g(0), \ldots, g(2^{l+2})\} \\ r & k \in \{g(0), \ldots, g(2^{l+2}) \text{ and } g(r) = k \end{cases}. \tag{5.59}$$

Note that r above is unique. Since $\mathcal{A}$ is recursively enumerable, the function q is indeed recursive and therewith computable.

Next we consider the double sequence $\{\epsilon_{q(k,l)}\}_{(k,l) \in \mathbb{N} \times \mathbb{N}}$ of errors, that is effectively computable, given that q is a recursive function. The idea is, that for each $k \in \mathbb{N}$, $\{\epsilon_{q(k,l)}\}_{l \in \mathbb{N}}$, converges effectively to some $\hat{\epsilon}_k \in \mathbb{I}_c$. Furthermore, the sequence $\{\hat{\epsilon}_k\}_{k \in \mathbb{N}}$ is a sequence in $\mathbb{I}_c$. Now we construct for every $k \in \mathbb{N}$, a computable function ϕ_k such that for all $k \in \mathbb{N}$, it holds

$$|\epsilon_{q(k,l)} - \hat{\epsilon}_k| < \frac{1}{2^m},$$

for all $l \geqslant \phi_k(m)$. We particularly note that the function $\phi(k, m) = \phi_k(m)$ with $\phi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is a computable function in both arguments.

For $k \in \mathcal{A}$ let l_0 be the smallest natural number such that $k \in \{g(0), \ldots, g(2^{l_0+2})\}$ is satisfied. Now, for all $l \geqslant l_0$ we have $q(k,l) = r$ and therewith also

$$|\epsilon_{q(k,l)} - \epsilon_r| = 0 < \frac{1}{2^l}. \tag{5.60}$$

On the other hand, for $l < l_0$ we have $k \notin \{g(0), \ldots, g(2^{l+2})\}$ so that $\epsilon_{q(k,l)} = \epsilon_{2^{l+2}}$.

Accordingly,

$$\begin{aligned}
|\epsilon_{q(k,l)} - \epsilon_r| &= |\epsilon_{2^{l+2}} - \epsilon_r| \\
&= |\epsilon_{2^{l+2}} - \epsilon_* + \epsilon_* - \epsilon_r| \\
&\leqslant |\epsilon_{2^{l+2}} - \epsilon_*| + |\epsilon_* - \epsilon_{2^{l+2}}| \qquad (5.61) \\
&= \frac{1}{2^{2^{l+2}}} + \frac{1}{2^r} \\
&< \frac{2}{2^{2^{l+2}}} < \frac{1}{2^l}, \qquad (5.62)
\end{aligned}$$

since for $l < l_0$ we have $g(r) = k$ and it must hold $r > 2^{l+2}$, and $2^{l+2} > l$.
Now we consider the case $k \in \mathcal{A}^c$. Here we have $q(k,l) = 2^{l+2}$ so that

$$|\epsilon_{q(k,l)} - \epsilon_*| = \frac{1}{2^{2^{l+2}}} < \frac{1}{2^l}. \tag{5.63}$$

We consider the sequence $\{\hat{\epsilon}_k\}_{k\in\mathbb{N}}$ with

$$\hat{\epsilon}_k = \begin{cases} \epsilon_* & \text{if } k \in \mathcal{A}^c \\ \epsilon_r & \text{if } k \in \mathcal{A} \text{ and } g(r) = k. \end{cases} \tag{5.64}$$

From (5.60),(5.61) and (5.63), we therefore have for arbitrary $k \in \mathbb{N}$ and arbitrary $l \in \mathbb{N}$,

$$|\epsilon_{q(k,l)} - \hat{\epsilon}_k| \leqslant \frac{1}{2^l},$$

i.e. we have for sequence $\{\epsilon_{q(k,l)}\}_{l\in\mathbb{N}} \subset \mathbb{I}_c$, computable convergence to $\hat{\epsilon}_k$ for every $k \in \mathbb{N}$. With this we have $\phi_k(m) = m$, since for all $l \geqslant \phi_k(m)$, we obtain

$$|\epsilon_{q(k,l)} - \hat{\epsilon}_k| \leqslant \frac{1}{2^l} \leqslant \frac{1}{2^m}, \tag{5.65}$$

i.e. we have effective convergence and the function ϕ_k is further, independent of k. This immediately implies that $\hat{\epsilon}_k \in \mathbb{I}_c$. Furthermore, we observe that the sequence $\{\hat{\epsilon}_k\}_{k\in\mathbb{N}}$ is computable as well and that the effective convergence speed can be bounded universally, i.e. independent of k.
Accordingly,$\Phi : \mathbb{N} \to \{0,1\}$, with $\Phi(k) = \frac{1}{b}[C(\mathcal{W}, \hat{\epsilon}_k) - \log(5/4) + b], k \in \mathbb{N}$, is a recursive function. This comes from the fact that multiplication and addition of a recursive function, namely here by assumption $C_{Alt}(\mathcal{W}, \cdot)$, is recursive. For arbitrary $k \in \mathbb{N}$, we have

$$\Phi(k) = 1 \iff k \in \mathcal{A}^c \tag{5.66}$$

$$\Phi(k) = 0 \iff k \in \mathcal{A}. \tag{5.67}$$

This means, that the characteristic function of the set $\mathcal{A}^c$, and hence that of $\mathcal{A}$ are recursive, which is a contradiction, as we started with the assumption that $\mathcal{A}$ was recursively

enumerable and not recursive. This proves the desired result for $|\mathcal{X}| = |\mathcal{Y}| = 2$. The proof immediately extends to general case $|\mathcal{X}|, |\mathcal{Y}| \geqslant 2$ by zero adding to the channel.
For function $C(\cdot, \mathcal{W})$, we follow the same strategy, only using Example 164 instead.

5.5. Computable upper and lower bound for ϵ-capacity

From a practical point of view, it is of course interesting and relevant to obtain computable lower bounds (e.g. based on improved coding schemes) and computable upper bounds on the ϵ-capacity.
Since $(\mathfrak{W}, D_{\mathfrak{W}})$ is a compact Hausdorff space [Rudin(1987)], we can apply our previous findings in Section 5.2.4 to the compound channel. In what follows, we consider computability of upper and lower bounds on the capacity function, for a fixed compound channel input and as a function of ϵ We will see that for the ϵ-capacity of the compound channel, there is either no computable achievability or no computable converse.
Here, the functions $\{F_N\}$ can be interpreted as lower bounds for achievable rates and the ϵ-capacity respectively. Of course, such bounds should be effectively computable so that they can be evaluated on a digital computer. These bounds should improve with increasing $N \in \mathbb{N}$, i.e., $F_N(\mathcal{W}, \epsilon) \leqslant F_{N+1}(\mathcal{W}, \epsilon)$, $(\mathcal{W}, \epsilon) \in \mathfrak{W}_c$, and further should be asymptotically tight, i.e., for $N \to \infty$ the sequence $\{F_N\}_{N\in\mathbb{N}}$ should converge point-wise to $C(\mathcal{W}, \epsilon)$.
Accordingly, the functions $\{G_N\}$ can be interpreted as upper bounds on the achievable rates and the ϵ-capacity, respectively. Similarly, one would like to have these bounds to be effectively computable and further $C(\mathcal{W}, \epsilon) \leqslant G_{N+1}(\mathcal{W}, \epsilon) \leqslant G_N(\mathcal{W}, \epsilon)$, $(\mathcal{W}, \epsilon) \in \mathfrak{W}_c$, i.e., the bounds should improve with increasing $N \in \mathbb{N}$.
Let $\hat{\epsilon} \in (0, 1)$ be an arbitrary computable real number and we will study the behavior of $C(\mathcal{W}, \epsilon)$ for $\epsilon \in \mathbb{I}_c(\hat{\epsilon}) = [\hat{\epsilon}, 1]$. We set

$$\mathfrak{W}_c(\hat{\epsilon}) = \quad (\mathcal{W}, \epsilon) : \mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, \mathcal{S}; \mathcal{Y}), \epsilon \in \mathbb{I}_c(\hat{\epsilon})\big\}.$$

Clearly if the desirable bounds cannot be computable as a function of ϵ, they cannot be computable as a function of $(\epsilon, \mathcal{W})$.

Theorem 167 *Let $\hat{\epsilon} \in (0, 1/2)$ be a computable number. There exists a computable compound channel $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, S; \mathcal{Y})$, such that for all sequences $\{G_N\}_{N\in\mathbb{N}}, \{F_N\}_{N\in\mathbb{N}}$ of Banach-Mazur computable functions with*

- $G_N : [\hat{\epsilon}, 1] \to \mathbb{R}_c$, $F_N : [\hat{\epsilon}, 1] \to \mathbb{R}_c$, $N \in \mathbb{N}$ *and*
- $F_N \leqslant C(\mathcal{W}, \epsilon) \leqslant G_N(\epsilon)$, $\epsilon \in [\hat{\epsilon}, 1]$, $N \in \mathbb{N}$,

we have

$$\inf_{N\in\mathbb{N}} \big(G_N(1/2) - F_N(1/2)\big) > 0.$$

Proof 168 *Consider $\mathcal{W}$ from Example 164. Let $\{G_N\}_{N\in\mathbb{N}}, \{F_N\}_{N\in\mathbb{N}}$ be arbitrary BM computable sequences as described in the statement. We have for $\epsilon > 1/2$*

$$F_N(\epsilon) \leqslant C(\mathcal{W}, \epsilon) = C(\mathcal{W}, 1/2) = \log 3. \tag{5.68}$$

Since F_N is BM computable on $[\hat{\epsilon}, 1] \cap \mathbb{R}_c$, we have for all $N \in \mathbb{N}$, for sequence $(\epsilon_k)_{k=1}^{\infty}, \epsilon_k := 1/2 - (1/2^k)$:

$$F_N(1/2) = \lim_{k\to\infty} F_N(\epsilon_k) \leqslant \lim_{k\to\infty} C(\mathcal{W}, \epsilon_k) = \log 3 - b, b > 0. \tag{5.69}$$

We also have for all $N \in \mathbb{N}$

$$G_N(1/2) \geqslant C(\mathcal{W}, 1/2) = \log 3. \tag{5.70}$$

From (5.69) and (5.70) we conclude $\forall N \in \mathbb{N}$,

$$G_N(1/2) - F_N(1/2) = b > 0. \tag{5.71}$$

Remark 169 *In the proof of Theorem 167, we can use 163 to prove the same statement about $C_{Alt}(\cdot, \mathcal{W})$.*

Theorem 170 *Let $|\mathcal{X}| \geqslant 2, |\mathcal{Y}| \geqslant 2$ and $|S| \geqslant 2$. Let $\hat{\epsilon} \in (0, 1/2)$ be a computable number. There exists a computable compound channel $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, S; \mathcal{Y})$, such that for all sequences $\{G_N\}_{N\in\mathbb{N}}, \{F_N\}_{N\in\mathbb{N}}$ of Banach-Mazur computable functions with*

- *$G_N : [\hat{\epsilon}, 1] \to \mathbb{R}_c$, $F_N : [\hat{\epsilon}, 1] \to \mathbb{R}_c$, $N \in \mathbb{N}$ and*
- *$F_N \leqslant C_{Alt}(\mathcal{W}, \epsilon) \leqslant G_N(\epsilon)$, $\epsilon \in [\hat{\epsilon}, 1]$, $N \in \mathbb{N}$,*

we have

$$\inf_{N\in\mathbb{N}} \big(G_N(1/2) - F_N(1/2)\big) > 0.$$

As a consequence, we cannot have a capacity theorem, where the computable achievability bound could be equal to that of the converse. This in turn makes the entropic characterization of the capacity function impossible. However, Theorem 167, does not rule out the possibility, that at least one of the bounds (either achievability or converse) would converge to the ϵ-capacity of the compound channel.

5.6. Decision problem

Algorithmic computability of the ϵ-capacity, as considered so far in this work, is a strong condition that may be more that what is required to be satisfied in many applications. Relaxing this condition, we might only want to know if it is possible to algorithmically decide, if the ϵ-capacity of the channel is below or above a certain threshold $\lambda > 0$. More precisely, we ask

Question 1: Is there an algorithm (or Turing machine) TM, that for all $\epsilon \in (0,1) \cap \mathbb{R}_c$, takes the compound channel $\mathcal{W}$ and the threshold requirement $\lambda > 0$ as inputs and outputs *yes*, if the channel satisfies the threshold requirement, i.e. whenever $C(\mathcal{W}, \epsilon) > \lambda$, and outputs *no*, if the channel does not satisfy the threshold requirement, i.e. whenever $C(\mathcal{W}, \epsilon) < \lambda$?

Similarly, we can ask the question, about the ϵ input:

Question 1': Is there an algorithm (or Turing machine) TM, that for all $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, S; \mathcal{Y})$, takes $0 < \epsilon < 1$ and the threshold requirement $\lambda > 0$ as inputs and outputs *yes*, if the channel satisfies the threshold requirement, i.e. whenever $C(\mathcal{W}, \epsilon) > \lambda$, and outputs *no*, if the channel does not satisfy the threshold requirement, i.e. whenever $C(\mathcal{W}, \epsilon) < \lambda$?

This is a decision problem where we look for a Turing machine that decides whether or not a compound channel, or an average error upper-bound, satisfy a certain requirement. As such the question of decidability has important implications for resource-allocation. [1] Here, the Turing machine needs to stop for every channel-error pair, outputting the correct answer. However, such a Turing machine may not always exist and the question above may then be undecidable. In this case, one may be inclined to modify the question by weakening the assumptions. We only ask the weaker question for the channel input here, noticing that as before, similar question can be asked about the error input.

Question 2a: Is there an algorithm (or Turing machine) TM that for $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, \mathcal{S}; \mathcal{Y})$, takes error input $0 < \epsilon < 1$ and the threshold requirement $\lambda > 0$ as inputs, and stops if $C(\mathcal{W}, \epsilon) > \lambda$?

While Question 1 (equivalently Question 1') asks whether or not the problem is decidable, the modified Question 2a asks whether or not the problem is *semi-decidable*. This means the Turing machine is only required to stop and to provide the correct answer whenever the capacity does satisfy the requirement, i.e. whenever $C(\mathcal{W}, \epsilon) > \lambda$. In the other case, the Turing machine does not stop at all. Obviously, this is not the only way to pose the semi-decidability. One can also modify the initial question in the opposite way by requiring the Turing machine to stop only whenever the channel (or equivalently the error input) does not satisfy the requirement $C(\mathcal{W}, \epsilon) < \lambda$. Accordingly, in the other

[1] Note that in Question 1 and Question 1', all pairs $(\mathcal{W}, \epsilon)$ with $C(\mathcal{W}, \epsilon) = \lambda$ are excluded on purpose, as input to the Turing machine TM. The problem of deciding whether or not an expression is equal to a given real number is undecidable in general [Pour-El and Richards(2017)].

case, the Turing machine does not stop.

Question 2b: Is there an algorithm (or Turing machine) TM that for $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, \mathcal{S}; \mathcal{Y})$, takes error input $0 < \epsilon < 1$ and the threshold requirement $\lambda > 0$ as inputs and stops if the channel does not support the requirement $C(\mathcal{W}, \epsilon) < \lambda$?

Remark 171 *Note that if both Questions 2a and 2b have a positive answer, i.e. both problems are semi-decidable, it immediately implies that the initial problem is decidable and Question 1(a,b) has a positive answer as well.*

Again, in much the same way that Question 1' related to Question 1, we might ask Questions 2a' and 2b', that ask the same as in Questions 2a and 2b about the channel input. We answer the question of semi-decidability, when asked about the error input, negatively.

Theorem 172 *There exists a compound channel $\mathcal{W} \in \mathcal{CC}_c(\mathcal{X}, S; \mathcal{Y})$ and $\lambda \in \mathbb{R}_c$, such that $\mathcal{E}(\lambda) := \{\epsilon : C(\mathcal{W}, \epsilon) > \lambda\}$ and $\mathcal{E}'(\lambda) := \{\epsilon : C_{Alt}(\mathcal{W}, \epsilon) > \lambda\}$ are not semi-decidable.*

Proof 173 *We will prove the above theorem for infinitely many $\lambda \in \mathbb{R}_c$. We prove the statement for $C(\mathcal{W}, \epsilon)$ using the compound channel from Example 164. We remark that using the same construction, given Example 163, the same statement can be shown for $C_{Alt}(\mathcal{W}, \epsilon)$. Let $\lambda \in \mathbb{R}_c$ be such that*

$$C(\mathcal{W}, \epsilon < 1/2) < \lambda < C(\mathcal{W}, 1/2) = \log(3). \tag{5.72}$$

Assume that $\mathcal{E}(\lambda)$ is semi-decidable. For $\epsilon \in [1/2, 1)$, we construct a Turing machine TM_* that runs the following Turing machines in parallel.

1. *TM_1, that stops exactly when $\epsilon < 1/2$. This machine exists and runs the following algorithm. Since λ is a computable real, there exists a computable sequence of rational numbers $\{r(n)\}_{n\in\mathbb{N}}$ such that $|\lambda - r(n)| < 2^{-n}$ for all $n \in \mathbb{N}$. Furthermore, there exist recursive functions $a, b : \mathbb{N} \to \mathbb{N}$ with $b(n) \neq 0$ for all $n \in \mathbb{N}$ and*

$$r(n) = \frac{a(n)}{b(n)}, \quad n \in \mathbb{N}. \tag{5.73}$$

We now specify an algorithm that stops at $\epsilon < 1/2$. It holds that $\lambda - C(\mathcal{W}, \epsilon < \frac{1}{2}) \in \mathbb{R}_c$, as both arguments of the subtraction are computable reals. We set

$$\Delta(n) := \frac{a(n)}{b(n)} - C(\mathcal{W}, \epsilon).$$

The algorithm run by TM_1 computes $\Delta(1)$. If $\Delta(1) > 1/2$, then it stops. Otherwise the algorithm computes $\Delta(2)$. If $\Delta(2) > \frac{1}{2^2}$, it stops. Assuming the algorithm has not stopped in l steps, it computes $\Delta(l+1)$. If $\Delta(l+1) > \frac{1}{2^{l+1}}$, then it stops. With this

description, the following holds. The algorithm stops if and only if $\lambda - C(\mathcal{W}, \epsilon) > 0$ because:
$\Rightarrow$: if algorithm stops, we have found n_0 such that

$$\begin{aligned}\frac{1}{2^{n_0}} < \varDelta(n_0) &= \frac{a(n_0)}{b(n_0)} - C(\mathcal{W}, \epsilon)\\ &= -\lambda + \lambda + \frac{a(n_0)}{b(n_0)} - C(\mathcal{W}, \epsilon)\\ &< \lambda + |\frac{a(n_0)}{b(n_0)} - \lambda| - C(\mathcal{W}, \epsilon)\\ &< \lambda - C(\mathcal{W}, \epsilon) + \frac{1}{2^{n_0}},\end{aligned}$$

where in the last line, we have inserted $|\lambda - \frac{a(n_0)}{b(n_0)}| < \frac{1}{2^{n_0}}$. Therefore $\lambda - C(\mathcal{W}, \epsilon) > 0$.
$\leftarrow$: if $\lambda - C(\mathcal{W}, \epsilon) > 0$, then there is an n_0 such that

$$\lambda - C(\mathcal{W}, \epsilon) > \frac{2}{2^{n_0}}.$$

Therefore,

$$\begin{aligned}\frac{2}{2^{n_0}} < \lambda - C(\mathcal{W}, \epsilon) &= \frac{a(n_0)}{b(n_0)} - \frac{a(n_0)}{b(n_0)} + \lambda - C(\mathcal{W}, \epsilon)\\ &\leqslant \frac{a(n_0)}{b(n_0)} + |\lambda - \frac{a(n_0)}{b(n_0)}| - C(\mathcal{W}, \epsilon)\\ &< \frac{a(n_0)}{b(n_0)} - C(\mathcal{W}, \epsilon) + \frac{1}{2^{n_0}}.\end{aligned}$$

Therefore $\frac{a(n_0)}{b(n_0)} - C(\mathcal{W}, \epsilon) > \frac{1}{2^{n_0}}$, which stops the algorithm.

2. *TM_2, that stops when $\epsilon = 1/2$. This Turing machine exists because by assumption, $\mathcal{E}(\lambda)$ is semi-decidable. This means there exists a Turning machine that stops, when $C(\mathcal{W}, \epsilon) > \lambda$ and hence by our choice of λ and $\mathcal{W}$, $\epsilon = 1/2$.*

We define for $\epsilon \in [1/2, 1)$

$$TM_* := \begin{cases} yes & \text{if } TM_1 \text{ stops} \\ no & \text{if } TM_2 \text{ stops} . \end{cases} \tag{5.74}$$

Therefore we have

$$TM_*(\epsilon) = \begin{cases} yes & \text{if } \epsilon < 1/2 \\ no & \text{if } \epsilon = 1/2. \end{cases} \tag{5.75}$$

Taking the recursively enumerable set $\mathcal{A}$ and $\hat{\epsilon}_k$ from Theorem 165, we obtain

$$TM_*(\hat{\epsilon}_k) = \begin{cases} \textit{yes} & \textit{if } k \in \mathcal{A} \\ \textit{no} & \textit{if } k \in \mathcal{A}^c, \end{cases} \tag{5.76}$$

which is only the case if $\mathcal{A}$ is recursive, that runs contrary to our assumption. Therefore TM_* cannot exist. This can be extended to $|\mathcal{X}| \geqslant 5, |\mathcal{Y}| \geqslant 5, |S| \geqslant 2$.

is not semi-decidable.

5.7. Common randomness and entanglement assisted ϵ-capacities

In this section, we prove some important implications of our results for assisted scenarios. Here, we consider the ϵ-capacity of the channel, when the communicating parties have at their disposals pre-shared *common randomness* and *entanglement.* In both scenarios, we make use of the equivalence between a maximal and average error bound requirement.

Definition 174 *An (n, M_n) randomness assisted code for compound channel $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, S; \mathcal{Y})$, is a probability measure μ_n on $(\sigma_n \times \Omega_n, \sigma_n)$, where*

- $\sigma_n := \{E : [M_n] \to \mathcal{P}(\mathcal{X}^n)\}$,
- $\Omega_n := \{\phi : \mathcal{Y}^n \to \mathcal{P}([M_n])\}$,
- *and the sigma-algebra σ_n is chosen such that the function*

$$e_{m,s,n}(E, \phi) := 1 - \sum_{y^n, x^n} W_n^n(y^n|x^n)E(x^n|m)\phi(m|y^n) \tag{5.77}$$

  is measurable with respect to μ_n for all $s \in S$ and $m \in [M_n]$.[2]

- *We further require that the sigma-algebra contains all the singleton sets. The deterministic codes defined in Section 5.3 are then a specification of μ_n to a probability distribution that is equal to unity at a singleton element and zero otherwise.*

Let each singleton member be notified by (E, ϕ). The following is our average performance (error) function for the common randomness assisted task

$$\bar{e}_{cr,n}(\mathcal{W}) := \max_{s \in S} \int \frac{1}{M_n} \sum_{m \in [M_n]} e_{m,s,n}(E, \phi) d\mu(E, \phi). \tag{5.78}$$

We also consider the maximal error function defined for each $s \in S$ and $m \in [M_n]$ by

[2] Notice that $e_{m,s,n}(E, \phi) = e_{m,s,n}$ where the right hand side is defined by (5.17) with $E(x^n|m) = 1$ iff $x^n = x_m$ and $\phi(m|y^n) > 0$ iff $y^n \in D_m$.

$$e_{m,s,cr,n} := \int e_{m,s,n}(E,\phi)d\mu(E,\phi). \tag{5.79}$$

We define the following numbers for $n \in \mathbb{N}$ and $0 \leqslant \epsilon < 1$, corresponding to the common randomness assisted (CR) ϵ-capacity under average and maximal error criteria.

1. $N(\mathcal{W},\epsilon,cr,n) := \max\{N \in \mathbb{N} : \exists(n,N)$ CR *code* for $\mathcal{W}$ with $\overline{e}_{cr,n}(\mathcal{W}) \leqslant \epsilon\}$,
2. $N^{\max}(\mathcal{W},\epsilon,n) := \max\{N \in \mathbb{N} : \exists(n,N)$ CR *code* for $\mathcal{W}$ with $\max_{s\in\mathcal{S}} \max_{m\in[N]} e_{m,s,cr,n} \leqslant \epsilon\}$.

There could be two notions of capacity corresponding to these numbers that are defined in the following.

Definition 175 *Let $\mathcal{W} \in \mathcal{CC}(\mathcal{X},\mathcal{S};\mathcal{Y})$ and $0 \leqslant \epsilon < 1$. Then*

- $\overline{C(\mathcal{W},\epsilon,cr)} := \limsup_{n\to\infty} \frac{1}{n} \log N(\mathcal{W},\epsilon,cr,n)$ *and* $\overline{C^{max}(\mathcal{W},\epsilon,cr)} := \limsup_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W},\epsilon,cr,n)$ *are the optimistic CR ϵ-capacities of $\mathcal{W}$ under average and maximal error criteria respectively.*

- *Finally, $C(\mathcal{W},\epsilon,cr)$ and $C^{\max}(\mathcal{W},\epsilon,cr)$ are the CR ϵ-capacities of $\mathcal{W}$ under average and maximal error criteria, if*

$$\limsup_{n\to\infty} \frac{1}{n} \log N(\mathcal{W},\epsilon,cr,n) = \liminf_{n\to\infty} \frac{1}{n} \log N(\mathcal{W},\epsilon,cr,n) = C(\mathcal{W},\epsilon,cr)$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W},\epsilon,cr,n) = \liminf_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W},\epsilon,cr,n) = C^{\max}(\mathcal{W},\epsilon,cr)$$

respectively.

The definitions for zero-error CR capacity $C_0(\mathcal{W},cr)$ are given by setting $\epsilon = 0$ in the above given definitions.

Definition 176 *An entanglement-assisted (n,M)-Code for compound channel $\mathcal{W} \in \mathcal{CC}(\mathcal{X},S;\mathcal{Y})$ is a triple $(\mathcal{E},\mathcal{D},\Psi)$ with $\mathcal{E} = (E^m_{\mathbf{x}})_{\mathbf{x}\in\mathcal{X}^n,m\in[M]}$ being a family of M POVMs on Hilbert space $\mathcal{K}_A$ with outcomes in $\mathcal{X}^n$, $(D^{\mathbf{y}}_m)_{m\in[M],\mathbf{y}\in\mathcal{Y}^n}$ being a family of $|\mathcal{Y}^n|$ POVMs on Hilbert space $\mathcal{K}_B$ with outcomes in $[M]$ and Ψ being a quantum state on Hilbert space $\mathcal{K}_A \otimes \mathcal{K}_B$.*

Remark 177 *Since we do not assume any restrictions on the dimension of $\mathcal{K}_A \otimes \mathcal{K}_B$ other than being finite, we can without loss of generality assume Ψ to be a pure state.*

Otherwise, the communicating parties may resort to a purification of Ψ on a Hilbert space of larger dimension.

Similar to the previously discussed communication scenarios, we can assign a probability to the occurrence of an erroneous decoding. Given a compound channel $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, S; \mathcal{Y})$ with corresponding (n, M)-Code $(\mathcal{E}, \mathcal{D}, \Psi)$, define

$$\overline{e} := \max_{s\in\mathcal{S}} \frac{1}{M} \sum_{m\in[M]} \sum_{\mathbf{x}\in\mathcal{X}^n} \sum_{\mathbf{y}\in\mathcal{Y}^n} \operatorname{tr}\left(\left(E_{\mathbf{x}}^m \otimes D_m^{\mathbf{y}}\right)\Psi\right) W_s(\mathbf{y}|\mathbf{x}) \tag{5.80}$$

$$e := \max_{s\in\mathcal{S}} \max_{m\in[M]} \sum_{\mathbf{x}\in\mathcal{X}^n} \sum_{\mathbf{y}\in\mathcal{Y}^n} \operatorname{tr}\left(\left(E_{\mathbf{x}}^m \otimes D_m^{\mathbf{y}}\right)\Psi\right) W_s(\mathbf{y}|\mathbf{x}) \tag{5.81}$$

We define the following numbers for $n \in \mathbb{N}$ and $0 \leqslant \epsilon < 1$, one corresponding to the average-error criterion and one corresponding to the maximum-error criterion:

$$N(\mathcal{W}, \epsilon, n, EA) := \max\{M \in \mathbb{N} : \exists (n, M)\text{-Code for } \mathcal{W} \text{ satisfying } \overline{e} \leqslant \epsilon\}, \tag{5.82}$$

$$N^{\max}(\mathcal{W}, \epsilon, n, EA) := \max\{M \in \mathbb{N} : \exists (n, M)\text{-Code for } \mathcal{W} \text{ satisfying } e \leqslant \epsilon\}. \tag{5.83}$$

In Definition 138, we have introduced the optimistic and the pessimistic ϵ-capacity for the standard non-assisted compound channel. The definition of optimistic and pessimistic ϵ-capacities for the entanglement-assisted communication scheme follows the same line of reasoning.

Definition 178 *Let $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, \mathcal{S}, \mathcal{Y})$ be a compound channel and $0 \leqslant \epsilon < 1$ a real number. Then,*

- $\overline{C(\mathcal{W}, \epsilon, EA)} := \limsup_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, EA)$ *and* $\overline{C^{\max}(\mathcal{W}, \epsilon, EA)} := \limsup_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W}, \epsilon, EA)$ *are called the* optimistic ϵ-capacity *of $\mathcal{W}$ with respect to the average- and the maximum-error criterion, respectively;*

- $\underline{C(\mathcal{W}, \epsilon, EA)} := \liminf_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, EA)$ *and* $\underline{C^{\max}(\mathcal{W}, \epsilon, EA)} := \liminf_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W}, \epsilon, EA)$ *are called the* pessimistic ϵ-capacity *of $\mathcal{W}$ with respect to the average- and the maximum-error criterion, respectively;*

- *if the corresponding limit exists,* $C(\mathcal{W}, \epsilon, EA) := \lim_{n\to\infty} \frac{1}{n} \log N(\mathcal{W}, \epsilon, EA)$ *and* $C^{\max}(\mathcal{W}, \epsilon, EA) := \lim_{n\to\infty} \frac{1}{n} \log N^{\max}(\mathcal{W}, \epsilon, EA)$ *are called the* ϵ-capacity *of $\mathcal{W}$ with respect to the average- and the maximum-error criterion, respectively.*

Remark 179 *The capacities $C(\mathcal{W}, \epsilon, EA)$ and $C^{\max}(\mathcal{W}, \epsilon, EA)$ exist if and only if $\underline{C(\mathcal{W}, \epsilon, EA)} = \overline{C(\mathcal{W}, \epsilon, EA)}$, $\underline{C^{\max}(\mathcal{W}, \epsilon, EA)} = \overline{C^{\max}(\mathcal{W}, \epsilon, EA)}$, respectively, holds true.*

Setting ϵ to equal zero, we obtain the entanglement-assisted zero-error compound capacity
$C_0(\mathcal{W}, EA) := C(\mathcal{W}, 0, EA) = C^{max}(\mathcal{W}, 0, EA)$. This capacity exists in the sense of Remark 179, following from Fekete's lemma.

Based on these definitions, we prove the following results.

Lemma 180 *For $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, S; \mathcal{Y})$ it holds*

$$C(\mathcal{W}, \epsilon, cr) = C(\mathcal{W}, \epsilon), \epsilon \in [0, 1). \tag{5.84}$$

Proof 181 *The inequality $C(\mathcal{W}, \epsilon, cr) \geqslant C(\mathcal{W}, \epsilon)$ is operationally clear for all $\epsilon \in [0, 1)$. This is because the communicating parties can choose to ignore the common randomness resource available to them. To see the other inequality, we notice that for all $\epsilon \in [0, 1)$, if there exists a randomness assisted code μ_n, such that (5.78) is satisfied, then there must exist one singleton (E, ϕ) that fulfills the same upper bound.*

Theorem 182 *For $\mathcal{W} \in \mathcal{CC}(\mathcal{X}, S; \mathcal{Y})$ it holds*

$$C^{max}(\mathcal{W}, \epsilon, cr) = C(\mathcal{W}, \epsilon), \epsilon \in [0, 1). \tag{5.85}$$

Proof 183 *Let (E, ϕ) be an (n, M)-code for $\mathcal{W}$ with the average error upper bounded as*

$$\overline{e}_n(\mathcal{W}) = \max_{s \in \mathcal{S}} \frac{1}{M} \sum_{m \in [M]} e_{m,s,n}(E, \phi) \leqslant \epsilon. \tag{5.86}$$

We define the CR code $\tilde{\mu}$ on $(\tilde{\sigma} \times \tilde{\Omega}, \tilde{\sigma})$ with $\tilde{\sigma} := \{E^{(m)} : m \in [M]\}$,$\Omega := \{\phi^{(m')} : m' \in [M]\}$, defined by

- $E^{(i)}(\cdot|j) = E(\cdot|i)$,

- $\phi^{(i)}(j|\cdot) = \phi(i|\cdot)$ *and*

- $\tilde{\mu}(E^{(i)}, \phi^{(j)}) = \frac{\delta_{i,j}}{M}$,

for all $i, j \in [M]$ and $\delta_{i,j}$ the Kronecker delta function. Calculating the error due to this *code, for each $m \in [M]$ we obtain*

$$\begin{aligned}\max_{s\in\mathcal{S}} e_{m,s,cr,n} &= \max_{s\in\mathcal{S}} \int e_{m,s,n}(E^{(i)},\phi^{(j)})d\tilde{\mu}_{i,j} \\ &= \max_{s\in\mathcal{S}} \sum_{i,j} e_{m,s,n}(E^{(i)},\phi^{(j)})\tilde{\mu}(E^{(i)},\phi^{(j)}) \\ &= 1-\min_{s\in\mathcal{S}} \sum_{i,j}\sum_{x^n,y^n} W^n(y^n|x^n)E^{(i)}(x^n|m)\phi^{(j)}(m|y^n)\tilde{\mu}(E^{(i)},\phi^{(j)}) \\ &= 1-\min_{s\in\mathcal{S}} \frac{1}{M}\sum_{i\in[M]}\sum_{x^n,y^n} W_s^n(y^n|x^n)E^{(i)}(x^n|i)\phi^{(i)}(j|y^n) \\ &= 1-\min_{s\in\mathcal{S}} \frac{1}{M}\sum_{i\in[M]}\sum_{x^n,y^n} W_s^n(y^n|x^n)E(x^n|i)\phi(i|y^n) = \overline{e}_n(\mathcal{W}) \leqslant \epsilon. \qquad (5.87)\end{aligned}$$

The first line is the definition of the error function given by (5.79), the second line follows because of the discrete nature of the probability space, the second last line comes from the definition of $\tilde{\mu}$ and the last line comes from the defining property of our code with respect to (E,ϕ). Therefore, we have concluded that for every code for the compound channel that satisfies the upper bound on the average error criterion, we find a CR code that satisfies the upper bound on maximal error criterion. This shows the statement $C(\mathcal{W},\epsilon) \leqslant C^{\max}(\mathcal{W},\epsilon,cr)$. Given that the inequality $C^{\max}(\mathcal{W},\epsilon,cr) \leqslant C(\mathcal{W},\epsilon,cr)$ is operationally clear, the statement follows from Lemma 180.

The following statement, results from Theorem 182 and our results on the computability of the ϵ-capacity in Section 5.4.

Corollary 184 *The capacity function $C^{\max}(\cdot,\cdot,cr) : \mathcal{CC}(\mathcal{X},S;\mathcal{Y}) \times (0,1) \to \mathbb{R}$ is not in general Turing computable.*

Remark 185 *We note that for the compound channel $\mathcal{W}$ and $\epsilon \in (0,1)$, the capacity function $C^{\max}(\mathcal{W},\epsilon) = C(\mathcal{W})$ (see [Ahlswede(2015)]). Theorem 182 and Examples 163 and 164, show that pr-shared randomness improves the ϵ-capacity of the compound channel under maximal error criterion.*

We continue this section by deriving analogous results for entanglement assisted ϵ-capacity of the channel. The next lemma states that average and maximal error criteria yield the same EA assisted ϵ capacities.

Lemma 186 *Let $\mathcal{W} \in \mathcal{CC}(\mathcal{X},S;\mathcal{Y})$ and $\epsilon \in (0,1)$. It holds*

$$C^{\max}(\mathcal{W},\epsilon,EA) = C(\mathcal{W},\epsilon,EA). \qquad (5.88)$$

Proof 187 *See [H. Boche(2017)] Lemma 4. for a proof.*

Remark 188 *We note that for the compound channel $\mathcal{W}$ and $\epsilon \in (0,1)$, the capacity function $C^{\max}(\mathcal{W},\epsilon) = C(\mathcal{W})$ (see [Ahlswede(2015)]). Given Lemma 186 and Examples 164 or 163, entanglement can indeed improve the ϵ-capacity under maximal error criterion.*

6. Conclusions and outlooks

6.1. Chapter 2: Randomness cost of symmetric twirling

In this chapter, we derived upper and lower bounds of randomness consumption of universal symmetrization of states by applying averaged permutations. Our bounds lead to positive and negative conclusions when applied to information-theoretic modelling. First, the encouraging implication of our upper bound on the support of weighted designs shows, that there are always protocols which universally symmetrize quantum states on a given n-partite quantum system, consuming reasonable common randomness resource. Specifically, the number of coordinated random choices of permutations used for symmetrizing arbitrary quantum states on that system can be always restricted to being exponentially growing (with number of systems). The lower bounds on the common randomness needed for permutation-based symmetrization of arbitrary quantum states proven in this chapter, enforce a rather disillusioning conclusion. To universally symmetrize quantum states on the n-fold tensor extension of a given system with Hilbert space dimension d, one asymptotically needs at least a common randomness rate of $\log d$. Since this number marks the trivial upper bound for common randomness rates generated from repetitions of an ideal system of that dimensionality, the common randomness consumption seems exorbitant in some situations.

Since universal symmetrization of communication attacks is a vital ingredient of a broad class of security proofs for quantum key distribution protocols, our findings strongly motivate further research for finding more efficient protocols for symmetrizing quantum states and channels.

6.2. Chapter 3: Simultaneous transmission of classical and quantum information under channel uncertainty

We have developed universal codes for simultaneous transmission of classical information and entanglement under possible jamming attacks by a third malignant party. In the compound channel model, the quantum part of information transmission was done under two important scenarios of entanglement transmission and entanglement generation. The present random codes differ from those used for the perfectly known channel in [Devetak(2005)]. We therefore did not need to approximate our input random codes by an i.i.d state (one with tensor product structure). Also, we evaded BSST type lemmas used in [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] by using basic concavity properties of von Neumann entropy. Future work will hopefully include another important scenario under which quantum information is transmitted, namely subspace transmission. An equivalence statement between the strong subspace transmission and entanglement transmission has been proven in [Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel]. Recently in [Boche et al.(2019d)Boche, Janßen, and Saeedinaeeni], an instance of the present classically enhanced codes was used for universal coding of multiple access quantum channels, where one of the senders shares classical messages with the receiver while the other sends quantum information.

Theorem 32 does not make a positive statement about the structure of $\overline{\mathcal{A}}_{r,CET}$ in the case where $\overline{\mathcal{A}}_{d,CET} = \{(0,0)\}$. In [Boche and Nötzel(2014)], the authors have constructed an example of a channel where the intersection of $\overline{\mathcal{A}}_{r,CET}$ with the x-axis is positive and $\overline{\mathcal{A}}_{d,CET} = \{(0,0)\}$. Future work will consider the structure of the non-zero region in this case along both axes.

The capacity region characterized in Theorem 25 is of a multi-letter nature (requiring a limit over many uses of the channel) but might reduce to a single-letter formula for specific cases of compound channels, which is in itself an interesting question to be considered in future work. Ensuing this question, one might suggest formulas for these capacity regions that offer a more useful characterization. This means that the alternative characterization could entail larger one-shot regions compared to our $\hat{C}(\mathcal{N}_s, p, \Psi)$. An instance of such a characterization in the case of perfectly known quantum channels exists in [Hsieh and Wilde(2010a)] Theorem 5. Therein however, the authors note that their one-shot trapezoids is the same as rectangular regions offered in [Devetak and Shor(2005)], when one considers the union over all the one-shot, one-state regions. The converse statement for compound channels implies that other such characterizations, must also reduce to ours. Reduction to single-letter formulas is nevertheless an important criterion when comparing different characterizations.

Today, in classical systems, secure communication is obtained by applying cryptographic methods upon available reliable- communication schemes. Security of the resulting protocol, that can hence be separated into two protocols (one responsible for reliability and the other for security), relies on assumptions such as non-feasibility of certain tasks or the limited computational capabilities of illegal receivers. For the next generation of classical communication systems, it is expected that different applications (e.g. secure message transmission, broadcasting of common messages and message transmission), are all implemented by physical coding or " physical layer service integration " schemes (see [Schaefer and Boche(2014c)]). For quantum systems that offer a larger variety of services, [Devetak and Shor(2005), Hsieh and Wilde(2010a), Hsieh and Wilde(2010b)] were the first papers in this line of research. The present chapter develops solutions for different models of channel uncertainty that are unavoidable when implementing such integrated services in real-world communication. Following up on the results of [Boche et al.(2019d)Boche, Janßen, and Saeedinaeeni], an interesting direction for future work is towards finding the solution to the arbitrarily varying model for multiple access and broadcast channels as a key step in development of quantum networks.

6.3. Chapter 4: Universal superposition codes: capacity regions for quantum broadcast channel

To construct private codes for the broadcast channel, we first generated suitable random message transmission codes for the broadcast channel without imposing privacy constraints (Lemma 90). This was done by establishing suitable bounds for random universal "superposition codes". With subsequent application of a covering principle, these codes were transformed to fulfill the security criterion in Lemma 89.

As a possible alternative technique to generate such codes, we mention the rather recent "position decoding" and "convex split" techniques [Anshu et al.(2017)Anshu, Devabathini, and Jain,Anshu et al.(2019a)Anshu, Jain, and Warsi]. This approach proved to be powerful yet elegant and was successfully applied to determine "one-shot capacities" or "second order rates" in several scenarios. However, these techniques need still to be further developed, to also be suitable when dealing with channel uncertainties as in the scenarios considered in the present chapter. A partial result in that directions is [Anshu et al.(2019b)Anshu, Jain, and Warsi], where near-optimal universal codes for entanglement assisted message transmission over compound quantum channels with finitely many channel states are constructed. Recently, convex split and position-decoding have been applied in [Wilde et al.(2019)Wilde, Khatri, Kaur, and Guha] to determine the second-order capacity of a cqq compound wiretap channel under the restriction, that the channel state does not vary for the legitimate receiver. For establishing this result, only the "convex split" part has to be universal, while "position- decoding" is applied on a channel with fixed state. As a future research goal, it is desirable to close the gap and establish a fully universal version of these protocol steps.

As mentioned in the introduction, a strong converse cannot be established for the message transmission capacity of the compound cq channel under average error criterion, even when considering $|S| = 2$. When considering a fixed non-vanishing upper bound on the average of decoding error, calculation of capacity for the compound channel is further problematic as there are examples where the optimistic definition of the ϵ-capacity yields a strictly larger number than the one yielded by its pessimistic definition (see [Boche et al.(2018a)Boche, Schaefer, and Poor] Remark 13). This implies that in general there is no second rate capacity theorem possible. The implications of these negative statements are highly interesting in practice, as channel coding in all existing communication systems (such as wireless cellular and WiMax systems), is done given a fixed error probability. It is

therefore important to design channel codes corresponding to ϵ-capacity of the compound channel, that is in general larger than its message transmission capacity.
When considering the one-shot approach ([Anshu et al.(2017)Anshu, Devabathini, and Jain, Anshu et al.(2019a)Anshu, Jain, and Warsi, Salek et al.(2020)Salek, Anshu, Hsieh, Jain, and Fonollosa]) as an alternative to proving capacity results derived here, one must take certain consequences into account. In this approach, one tries to obtain lower and upper-bounds for the ϵ-capacity, and then consider the limit $\epsilon \to 0$ of these bounds. For the compound channel however, the capacity is in general strictly smaller than the ϵ-capacity and hence, it is not clear how these bounds will help, as a lower bound on the ϵ-capacity is not a priori a lower bound on the capacity of the channel. Furthermore, there are some additional highly interesting properties of the ϵ-capacity and the capacity, even when one considers *finite* compound channels ($|S| < \infty$):

- The capacity of the finite compound channel is, as a function of the computable compound channel, a Turing computable function. This is no longer true about infinite compound channels (see [Boche et al.(2020a)Boche, Schaefer, and Poor]).
- The ϵ-capacity of the finite compound channel, as a function of ϵ, is not Banach-Mazur computable, which in turn means that it is not Turing computable either, as the latter condition is a stronger one on computability than former.

These results have of course an impact on the effectiveness of the one-shot approach to achieving capacity results in classical and quantum information theories [Boche et al.(2020a)Boche, Schaefer, and Poor].

A direction for future work given the results derived here, is considering a three dimensional capacity region, establishing a trade-off between the ability of the quantum channel in transmitting common, public and confidential messages under assumptions of the compound channel model. One must pay attention to the operational difference between public messages (belonging to the set $[M_{1,n}]$) and those used for equivocation by Alice (belonging to the set $[L_n]$).
Another direction for future work given the results derived here, is considering the arbitrarily varying quantum channel (AVQC) model for the broadcast channel with confidential messages. Given that in all instances, our error and security requirements, achieve exponential rates of decay, it is perceivable that using the well known robustification and elimination techniques developed in [Ahlswede(1978)], capacity results including dichotomy statements can be made for the AVQC model.

6.4. Chapter 5: Computability aspects

Using channel examples where the ϵ-capacity of the channel exhibits discontinuity as a function of the error input, we were able to show that the function is not in general computable. Further, we showed that in general, there could not be matching computable lower and upper-bound sequences converging to the capacity function, that in turn refutes the existence of a single letter formula and effectiveness of second order rate theorems that intend to achieve the asymptotically optimal rates from those available in the one-shot regime (see [Boche et al.(2020b)Boche, Janßen, and Saeedinaeeni] for more discussions).

The statements proven here have concerned the ϵ-capacity of the channel as a function of the error input. An interesting direction for future work would be to consider this number as a function of the channel input. Our statement of achievability given by Lemma 143, suggests that zero-error codes might play a role in the traditional definition of ϵ-capacity at the discontinuity point of $\epsilon = \frac{1}{2}$. This could be promising as there are examples of compound channels where $C(\mathcal{W}) < \min_{s=1,2} C_0(W_s) < \min_{s=1,2} C_0(W_s)$. Such an example can be readily perceived by appropriate perturbation of the channel given by Example 164. To see this consider the following channel. For $\delta \in [0,1)$, consider the compound channel generated by
$\mathcal{W}_\delta := \{W_{1,\delta}, W_{2,\delta}\} \subset \mathcal{CH}(\mathcal{X}, \{1,2\}; \mathcal{Y}), \mathcal{X} = \mathcal{Y} = \{1,\dots,5\}$,
defined by the following stochastic matrices:

$$W_1 := \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1-\delta & \delta & 0 \\ 0 & 0 & 1-\delta & 0 & \delta \end{pmatrix} \quad W_{2,\delta} := \begin{pmatrix} \delta & 0 & 1-\delta & 0 & 0 \\ 0 & \delta & 1-\delta & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6.1}$$

It is clear that $C_0(W_1) = C_0(W_2)$. We show that for this example, we have $C_0(\mathcal{W}_\delta) > 0$ for $\delta \in [0,1)$, and hence the hypothesis of Lemma 143 is satisfied. For this purpose, consider the $(2,2)$-code (blocklength 2 and 2 messages), with input strings $m_1 := (3,3), m_2 := (1,5)$ and decoder sets $D := \{(3,3)\}$ and D^c respectively. Calculation of error for the $W_{1,\delta}$ yields:

$$e_{m_1,s=1} = \sum_{y \in D^c} W_{1,\delta}^{\otimes 2}(y|(3,3)) = 0, \tag{6.2}$$

$$e_{m_2,s=1} = \sum_{y \in D} W_{1,\delta}^{\otimes 2}(y|(1,5)) = 0. \tag{6.3}$$

Similarly for the second channel we obtain $e_{m_1,s=2} = e_{m_2,s=2} = 0$ and hence we have $C_0(\mathcal{W}_\delta) > 0$.
We observe for $\delta \in (0,1)$.

$$C_0(W_{2,\delta}) = C(W_1,\delta) = \log 3. \tag{6.4}$$

It is also clear that

$$C(W_{2,\delta}) > C(W_s) = \log 3, (s = 1, 2). \tag{6.5}$$

Also, given Theorem 211, we have

$$|C(\mathcal{W}_0) - C(\mathcal{W}_\delta)| \leqslant f(\delta), \tag{6.6}$$

with some $f(\delta)$ such that $f(\delta) \to 0$ as $\delta \to 0$. Let $\min_{s=1,2} C(W_{s,0}) - C(\mathcal{W}_0) = \lambda$. From [Ahlswede(2015)], we know that $\lambda > 0$. Choose δ such that $f(\delta) < \lambda$. From (6.6) we have

$$C(\mathcal{W}_\delta) < \log 3. \tag{6.7}$$

With this choice of $\delta > 0$, we have from (6.4) and (6.5),

$$C(\mathcal{W}_\delta) < \log 3 = \min_{s=1,2} C_0(W_{s,\delta}) < \min_{s=1,2} C(W_{s,\delta}). \tag{6.8}$$

Our capacity results for the case with informed decoder in Section 5.3.3 can be used to derive similar results for the ϵ-capacity of the Broadcast Channel. In this channel model, a sender transmit messages that will be received by two receivers, each in control of the output of a different channel. A specific scenario is when the transmitter wishes to send public messages (decoded by both receivers) and two individual messages. The rates corresponding to the public message transmission are then exactly those achieved in the compound model with two channel states. The negative computability analysis for this specific model would then extend to the general case.

A. Frequency typical sets

In this appendix we give the basic definitions regarding types and frequency typical sets. For a broad as well as concise introduction to the concept of types the reader is referred to [Csiszár and Körner(2011b)], where the bounds stated in this appendix can be found without exception.

Let $\mathcal{X}$ be a finite set, p a probability distribution on $\mathcal{X}$. We define the set of p-typical words in $\mathcal{X}^n$ by

$$T_p^n := \{\mathbf{x} : \ \forall a \in \mathcal{X} : \tfrac{1}{n} N(a|\mathbf{x}) = p(a)\}.$$

If this set in nonempty, we call p a *type of sequences in* $\mathcal{X}^n$ (or *n-type* for short). The concept of types is a powerful tool in classical as well as quantum Shannon theory. In this chapter, use some cardinality bounds on the entities introduced which are stated in the next two lemmas. If we denote, for $n \in \mathbb{N}$ the set of n-types by $\mathfrak{T}(n, \mathcal{X})$, the following statement is true.

Lemma 189 (cf. [Csiszár and Körner(2011b)], Lemma 2.2) *For each $n \in \mathbb{N}$, it holds*

$$|\mathfrak{T}(n, \mathcal{X})| \leqslant (n+1)^{|\mathcal{X}|}.$$

Lemma 190 (cf. [Csiszár and Körner(2011b)], Lemma 2.3) *For each $n \in \mathbb{N}$, and each n-type $\lambda \in \mathfrak{T}(n, \mathcal{X})$, it holds*

$$(n+1)^{-|\mathcal{X}|} \cdot 2^{nH(\lambda)} \ \leqslant \ |T_\lambda^n| \ \leqslant 2^{nH(\lambda)}.$$

Lemma 191 *For each $n \in \mathbb{N}$, there exists a type μ_* of sequences in $\mathcal{X}^n$, such that*

$$H(\mu_*) \geqslant \log |\mathcal{X}| - |\mathcal{X}| \frac{\log(n+1)}{n} \tag{A.1}$$

holds.

Proof 192 *Let μ_* be a type of sequences in $\mathcal{X}^n$, which maximizes the Shannon entropy, i.e.*

$$H(\mu_*) \geqslant H(\lambda)$$

holds for each type λ. *Then, by standard bounds for the frequency typical sets [Csiszár and Körner(2011b)]*

$$|T_\lambda^n| \leqslant 2^{nH(\lambda)} \leqslant 2^{nH(\mu_*)}.$$

holds for each type λ. *Since there are not more than* $(n+1)^{|\mathcal{X}|}$ *different types of sequences in* $\mathcal{X}^n$, *the bound*

$$|\mathcal{X}^n| \leqslant (n+1)^d \cdot 2^{nH(\mu_*)}$$

is valid, which with some rearrangements proves the lemma.

B. Approximation of quantum compound channels using nets

Definition 193 *A τ-net in $\mathcal{C}(\mathcal{H},\mathcal{K})$ is a finite set $\{\mathcal{N}_i\}_{i=1}^T$ with the property that for each $\mathcal{N} \in \mathcal{C}(\mathcal{H},\mathcal{K})$ there is at least one $i \in \{1,...,T\}$ 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 an upper bound on the minimal cardinality of τ-nets.

Lemma 194 *(see e.g. [Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] Lemma 7) For any $\tau \in (0,1]$, there is a τ-net $\{\mathcal{N}_i\}_{i=1}^T$ in $\mathcal{C}(\mathcal{H},\mathcal{K})$ with $T \leqslant (\frac{3}{\tau})^{2(d\cdot d')^2}$, where $d = \dim \mathcal{H}$ and $d' = \dim \mathcal{K}$.*

Given a net in $\mathcal{C}(\mathcal{H},\mathcal{K})$, any compound channel generated by $\mathcal{J} \subset \mathcal{C}(\mathcal{H},\mathcal{K})$ can be approximated by one of its finite subsets. This is the subject of the following lemma.

Lemma 195 *(see e.g Lemma 13 [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel]) Given any compound channel generated by $\mathcal{J} \subset \mathcal{C}(\mathcal{H},\mathcal{K})$, one can construct a finite set $\mathcal{J}_\tau$ with the following properties:*

1. *$\mathcal{J}_\tau \subset \mathcal{J}$,*
2. *$|\mathcal{J}_\tau| \leqslant (\frac{6}{\tau})^{2(d\cdot d')^2}$ with d, d' the dimensions of $\mathcal{H},\mathcal{K}$ respectively and*
3. *for all $\mathcal{N} \in \mathcal{J}, \exists \mathcal{N}' \in \mathcal{J}_\tau$ such that $\| \mathcal{N} - \mathcal{N}' \|_\diamond \leqslant 2\tau$.*

The net approximation can also be performed for arbitrary sets of classical-quantum channels.

Definition 196 *For $\mathcal{W} \subset CQ(\mathcal{X},\mathcal{H})$ and $\tau > 0$, a τ-net is a finite set $\mathcal{W}_\tau := \{W_i\}_{i \in S_\tau} \subset CQ(\mathcal{X},\mathcal{H})$, with property that for every $W \in \mathcal{W}$, there exists and index $i \in [S_n]$ such that*

$$\| W - W_i \|_{CQ} < \tau. \tag{B.1}$$

The existence of such τ-net does not readily guarantee that $\mathcal{W}_\tau \subset \mathcal{W}$. The following lemma gives the existence of a good τ-net contained in the given channel set.

Lemma 197 *(cf. [Bjelaković et al.(2013)Bjelaković, Boche, Janßen, and Nötzel] Lemma 6) Let $\mathcal{W} := \{W_i\}_{i \in S} \subset CQ(\mathcal{X},\mathcal{H})$ and $\tau \in (0, 1/e)$. There exists a set $\mathcal{W}_\tau := \{W_i\}_{i \in S_\tau} \subset \mathcal{W}$ with such that*

1. *$|S_\tau| < (\frac{6}{\tau})^{2|\mathcal{X}|\dim(\mathcal{H})^2}$,*

2. *given any $n \in \mathbb{N}$, for every $i \in S$, there exists $i' \in S_n$ such that*

$$\parallel W_i^{\otimes n}(\mathbf{x}) - W_{i'}^{\otimes n}(\mathbf{x}) \parallel_1 \leqslant 2n\tau, \ (\forall \mathbf{x} \in \mathcal{X}^n). \tag{B.2}$$

C. Auxiliary results

In this section we state some results for reader's convenience.

Lemma 198 *([Yard et al.(2005)Yard, Devetak, and Hayden]) Let $\Psi, \rho, \sigma \in \mathcal{S}(\mathcal{K})$ and let Ψ be pure. Then*

$$F(\Psi, \rho) \geqslant F(\Psi, \sigma) - \frac{1}{2} \parallel \rho - \sigma \parallel_1$$

Lemma 199 *([Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter]) Let L and D be $N \times N$ matrices with non-negative entries which satisfy*

$$L_{jl} \leqslant L_{jj}, L_{jl} \leqslant L_{ll}$$

and

$$D_{jl} \leqslant \max\{D_{jj}, D_{ll}\}$$

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

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

Lemma 200 *([Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] Lemma 3) Let $\rho \in \mathcal{S}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Let, for some other Hilbert space $\mathcal{K}$, $\mathcal{A} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$, $\mathcal{D} \in \mathcal{C}(\mathcal{K}, \mathcal{H})$, $q \in \mathcal{L}(\mathcal{K})$ be an orthogonal projection. If for some $\epsilon > 0$, $F_e(\rho, \mathcal{D} \circ Q \circ \mathcal{A}) \geqslant 1 - \epsilon$ holds, then we have*

$$F_e(\rho, \mathcal{D} \circ \mathcal{A}) \geqslant 1 - 3\epsilon.$$

Lemma 201 *(see e.g. [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] Lemma 5) There is a real number $\bar{c} > 0$ such that for every Hilbert space $\mathcal{H}$, there exist functions $h' : \mathbb{N} \to \mathbb{R}^+$, $\phi : (0, 1/2) \to \mathbb{R}^+$ with $\lim_{l\to\infty} h'(l) = 0$ and $\lim_{\delta\to 0} \phi(\delta) = 0$, such that for $\rho \in \mathcal{S}(\mathcal{H})$, $\delta \in (0, 1/2)$, $l \in \mathbb{N}$, there is an orthogonal projection $q_{\delta,l}$ called the frequency typical projection satisfying*

1. $\mathrm{tr}(\rho^{\otimes l} q_{\delta,l}) \geqslant 1 - 2^{-l(\bar{c}\delta^2 - h'(l))}$
2. $q_{\delta,l} \rho^{\otimes l} q_{\delta,l} \leqslant 2^{-(S(\rho^{\otimes l}) - l\phi(\delta))} q_{\delta,l}$.

Lemma 202 *(see e.g. [Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] Lemma 6) Let $\mathcal{H}$ and $\mathcal{K}$ be finite dimensional Hilbert spaces. There are functions $\gamma : (0, 1/2) \to \mathbb{R}^+$ and $h' : \mathbb{N} \to \mathbb{R}^+$ 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 $\mathcal{G} \subset \mathcal{H}$, there is an operation $\mathcal{N}_{\delta,l} \in \mathcal{C}^{\downarrow}(\mathcal{H}^{\otimes l}, \mathcal{K}^{\otimes l})$, called the reduced operation with respect to $\mathcal{N}$ and $\pi_{\mathcal{G}}$, satisfying

1. *$tr(\mathcal{N}_{\delta,l}(\pi_{\mathcal{G}}^{\otimes l})) \geqslant 1 - 2^{-l(c'\delta^2 - h'(l))}$, with universal constant $c' > 0$.*

2. *$\mathcal{N}_{\delta,l}$ has a Kraus representation with at most $n_{\delta,l} \leqslant 2^{S_e(\pi_{\mathcal{G}}^{\otimes l}, \mathcal{N}^{\otimes l}) + l(\gamma(\delta) + h'(l))}$ Kraus operators.*

3. *For every state $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l})$ and every two channels $\mathcal{M} \in \mathcal{C}^{\downarrow}(\mathcal{H}^{\otimes l}, \mathcal{H}^{\otimes l})$ and $\mathcal{L} \in \mathcal{C}^{\downarrow}(\mathcal{K}^{\otimes l}, \mathcal{H}^{\otimes l})$, the inequality*

$$F_e(\rho, \mathcal{L} \circ \mathcal{N}_{\delta,l} \circ \mathcal{M}) \leqslant F_e(\rho, \mathcal{L} \circ \mathcal{N}^{\otimes l} \circ \mathcal{M})$$

is fulfilled.

Lemma 203 (Gentle measurement) *(see e.g. [Wilde(2017)]) Let $\rho \in \mathcal{S}(\mathcal{H})$ and $0 \leqslant \Lambda \leqslant \mathbb{I}$ with*

$$\mathrm{tr}(\Lambda\rho) \geqslant 1 - \epsilon$$

for some $0 \leqslant \epsilon < 1$. Then for $\rho' := \frac{\sqrt{\Lambda}\rho\sqrt{\Lambda}}{\mathrm{tr}(\Lambda\rho)}$ we have

$$\parallel \rho - \rho' \parallel_1 \leqslant 2\sqrt{\epsilon}.$$

Lemma 204 *([Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] proof of Theorem 3.2 equation (16)) Let $\mathcal{F} \subset \mathcal{G} \subset \mathcal{H}$ with $\dim(\mathcal{F}) = k$ be given. Also let any member of the set $\{\mathcal{N}_1, \ldots, \mathcal{N}_{|S|}\} \subset \mathcal{C}^{\downarrow}(\mathcal{H}, \mathcal{K})$ have a Kraus representation with n_j operators for $j \in \{1, \ldots, S\}$ and set*

$$\overline{\mathcal{N}} := \frac{1}{|S|} \sum_{j=1}^{|S|} \mathcal{N}_j.$$

Then there exists a recovery operation $\mathcal{R} \in \mathcal{C}(\mathcal{K}, \mathcal{H})$ such that

$$F_e(\pi_{\mathcal{F}}, \mathcal{R} \circ \overline{\mathcal{N}}) \geqslant w - \parallel D(p) \parallel_1, \tag{C.1}$$

where $w := tr(\overline{\mathcal{N}}(\pi_{\mathcal{F}}))$, $p := k\pi_{\mathcal{F}}$ and

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

In the above

$$D_{(ij)(rl)}(p) := \frac{1}{k}(pa_{j,i}a_{l,r}^{\dagger}p - \frac{1}{k}tr(pa_{j,i}^{\dagger}a_{l,r}p)p).$$

where $\{|f_1\rangle, ..., |f_{|S|}\rangle\}$ and $\{|e_1\rangle, ..., |e_{n|S|}\rangle\}$ are arbitrary orthonormal bases for $\mathbb{C}^{|S|}$ and $\mathbb{C}^{n|S|}$, and where $\{a_{j,i}\}_{i=1}^{n_j}$ is the set of Kraus operators for $\mathcal{N}_j$.

Lemma 205 *(see [Ahlswede(1978)]) If a function $f : S^l \to [0,1]$, satisfies*

$$\sum_{s^l \in S^l} f(s^l) q^l(s^l) \geqslant 1 - \gamma \tag{C.2}$$

with $q^l(s^l) := \prod_{i=1}^{l} q(s_i)$, for all $q \in \mathcal{T}(l,S)$ and some $\gamma \in [0,1]$, then

$$\frac{1}{l!} \sum_{\sigma \in \mathfrak{S}_l} f(\sigma(s^l)) \geqslant 1 - (l+1)^{|S|} \cdot \gamma \quad \forall s^l \in S^l. \tag{C.3}$$

Lemma 206 *(see [Ahlswede(1978)]) Let $K \in \mathbb{N}$ and numbers $a_1, \ldots, a_K, b_1, \ldots, b_K \in [0,1]$ be given. Assume that*

$$\frac{1}{K} \sum_{i=1}^{K} a_i \geqslant 1 - \epsilon$$

and

$$\frac{1}{K} \sum_{i=1}^{K} b_i \geqslant 1 - \epsilon$$

hold. Then

$$\frac{1}{K} \sum_{i=1}^{K} a_i b_i \geqslant 1 - 2\epsilon. \tag{C.4}$$

Lemma 207 *Let $\{W_s : \mathcal{X} \to \mathcal{S}(\mathcal{H})\}_{s \in S}$ be a set of cq channels and let $p \in \mathcal{P}(\mathcal{X})$. Then*

$$\lim_{\alpha \to 1} \inf_{s \in S} \chi_\alpha(W_s, p) = \inf_{s \in S} \chi(p, W_s).$$

Lemma 208 *Let $p, q \in \mathcal{L}(\mathcal{H})$, $0 \leqslant p, q \leqslant \mathbb{1}_{\mathcal{H}}$ and $\tau \in \mathcal{S}(\mathcal{H})$. It holds*

$$\operatorname{tr}(\tau p q p) \geqslant \operatorname{tr}(\tau q) - 2\sqrt{\operatorname{tr}(\tau(\mathbb{1} - p))}.$$

Lemma 209 (cf. [Audenaert(2007)]) *For any two states ρ and σ on Hilbert space $\mathcal{H}$, let $\delta = \parallel \rho - \sigma \parallel_1$ and $\dim(\mathcal{H}) = d$. Then*

$$|S(\rho) - S(\sigma)| \leqslant \delta \log(d-1) + h(\delta) \tag{C.5}$$

hold, with $h(x) = -(1-x)\log(1-x) - x \log x$, for $x \in (0,1]$ and $h(0) = 0$, the binary entropy.

Lemma 210 *[cf. [Shirokov(2017)], Corollary 2] Let $\rho, \sigma \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, with $\parallel \rho - \sigma \parallel_1 = \delta$ and $\dim(\mathcal{H}_B) = d$. It holds*

$$|I(A;B|C,\rho) - I(A;B|C,\sigma)| \leqslant 2\left(\delta \log d + (1+\delta) h(\frac{\delta}{1+\delta})\right),$$

wit h, the binary entropy, as defined in the previous lemma.

We also state a continuity theorem for the compound capacity of the classical channel.

Heuristically, this theorem states, that when two compound channels are close to each other in terms of their distance defined by (5.16), then their compound capacities are also close to each other.

Theorem 211 ([A. Grigorescu and Poor(2015)]) *Let $\mu \in (0,1)$ and $\mathcal{W}, \tilde{\mathcal{W}} \in \mathcal{CC}(\mathcal{X}, S; \mathcal{Y})$ with $D(\mathcal{W}, \tilde{\mathcal{W}}) \leqslant \mu$. It holds $|C(\mathcal{W}) - C(\tilde{\mathcal{W}})| \leqslant f(\mu, |\mathcal{Y}|)$, with*

$$f(\mu, |\mathcal{Y}|) := 12h_2(\mu) + 8\mu \log |\mathcal{Y}|,$$

where $h_2(x)$ is the binary entropy of x.

D. Universal classical-quantum superposition coding

In this appendix, we establish a random coding construction of superposition codes for classical-quantum channels which are a major ingredient for the achievability proofs in Section 4.3. In particular a detailed proof of Lemma 84 is provided.

Over the years several code constructions for message transmission over compound cq channels have been established (see [Bjelakovic and Boche(2007), Hayashi(2009), Datta and Hsieh(2010), Mosonyi(2015)]). The arguments we invoke below for proving Lemma 84 rely heavily on the techniques Mosonyi's work [Mosonyi(2015)]. Therein properties of the quantum Renyi Divergences and the closely related "sandwich Renyi divergences" are used to derive universal random coding results for classical-quantum channels. Below we further elaborate on that approach and extend it by suitable superposition codes.

To facilitate connecting the discussion below with the arguments in [Mosonyi(2015)] we introduce some notation from there. For a probability distribution $p \in \mathcal{P}(\mathcal{Y})$ and a cq channel $W : \mathcal{Y} \to \mathcal{S}(\mathcal{H})$, we define quantum states

$$W(p) := \sum_{y \in \mathcal{X}} p(y) \cdot W(y), \qquad \mathbb{W}(p) := \sum_{y \in \mathcal{X}} p(y) \left|y\right\rangle\left\langle y\right| \otimes W(y), \text{and} \qquad \hat{p} := \sum_{y \in \mathcal{X}} p(y) \left|y\right\rangle\left\langle y\right| .$$

For each pair of non-zero positive semi-definite operators ρ, σ and every $\alpha \in (0,1)$ we define

$$Q_\alpha(\rho||\sigma) := \mathrm{tr}(\rho^\alpha \sigma^{1-\alpha}), \qquad \text{and} \qquad D_\alpha(\rho||\sigma) := \frac{1}{1-\alpha} \log \mathrm{tr}(\rho^\alpha \sigma^{1-\alpha})$$

from

$$\chi_\alpha(p, W) \; := \; \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D_\alpha(\mathbb{W}(p)||\hat{p} \otimes \sigma) \tag{D.1}$$

derives. It is known, that the limit $\alpha \to 1$ of the above quantity exists and equals the Holevo quantity $\chi(p, W)$. Translating to the notation in the statement of Lemma 8, we notice, that $\chi(p, W) \; = \; I(Y; B)$ holds.

Lemma 212 *Let $\mathcal{W}$ be a set of cq channels each mapping $\mathcal{Y}$ to $\mathcal{S}(\mathcal{K})$, q a probability* *distribution on $\mathcal{X}$ and r_x a probability distribution on $\mathcal{Y}$ for each $x \in \mathcal{X}$. It holds*

$$\lim_{\alpha \mathbb{N}earrow 1} \inf_{V \in \mathcal{W}} \sum_{x \in \mathcal{X}} q(x) \cdot \chi_\alpha(r_x, V) \;=\; \inf_{V \in \mathcal{W}} \sum_{x \in \mathcal{X}} q(x) \cdot \chi(r_x, V)$$

The above statement slightly generalizes that of Lemma 3.13 in [Mosonyi(2015)] (regarding the limit from below). The proof is by a similar argument. We include a proof for the readers convenience.

Proof 213 *Set $f(\alpha, V) := \sum_{x \in \mathcal{X}} q(x) \cdot \chi_\alpha(r_x, V)$ for each $\alpha \in (0,1)$ and cq channel V. It holds*

$$\begin{aligned}
\lim_{\alpha \mathbb{N}earrow 1} \inf_{V \in \mathcal{W}} f(\alpha, V) &\overset{(a)}{=} \lim_{\alpha \mathbb{N}earrow 1} \min_{V \in \overline{\mathcal{W}}} f(\alpha, V) \\
&\overset{(b)}{=} \sup_{\alpha \in (0,1)} \min_{V \in \overline{\mathcal{W}}} f(\alpha, V) \\
&\overset{(c)}{=} \min_{V \in \overline{\mathcal{W}}} \sup_{\alpha \in (0,1)} f(\alpha, V) \\
&\overset{(d)}{=} \min_{V \in \overline{\mathcal{W}}} \lim_{\alpha \mathbb{N}earrow 1} f(\alpha, V) \\
&= \inf_{V \in \mathcal{W}} \sum_{x \in \mathcal{X}} q(x) \cdot \chi(V, r_x)
\end{aligned}$$

The equality in (a) holds by continuity of $f(\alpha, \cdot)$ for each $\alpha \in (0,1)$ (the closure might be performed in any norm, for example the $\|\cdot\|_{CQ}$), (b) is justified, because the argument of the limit is monotonously increasing on (0,1). The min-max exchange in (c) is an application of Lemma 2.3 from [Mosonyi(2015)], (d) by monotonicity of f in α, and in (e) the limit $\alpha \mathbb{N}earrow 1$ is performed according to Lemma B.3 in [Mosonyi and Hiai(2011)].

The starting point for our proof of Lemma 84 is the generic random coding bound from Hayashi and Nagaoka [Hayashi and Nagaoka(2003)] we state below.

Lemma 214 ([Hayashi and Nagaoka(2003)], cf. [Mosonyi(2015)], Lemma 4.15) *Let $V : \mathcal{Y} \to \mathcal{S}(\mathcal{K})$ be a cq channel, $M \in \mathbb{N}$ and $p \in \mathcal{P}(\mathcal{Y})$. There exists a map $(y_1, \dots, y_M) \mapsto (\Lambda_1(y), \dots, \Lambda_M(y))$, such that $(\Lambda_m(y))_{m \in [M]} \subset \mathcal{L}(\mathcal{K})$ is a POVM, and, given $Y^M := (Y_1, \dots, Y_M)$ of independent random variables each with distribution p, for each $\forall \alpha \in (0,1)$, the bound*

$$\mathbb{E}_{Y^M}\Big[\frac{1}{M} \sum_{m \in [M]} \mathrm{tr}(W(Y_m)\Lambda_m(Y^M)^c)\Big] \;\leqslant\; 8 \cdot M^{1-\alpha} \cdot Q_\alpha\big(\mathbb{W}(p) || \hat{p} \otimes W(q)\big)$$

holds.

Proof 215 (Proof of Lemma 84) *Fix $n \in \mathbb{N}$ and an n-word $\mathbf{x} \in T^n_{q,\delta}$ which we assume to be of type λ (i.e. $\mathbf{x} \in T^n_\lambda$). We approximate $\{W_s\}_{s \in S}$ by a finite τ_n-net $\{W_s\}_{s \in S_n} \subset \{W_s\}_{s \in S}$ with $\tau_n := 2^{-\frac{n\mathbb{N}u}{2}}$ with a constant positive number $\mathbb{N}u$ to be determined later. We choose the net small enough to fulfill the cardinality bound* $\log|S_n| \;\leqslant\; 2 \cdot |\mathcal{X}| \cdot d^2(\log 6 +$

$n\mathbb{N}u/2$) which is possible by Lemma 197. We introduce abbreviations $d := \dim \mathcal{K}_B$, $r_{\mathbf{x}}(\cdot) := r^{\otimes n}(\cdot|\mathbf{x})$ and $r'_{\mathbf{x}}(\cdot) := r'_{n,\delta}(\cdot|\mathbf{x})$ for each $\mathbf{x} \in \mathcal{X}^n$. Applying Lemma 214 on the cq channel $\overline{W}_n := \frac{1}{|S_n|}\sum_{s\in S_n} W_s^{\otimes n}$ with $p := r'_{\mathbf{x}}$, and

$$M := \lfloor \exp(n(\inf_{s\in S} I(Y;B|X,\sigma_s) - \delta \cdot |\mathcal{X}| \log d - \mathbb{N}u))\rfloor, \tag{D.2}$$

we know that choosing a codewords $Y_1, \dots, Y_M$ i.i.d. according to $r'_{\mathbf{x}}$ each, allows us to bound the expectation by

$$\mathbb{E}_{Y^M}\left[\frac{1}{M}\sum_{m\in[M]} \operatorname{tr}(\overline{W}_n(Y_m)\Lambda_m(Y^M)^c)\right] \leqslant$$
$$8 \cdot M^{1-\alpha} \cdot Q_\alpha\left(\frac{1}{|S_n|}\sum_{s\in S_n} \mathbb{W}_s^{\otimes n}(r'_{\mathbf{x}})||\hat{r'}_{\mathbf{x}} \otimes \overline{W}_n(r'_{\mathbf{x}})\right) \tag{D.3}$$

for each $\alpha \in (0,1)$. By linearity of the trace and the expectation, the above inequality implies

$$\mathbb{E}_{Y^M}\left[\min_{s\in S_n}\frac{1}{M}\sum_{m\in[M]} \operatorname{tr}(W_s^{\otimes n}(Y_m)\Lambda_m(Y^M)^c)\right] \leqslant$$
$$8 \cdot |S_n| \cdot M^{1-\alpha} \cdot Q_\alpha\left(\frac{1}{|S_n|}\sum_{s\in S_n} \mathbb{W}_s^{\otimes n}(r'_{\mathbf{x}})||\hat{r'}_{\mathbf{x}} \otimes \overline{W}_n(r'_{\mathbf{x}})\right). \tag{D.4}$$

The left hand side of the above inequality can be identified as the expected average error of a random code. We proceed to further bound the Function Q_α on the right hand side. We have

$$Q_\alpha\left(\frac{1}{|S_n|}\sum_{s\in S_n} \mathbb{W}_s^{\otimes n}(r'_{\mathbf{x}})||\hat{r'}_{\mathbf{x}} \otimes \overline{W_n}(r'_{\mathbf{x}})\right) \leqslant \frac{1}{r_{\mathbf{x}}(T_{r,\delta}(\mathbf{x}))^{2-\alpha}} Q_\alpha\left(\frac{1}{|S_n|}\sum_{s\in S_n} \mathbb{W}_s^{\otimes n}(r_{\mathbf{x}})||\hat{r}_{\mathbf{x}} \otimes \overline{W_n}(r_{\mathbf{x}})\right). \tag{D.5}$$

In (D.5) we have used definition of the pruned distribution, and observed operator monotonicity of the function $f(x) = x^\alpha$ for $\alpha \in [0,1]$ (cf. [Bhatia(1996)], Theorem 5.1.9). Following the arguments in proof of Lemma 4.16 [Mosonyi(2015)] we obtain

$$Q_\alpha\big(\frac{1}{|S_n|}\sum_{s\in S_n} \mathbb{W}_s^{\otimes n}(r_{\mathbf{x}})||\hat{r}_{\mathbf{x}} \otimes \overline{W_n}(q^{\otimes n})\big) \leqslant \frac{1}{|S_n|^\alpha}\sum_{s\in S_n} \exp\big((\alpha-1)\cdot\alpha\cdot\chi_\alpha(r_{\mathbf{x}}, W_s^{\otimes n})\big)\cdot d^{n(\alpha-1)^2}$$
$$\leqslant \exp\left((\alpha-1)\cdot\alpha\cdot\min_{s\in S_n}\chi_\alpha(r_{\mathbf{x}}, W_s^{\otimes n}) + n(\alpha-1)^2\log d + \log|S_n|\right) \tag{D.6}$$

In order to further estimate the error exponent above, we note that for each $s \in S$

$$\chi_\alpha(W_s^{\otimes n}, r_{\mathbf{x}}) \;=\; \sum_{x\in\mathcal{X}} \lambda(x)\cdot\chi_\alpha(W_s, r(\cdot|x)) \;\geqslant\; \sum_{x\in\mathcal{X}} q(x)\cdot\chi_\alpha(W_s, r(\cdot|x)) - \delta\cdot|\mathcal{X}|\log d.$$

In the above, we have used $|\lambda(x) - q(x)| \leqslant \delta$ *and* $\chi_\alpha(W_s, r(\cdot|x)) \leqslant \log d$. *By Lemma 212, choosing* α *close enough to one allows us to bound*

$$\alpha \inf_{s\in S}\sum_{x\in\mathcal{X}} q(x)\cdot\chi_\alpha(r(\cdot|x), W_s) \;\geqslant\; \inf_{s\in S}\sum_{x\in\mathcal{X}} q(x)\cdot\chi(r(\cdot|x), W_s) - \delta\cdot|\mathcal{X}|\log d \tag{D.7}$$

$$= \inf_{s\in S}\; I(Y;B|X,\sigma_s) - \delta\cdot|\mathcal{X}|\log d. \tag{D.8}$$

where we introduced the notation from the statement of Lemma 8 in the second line. Note, that with our choice of M*, we have*

$$\alpha \inf_{s\in S}\sum_{x\in\mathcal{X}} q(x)\cdot\chi_\alpha(r(\cdot|x), W_s) - \frac{1}{n}\log M \;\geqslant\; \mathbb{N}u > 0 \tag{D.9}$$

Combining the estimates from (D.4) - (D.9) and subsequent upper-bounding the right hand side of (D.3), we achieve the bound

$$\mathbb{E}_{Y^M}\left[\min_{s\in S_n}\frac{1}{M}\sum_{m\in[M]} \operatorname{tr}(W_s^{\otimes n}(Y_m)\Lambda_m(Y^M)^c)\right]$$
$$\leqslant 16\cdot\exp\left((\alpha-1)\cdot n(\mathbb{N}u + (\alpha-1)\log d + 2|\mathcal{X}|d^2[\frac{\log 6}{n} + \frac{\mathbb{N}u}{2}])\right) \;\leqslant\; 2^{-n\mathbb{N}u/4}$$

Where the last inequality above holds for a fixed choice of α *close enough to one and large enough* n. *By the property of the* τ_n *net and linearity of trace and expectation, we can conclude*

$$\mathbb{E}_{Y^M}\left[\inf_{s\in S}\frac{1}{M}\sum_{m\in[M]} \operatorname{tr}(W_s^{\otimes n}(Y_m)\Lambda_m(Y^M)^c)\right] \;\leqslant\; 2^{-n\mathbb{N}u/4} + n\cdot 2^{-n\mathbb{N}u}. \tag{D.10}$$

We are done.

Bibliography

[Grover(1996)] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *STOC '96*, 1996.

[Bennett et al.(1993)Bennett, Brassard, Crépeau, Jozsa, Peres, and Wootters] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Physical Review Letters*, vol. 70, no. 13, p. 1895, 1993.

[Wilde(2017)] M. W. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge, 2017.

[Holevo(1998)] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 269–273, 1998.

[Schumacher and Westmoreland(1997)] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, vol. 56, no. 1, p. 131, 1997.

[Devetak(2005)] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 44–55, 2005.

[Cai et al.(2004)Cai, Winter, and Yeung] N. Cai, A. Winter, and R. W. Yeung, "Quantum privacy and quantum wiretap channels," *Problems of Information Transmission*, vol. 40, no. 4, pp. 318–336, 2004.

[Devetak and Shor(2005)] I. Devetak and P. W. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information," *Communications in Mathematical Physics*, vol. 256, no. 2, pp. 287–303, 2005.

[Bennett et al.(1999)Bennett, Shor, Smolin, and Thapliyal] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted classical capacity of noisy quantum channels," *Physical Review Letters*, vol. 83, no. 15, p. 3081–3084, Oct. 1999. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.83.3081

[Hsieh and Wilde(2010a)] M.-H. Hsieh and M. M. Wilde, "Entanglement-assisted communication of classical and quantum information," *IEEE Transactions on*

Information Theory, vol. 56, no. 9, p. 4682–4704, Sep. 2010. [Online]. Available: http://dx.doi.org/10.1109/TIT.2010.2053903

[Hsieh and Wilde(2010b)] ——, "Trading classical communication, quantum communication, and entanglement in quantum shannon theory," *IEEE Transactions on Information Theory*, vol. 56, no. 9, p. 4705–4730, Sep. 2010. [Online]. Available: http://dx.doi.org/10.1109/TIT.2010.2054532

[Jones et al.(2012)Jones, Van Meter, Fowler, McMahon, Kim, Ladd, and Yamamoto] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, "Layered architecture for quantum computing," *Physical Review X*, vol. 2, no. 3, p. 031007, 2012.

[Wilde et al.(2012)Wilde, Hayden, and Guha] M. M. Wilde, P. Hayden, and S. Guha, "Quantum trade-off coding for bosonic communication," *Physical Review A*, vol. 86, no. 6, p. 062306, 2012.

[Schaefer and Boche(2014a)] R. F. Schaefer and H. Boche, "Physical layer service integration in wireless networks : Signal processing challenges," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 147–156, 2014.

[Liang et al.(2009)Liang, Poor, and Shamai] Y. Liang, H. V. Poor, and S. Shamai, *Information theoretic security.* Now Publishers Inc, 2009.

[Jorswieck et al.(2010)Jorswieck, Wolf, and Gerbracht] E. A. Jorswieck, A. Wolf, and S. Gerbracht, "Secrecy on the physical layer in wireless networks," *Trends in Telecommunications Technologies*, 2010.

[Liu and Trappe(2010)] R. Liu and W. Trappe, *Securing wireless communications at the physical layer.* Springer, 2010, vol. 7.

[Bloch and Barros(2011)] M. Bloch and J. Barros, *Physical-layer security: from information theory to security engineering.* Cambridge University Press, 2011.

[Wyner(1975)] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[Csiszár and Körner(1978)] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[Dynes et al.(2016)Dynes, Tam, Plews, Fröhlich, Sharpe, Lucamarini, et al.] J. F. Dynes, W. W. Tam, A. Plews, B. Fröhlich, A. W. Sharpe, Lucamarini *et al.*, "Ultra-high bandwidth quantum secured data transmission," *Scientific reports*, vol. 6, no. 1, pp. 1–6, 2016.

[Jacak et al.(2016)Jacak, Melniczuk, Jacak, Janutka, Jóźwiak, Gruber, and Jóźwiak] M. Jacak, D. Melniczuk, J. Jacak, A. Janutka, I. Jóźwiak, J. Gruber, and P. Jóźwiak, "Quantum key distribution security constraints caused by controlled quality of dark channel for non-entangled and entangled photon quantum cryptography setups," *Optical and Quantum Electronics*, vol. 48, no. 7, pp. 1–16, 2016.

[Boche et al.(2019a)Boche, Janßen, and Saeedinaeeni] H. Boche, G. Janßen, and S. Saeedinaeeni, "Universal random codes: Capacity regions of the compound quantum multiple-access channel with one classical and one quantum sender," *Quantum Information Processing*, vol. 18, no. 8, pp. 1–27, 2019.

[H. Boche(2020a)] H. V. P. H. Boche, R. F. Schaefer, "Denial-of-service attacks on communication systems: Detectability and jammer knowledge," *IEEE Transactions on Signal Processing*, vol. 68, May 2020.

[Ahlswede et al.(2013)Ahlswede, Bjelakovic, Boche, and Nötzel] R. Ahlswede, I. Bjelakovic, H. Boche, and J. Nötzel, "Quantum capacity under adversarial quantum noise: Arbitrarily varying quantum channels," *Communications in Mathematical Physics*, vol. 317, no. 1, pp. 103–156, Jan. 2013.

[H. Boche(2014)] J. N. H. Boche, "Positivity, discontinuity, finite resources, nonzero error for arbitrarily varying quantum channels," in *IEEE International Symposium on Information Theory*, Jul. 2014, pp. 541–545.

[H. Boche(2019)] N. C. H. Boche, M. Cai, "Message transmission over classical quantum channels with a jammer with side information, correlation as resource and common randomness generating," in *2019 IEEE International Symposium on Information Theory*. IEEE, Jul. 2019.

[H. Boche(2020b)] ——, "Message transmission over classical quantum channels with a jammer with side information: Correlation as resource, common randomness generation," *Journal of Mathematical Physics*, vol. 61, p. 062201 Online, 2020.

[Boche et al.(2019b)Boche, Janßen, and Saeedinaeeni] H. Boche, G. Janßen, and S. Saeedinaeeni, "Simultaneous transmission of classical and quantum information under channel uncertainty and jamming attacks," *Journal of Mathematical Physics*, vol. 60, no. 2, p. 022204, Feb. 2019. [Online]. Available: http://dx.doi.org/10.1063/1.5078430

[Ahlswede(1978)] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 44, no. 2, 1978.

[Avigad and Brattka(2014)] J. Avigad and V. Brattka, “Computability and analysis: The legacy of Alan Turing,” in *Turing’s Legacy: Developments from Turing’s Ideas in Logic*, R. Downey, Ed. Cambridge, UK: Cambridge University Press, 2014.

[Gödel(1930)] K. Gödel, “Die Vollständigkeit der Axiome des logischen Funktionenkalküls,” *Monatshefte für Mathematik*, vol. 37, no. 1, pp. 349–360, 1930.

[Gödel(1934)] ——, “On undecidable propositions of formal mathematical systems,” *Notes by Stephen C. Kleene and Barkely Rosser on Lectures at the Institute for Advanced Study, Princeton, NJ*, 1934.

[Kleene(1952)] S. C. Kleene, *Introduction to Metamathematics.* Van Nostrand, New York: Wolters-Noordhoffv, 1952.

[Minsky(1961)] M. Minsky, “Recursive unsolvability of Post’s problem of ’tag’ and other topics in theory of Turing machines,” *Ann. Math.*, vol. 74, no. 3, pp. 437–455, 1961.

[Boche et al.(2019c)Boche, Janßen, and Saeedinaeeni] H. Boche, G. Janßen, and S. Saeedinaeeni, “Universal superposition codes: capacity regions of compound quantum broadcast channel with confidential messages,” 2019.

[Ahlswede(1967a)] R. Ahlswede, “Certain results in coding theory for compound channels,” in *Proceedings of the Colloquium on Information Theory*, vol. 1, 1967, pp. 35–60.

[Ahlswede and Wolfowitz(1969)] R. Ahlswede and J. Wolfowitz, “The structure of capacity functions for compound channels,” in *Probability and Information Theory.* Springer, 1969, pp. 12–54.

[Bjelaković et al.(2013)Bjelaković, Boche, Janßen, and Nötzel] I. Bjelaković, H. Boche, G. Janßen, and J. Nötzel, “Arbitrarily varying and compound classical-quantum channels and a note on quantum zero-error capacities,” in *Information Theory, Combinatorics, and Search Theory.* Springer, 2013, pp. 247–283.

[Boche et al.(2018a)Boche, Schaefer, and Poor] H. Boche, R. F. Schaefer, and H. V. Poor, “Analytical properties of shannon’s capacity of arbitrarily varying channels under list decoding: Super-additivity and discontinuity behavior,” *Problems of Information Transmission*, vol. 54, no. 3, pp. 199–228, 2018.

[Boche et al.(2022)Boche, S. Saeedinaeeni, and Poor] H. Boche, R. S. S. Saeedinaeeni, and V. Poor, “On the algorithmic computability of achievability and converse:ϵ-capacity of compound channels and asymptotic bounds of error-correcting codes,” *Unpublished*, 2022.

[Holevo(2012)] A. S. Holevo, *Quantum systems, channels, information: a mathematical introduction.* Walter de Gruyter, 2012, vol. 16.

[Ludwig(1983)] G. Ludwig, *The Problem: An Axiomatic Basis for Quantum Mechanics.* Berlin, Heidelberg: Springer Berlin Heidelberg, 1983, pp. 1–11. [Online]. Available: https://doi.org/10.1007/978-3-642-86751-4_1

[Bose(1924)] Bose, "Plancks gesetz und lichtquantenhypothese," *Zeitschrift für Physik*, vol. 26, no. 1, pp. 178–181, Dec 1924. [Online]. Available: https://doi.org/10.1007/BF01327326

[Compton and Heisenberg(1984)] A. H. Compton and W. Heisenberg, *The Physical Principles of the Quantum Theory.* Berlin, Heidelberg: Springer Berlin Heidelberg, 1984, pp. 117–166. [Online]. Available: https://doi.org/10.1007/978-3-642-61742-3_10

[Heinosaari et al.(2016)Heinosaari, Miyadera, and Ziman] T. Heinosaari, T. Miyadera, and M. Ziman, "An invitation to quantum incompatibility," *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 12, p. 123001, feb 2016. [Online]. Available: https://doi.org/10.1088/1751-8113/49/12/123001

[Filippov et al.(2017)Filippov, Heinosaari, and Leppäjärvi] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, "Necessary condition for incompatibility of observables in general probabilistic theories," *Physical Review A*, vol. 95, no. 3, p. 032127, 2017.

[Plávala(2016)] M. Plávala, "All measurements in a probabilistic theory are compatible if and only if the state space is a simplex," *Phys. Rev. A*, vol. 94, p. 042108, Oct 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.94.042108

[Werner(2001)] R. F. Werner, *Quantum Information Theory — an Invitation.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 14–57. [Online]. Available: https://doi.org/10.1007/3-540-44678-8_2

[Wold(2012)] M. Wold, "Quantum channels and operators: Guided tour," *Unpublished lecture notes*, 2012.

[Wolf et al.(2009)Wolf, Perez-Garcia, and Fernandez] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, "Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory," *Phys. Rev. Lett.*, vol. 103, p. 230402, Dec 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.103.230402

[Holevo(1973)] A. Holevo, "Statistical decision theory for quantum systems," *Journal of Multivariate Analysis*, vol. 3, pp. 337–394, 1973.

[Hsieh and Wilde(2009)] M.-H. Hsieh and M. M. Wilde, “Public and private communication with a quantum channel and a secret key,” *Physical Review A*, vol. 80, no. 2, Aug. 2009. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.80.022306

[Kitaev et al.(2002)Kitaev, Shen, and Vyalyi] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and quantum computation*, ser. Graduate Studies in Mathematics. American Mathematical Soc., 2002, no. 47.

[Christandl et al.(2009)Christandl, König, and Renner] M. Christandl, R. König, and R. Renner, “Postselection technique for quantum channels with applications to quantum cryptography,” *Physical Review Letters*, vol. 102, no. 2, Jan. 2009. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.102.020504

[Gross et al.(2007)Gross, Audenaert, and Eisert] D. Gross, K. Audenaert, and J. Eisert, “Evenly distributed unitaries: On the structure of unitary designs,” *Journal of mathematical physics*, vol. 48, no. 5, p. 052104, 2007.

[Wakakuwa(2017)] E. Wakakuwa, “Symmetrizing cost of quantum states,” *Phys. Rev. A*, vol. 95, no. 032328, 2017.

[Barvinok(2003)] A. Barvinok, *A course in convexity, Graduate studies in mathematics.* merican Mathematical Society, 2003.

[Nielsen and Chuang(2010)] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, 2010.

[Audenaert(2007)] K. Audenaert, “A sharp continuity estimate for the von neumann entropy,” *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 28, 2007. [Online]. Available: https://doi.org/10.1088/1751-8113/40/28/S18

[Renner(2005)] R. Renner, “Security of quantum key distribution,” Ph.D. dissertation, ETH Zurich, 9 2005, available at http://arxiv.org/abs/quant-ph/0512258.

[Bjelaković et al.(2009)Bjelaković, Boche, and Nötzel] I. Bjelaković, H. Boche, and J. Nötzel, “Entanglement transmission and generation under channel uncertainty: Universal quantum channel coding,” *Communications in Mathematical Physics*, vol. 292, no. 1, p. 55–97, Aug. 2009. [Online]. Available: http://dx.doi.org/10.1007/s00220-009-0887-0

[Klesse(2007)] R. Klesse, “Approximate quantum error correction, random codes, and quantum channel capacity,” *Physical Review A*, vol. 75, no. 6, Jun. 2007. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.75.062315

[Mosonyi(2015)] M. Mosonyi, “Coding theorems for compound problems via quantum Rényi divergences,” *IEEE Transactions on Information Theory*, vol. 61, no. 6, p. 2997–3012, Jun. 2015. [Online]. Available: http://dx.doi.org/10.1109/TIT.2015.2417877

[Ahlswede et al.(2012)Ahlswede, Bjelaković, Boche, and Nötzel] R. Ahlswede, I. Bjelaković, H. Boche, and J. Nötzel, “Quantum capacity under adversarial quantum noise: Arbitrarily varying quantum channels,” *Communications in Mathematical Physics*, vol. 317, no. 1, p. 103–156, Nov. 2012. [Online]. Available: http://dx.doi.org/10.1007/s00220-012-1613-x

[Boche and Nötzel(2014)] H. Boche and J. Nötzel, “Positivity, discontinuity, finite resources, and nonzero error for arbitrarily varying quantum channels,” *Journal of Mathematical Physics*, vol. 55, no. 12, p. 122201, Dec. 2014. [Online]. Available: http://dx.doi.org/10.1063/1.4902930

[Wilde and Hsieh(2011a)] M. M. Wilde and M.-H. Hsieh, “The quantum dynamic capacity formula of a quantum channel,” *Quantum Information Processing*, vol. 11, no. 6, p. 1431–1463, Sep. 2011. [Online]. Available: http://dx.doi.org/10.1007/s11128-011-0310-6

[Boche et al.(2018b)Boche, Deppe, Nötzel, and Winter] H. Boche, C. Deppe, J. Nötzel, and A. Winter, “Fully quantum arbitrarily varying channels: Random coding capacity and capacity dichotomy,” *2018 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2018. [Online]. Available: http://dx.doi.org/10.1109/ISIT.2018.8437610

[Ericson(1985)] T. Ericson, “Exponential error bounds for random codes in the arbitrarily varying channel,” *IEEE Transactions on Information Theory*, vol. 31, no. 1, pp. 42–48, 1985.

[Karumanchi et al.(2016)Karumanchi, Mancini, Winter, and Yang] S. Karumanchi, S. Mancini, A. Winter, and D. Yang, “Quantum channel capacities with passive environment assistance,” *IEEE Transactions on Information Theory*, vol. 62, no. 4, p. 1733–1747, Apr. 2016. [Online]. Available: http://dx.doi.org/10.1109/TIT.2016.2522192

[Ahlswede and Winter(2002)] R. Ahlswede and A. Winter, “Strong converse for identification via quantum channels,” *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 569–579, 2002.

[Schaefer and Boche(2014b)] R. F. Schaefer and H. Boche, “Robust broadcasting of common and confidential messages over compound channels: Strong secrecy and decoding performance," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 10, pp. 1720–1732, 2014.

[Boche et al.(2014)Boche, Cai, Cai, and Deppe] H. Boche, M. Cai, N. Cai, and C. Deppe, "Secrecy capacities of compound quantum wiretap channels and applications," *Physical Review A*, vol. 89, no. 5, May 2014. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.89.052320

[Wilde and Hsieh(2011b)] M. M. Wilde and M.-H. Hsieh, "Public and private resource trade-offs for a quantum channel," *Quantum Information Processing*, vol. 11, no. 6, p. 1465–1501, Oct. 2011. [Online]. Available: http://dx.doi.org/10.1007/s11128-011-0317-z

[Salek et al.(2020)Salek, Anshu, Hsieh, Jain, and Fonollosa] F. Salek, A. Anshu, M.-H. Hsieh, R. Jain, and J. R. Fonollosa, "One-shot capacity bounds on the simultaneous transmission of classical and quantum information," *IEEE Transactions on Information Theory*, vol. 66, no. 4, p. 2141–2164, Apr. 2020. [Online]. Available: http://dx.doi.org/10.1109/TIT.2019.2945800

[Anshu et al.(2017)Anshu, Devabathini, and Jain] A. Anshu, V. K. Devabathini, and R. Jain, "Quantum communication using coherent rejection sampling," *Physical Review Letters*, vol. 119, no. 12, p. 120506, 2017.

[ANSHU(2018)] A. ANSHU, "One-shot protocols for communication over quantum networks: Achievability and limitations," Ph.D. dissertation, 2018.

[Anshu et al.(2019a)Anshu, Jain, and Warsi] A. Anshu, R. Jain, and N. A. Warsi, "Building blocks for communication over noisy quantum networks," *IEEE Transactions on Information Theory*, vol. 65, no. 2, p. 1287–1306, Feb. 2019. [Online]. Available: http://dx.doi.org/10.1109/TIT.2018.2851297

[Csiszár and Körner(2011a)] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[Cai(2018)] M. Cai, "Classical-quantum channels: Secret transmission under attacks," Ph.D. dissertation, 2018.

[Blackwell et al.(1959)Blackwell, Breiman, and Thomasian] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[Wolfowitz(1960)] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Analysis*, vol. 4, no. 4, pp. 371–386, 1960.

[Ahlswede(2015)] R. Ahlswede, *Transmitting and Gaining Data: Rudolf Ahlswede's Lectures on Information Theory 2*, A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds. Springer International Publishing, 2015.

[Tsfasman et al.(2007)Tsfasman, Vladut, and Nogin] M. Tsfasman, S. Vladut, and D. Nogin, *Algebraic Geometric Codes: Basic Notions.* Providence: American Mathematical Society, 2007.

[Joyner and Kim(2011)] D. Joyner and J.-L. Kim, *Selected Unsolved Problems in Coding Theory.* Basel: Birkhäuser, 2011.

[Pour-El and Richards(2017)] M. B. Pour-El and J. I. Richards, *Computability in Analysis and Physics.* Cambridge: Cambridge University Press, 2017.

[Turing(1936)] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proc. London Math. Soc.*, vol. 2, no. 42, pp. 230–265, 1936.

[Turing(1937)] ——, "On computable numbers, with an application to the Entscheidungsproblem. A correction," *Proc. London Math. Soc.*, vol. 2, no. 43, pp. 544–546, 1937.

[Church(1936)] A. Church, "An unsolvable problem of elementary number theory," *American Journal of Mathematics*, vol. 58, no. 2, pp. 345–363, 1936. [Online]. Available: http://www.jstor.org/stable/2371045

[Soare(1987)] R. I. Soare, *Recursively Enumerable Sets and Degrees.* Berlin, Heidelberg: Springer-Verlag, 1987.

[Weihrauch(2000)] K. Weihrauch, *Computable Analysis - An Introduction.* Berlin, Heidelberg: Springer-Verlag, 2000.

[Rudin(1987)] W. Rudin, *Real and Complex Analysis*, 3rd ed. Mcgraw-Hill Higher Education, 1987.

[Specker(1949)] E. Specker, "Nicht konstruktiv beweisbare Sätze der Analysis," *Journal of Symbolic Logic*, vol. 14, no. 3, pp. 145–158, Sep. 1949.

[Yagi and Nomura(2014)] H. Yagi and R. Nomura, "Single-letter characterization of epsilon-capacity for mixed memoryless channels," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 2874–2878.

[Csiszár and Körner(1981)] I. Csiszár and J. Körner, *Information Theory - Coding Theorems for Discrete Memoryless Systems*, 1st ed. Academic Press, 1981.

[Ahlswede(1967b)] R. Ahlswede, "Certain results in coding theory for compound channels," in *Proc. Colloquium Inf. Th.* Debrecen, Hungary: Bolyai Mathematical Society, 1967, pp. 35–60.

[H. Boche(2017)] S. H. Boche, G. Janß en, "Entanglement-assisted classical capacities of compound and arbitrarily varying quantum channels," *Quantum information processing*, vol. 16, 2017.

[Boche et al.(2019d)Boche, Janßen, and Saeedinaeeni] H. Boche, G. Janßen, and S. Saeedinaeeni, "Universal random codes: Capacity regions of the compound quantum multiple-access channel with one classical and one quantum sender," *Quantum Information Processing*, vol. 18, no. 246, 2019.

[Schaefer and Boche(2014c)] R. F. Schaefer and H. Boche, "Physical layer service integration in wireless networks: Signal processing challenges," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 147–156, May 2014.

[Anshu et al.(2019b)Anshu, Jain, and Warsi] A. Anshu, R. Jain, and N. Warsi, "A hypothesis testing approach for communication over entanglement-assisted compound quantum channel," *IEEE Transactions on Information Theory*, vol. 65, pp. 2623–2636, 2019.

[Wilde et al.(2019)Wilde, Khatri, Kaur, and Guha] M. M. Wilde, S. Khatri, E. Kaur, and S. Guha, "Second-order coding rates for key distillation in quantum key distribution," 2019, available online at https://arxiv.org/abs/1910.03883.

[Boche et al.(2020a)Boche, Schaefer, and Poor] H. Boche, R. F. Schaefer, and H. V. Poor, "Robust transmission over channels with channel uncertainty: An algorithmic perspective," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Barcelona, Spain, May 2020.

[Boche et al.(2020b)Boche, Janßen, and Saeedinaeeni] H. Boche, G. Janßen, and S. Saeedinaeeni, "Universal superposition codes: Capacity regions of compound quantum broadcast channel with confidential messages," *Journal of Mathematical Physics*, vol. 61, no. 4, p. 042204, 2020.

[Csiszár and Körner(2011b)] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, UK: Cambridge University Press, 2011.

[Yard et al.(2005)Yard, Devetak, and Hayden] J. Yard, I. Devetak, and P. Hayden, "Capacity theorems for quantum multiple access channels," *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, pp. 884–888, 2005.

[Shirokov(2017)] M. Shirokov, "Tight uniform continuity bounds for the quantum conditional mutual information, for the holevo quantity, and for capacities of quantum channels," *Journal of Mathematical Physics*, vol. 58, p. 102202, 2017.

[A. Grigorescu and Poor(2015)] R. F. S. A. Grigorescu, H. Boche and H. V. Poor, "Capacity region continuity of the compound broadcast channel with confidential messages," *IEEE Information Theory Workshop, Jerusalem*, pp. 1–5, 2015.

[Bjelakovic and Boche(2007)] I. Bjelakovic and H. Boche, "Classical capacities of averaged and compound quantum channels," *ArXiv*, vol. abs/0710.3027, 2007.

[Hayashi(2009)] M. Hayashi, "Universal coding for classical-quantum channel," *Communications in Mathematical Physics*, vol. 289, pp. 1087–1098, 2009.

[Datta and Hsieh(2010)] N. Datta and M.-H. Hsieh, "Universal coding for transmission of private information," *Journal of Mathematical Physics*, vol. 51, pp. 122 202–122 202, 2010.

[Mosonyi and Hiai(2011)] M. Mosonyi and F. Hiai, "On the quantum rényi relative entropies and related capacity formulas," *IEEE Transactions on Information Theory*, vol. 57, pp. 2474–2487, 2011.

[Hayashi and Nagaoka(2003)] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1753–1768, 2003.

[Bhatia(1996)] R. Bhatia, *Matrix Analysis.* Springer-Verlag, 1996.