TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Theoretische Informationstechnik

Quantum protocols for coherence and security under source uncertainties

Gisbert Janßen

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereichten Dissertation.

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

Die Dissertation wurde am 09.06.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 26.09.2016 angenommen.

Zusammenfassung

In dieser Arbeit werden Protokolle zur Bewältigung verschiedener Kommunikationsaufgaben unter Beteiligung quantenmechanischer Quellen unter Einbeziehung von Quellunsicherheit präsentiert.
Zugrunde liegen den Betrachtungen zwei aus der klassichen Informationstheorie wohlbekannte Modelle der Quellunsicherheit: *zusammengesetzte gedächtnislose* und *beliebig variierende* Quantenquellen. In beiden Modellen wird die Statistik der Quelle nicht, wie im Falle einer gedächtnislosen quelle mit perfekter Systemkenntnis, durch einen einzelnen generierenden Quantenzustand beschrieben. Die statistischen Eigenschaften der Quelle sind vielmehr durch eine Konfidenzmenge möglicher Zuständen auf dem Hilbertraum des Systems definiert.
Im ersten der genannten Quellenmodelle, der zusammengesetzten gedächtnislosen Quelle, erhalten die beteiligten Kommunikationsparteien zwar Ausgaben einer gedächtnislosen Quantenquelle. Der generierende Zustand ist den Parteien allerdings unbekannt und lediglich als ein Element der Konfidenzmenge identifiziert.
Das Modell der beliebig variierenden Quantenquelle geht von einem wesentlich höheren Grad der Unwissenheit der Systemnutzer aus. Die Statistik jedes der von der Quelle emittierten Systeme kann durch einen beliebigen, der in der generierenden Menge enthaltenenen Zustände, beschrieben sein.
In Kapitel 3 wird das Quanten-*State Merging* mit klassischer Vorwärtskommunikation unter den genannten Modellen von Quellunsicherheit betrachtet. Unter der Annahme einer zusammengesetzten Zweiparteien-Quantenquelle werden universelle, asymptotisch fehlerfreie, Protokolle für das Quanten-State Merging hergeleitet. Diese erreichen asymptotisch sowohl in ihrer Verschränkungsbilanz als auch hinsichtlich der klassischen Vorwärtskommunikation optimale Raten.
Unter Vorraussetzung einer beliebig variierenden Quantenquelle, erweisen sich die für viele andere Kodierungsprobleme generisch geeigneten Robustifizierungstechniken allerdings als suboptimal.
Unter Anwendung der Ergebnisse aus Kapitel 3 werden in Kapitel 4 Protokolle für die Verschränkungsdestillation mit unidirektionaler klassischer Kommunikation für zusammengesetzte Zweiparteien-Quantenquellen hergeleitet, die asymptotisch optimale Raten der Verschränkungsausbeute erreichen.
Wird das weitergehende Modell einer beliebig variierenden Zweiparteienquelle veranschlagt, erweist sich der Robustifizierungsansatz hier als erfolgreich. Dies ermöglicht die Herleitung von Protokollen, die auch für dieses Quellenmodell asymptotisch optimal bezüglich ihrer Verschränkungsausbeute sind.
In Kombination mit entsprechenden Beweisen der Umkehrungen ergeben sich daher vollständige Kodierungssätze für die Verschränkungsdestillation unter Assistenz unidirektionaler klas-

sischer Kommunikation für beide Modelle von Systemunsicherheit. Wie in der Quanteninformationstheorie häufig der Fall, werden die Kapazitäten funktionell durch *Multi-Letter*-Formeln charakterisiert, die im allgemeinen nicht vollständig durch die Statistik einer einzelnen Quellausgabe bestimmt sind. Allerdings erlauben die Kapazitätsformeln, eine Stetigkeit und damit Stabilität der Güte des Systems bei Störung der Konfidenzmenge nachzuweisen.
In Kapitel 5 wird die Erzeugung klassischer sicherer Schlüssel aus zusammengesetzten teilklassischen Dreiparteien-Quantenquellen untersucht. Neben zwei legitimierten Nutzern des Systems hat ausserdem eine dritte, abhörende, Kommunikationspartei Zugriff auf von der Quelle emittierte Teilsysteme.
Die legitimerten Nutzer haben die Aufgabe, eine möglichst perfekt gleichverteilte gemeinsame Zufallsvariable aus ihren Quellausgaben zu generieren. Darüberhinaus dürfen dem Abhörer Messungen auf seinen Systemen asymptotisch keine Rückschlüsse auf die Ergebnisse dieser Zufallsgrösse ermöglichen. Zur Quantifizierung dieser Geheimhaltungsanforderung wird ein starkes Sicherheitskriterium veranschlagt.
Unter der Annahme, dass einer der legitimierten Nutzer lediglich klassische Quellausgaben erhält und dieser als Sender zusätzlich die Freiheit öffentlicher klassischer Kommunikation hat, werden universelle Protokolle zur Erzeugung solch sicherer Schlüssel entwickelt. Unter zusätzlicher Vorraussetzung einer geeigneten Regularitätsbedingung an die Konfidenzmenge von Zuständen erweisen sich diese als asymptotisch optimal bezüglich der erreichten Schlüsselraten.
Kontrastierend dazu wird die Aufgabe der Erzeugung sicherer Schlüssel auch in dem Fall untersucht, dass der sendende, legitimierte, Nutzer perfekte Kenntnis der Marginalstatistik auf seinen Teilsystemen besitzt. Es wird eine funktionelle Charakterisierung der Güte für die Schlüsselerzeug hergeleitet, wobei hier auf eine zusätzliche Regularitätsbedingung verzichtet werden kann. Interessanterweise gleichen sich Schlüsselerzeugungskapazitäten mit und ohne diese Kenntnis des Marginalzustands, solange die generierende Konfidenzmenge die gennante Regularitätsbedingung erfüllt, während die Kapazitäten substantiell differieren können, falls die Quelle nicht regulär ist.
Anhand eines Beispiels wird demonstriert, dass perfekte Kenntnis der Marginalstatistik beim legitimierten Sender bisweilen extreme Vorteile gegenüber der Situation ohne diese Kenntnis zur Folge hat. Während mit Marginalkenntnis des legitimierten Senders auf einfachem Wege ein in Sicherheit und Gleichverteilung perfekter Schlüssel schon in endlicher Blocklänge generiert werden kann, ist ohne diese Kenntnis keine positive Rate der Schlüsselerzeugung erreichbar.
Abschließend werden die oben genannten Regularitätsbedingungen einer genaueren Betrachtung unterzogen. Es zeigt sich, dass allgemeine Resultate über Halbstetigkeiten mengenwertiger Abbildung in Kombination mit den gegeben Performanz- und Sicherheitskriterien sogar eine Abschächung der oben genannten Regularitätsbedingungen erlauben.

Abstract

In this thesis, protocols for several communication tasks involving quantum sources are presented, which are robust against statistical uncertainties of the involved systems.
Two prominent models of source uncertainty which are well known in classical as well as quantum information theory are considered, *compound memoryless* and *arbitrarily varying* quantum sources. For both models, the statistics of the source is - opposed to the case of perfect system knowledge - rather described by a set of generating states than by a single state.
If the first of the mentioned models, the compound memoryless quantum source, is considered, the communication parties receive systems emitted by a memoryless source, while the generating state can be any element from the set of possible states.
If an arbitrarily varying quantum source is considered, the users face system uncertainties on a substantially higher level. Each of the source outputs can have statistics according to any state from the set.
In Chapter 3 the quantum state merging task is considered for both of the mentioned models of source uncertainty. First, presence of a compound memoryless bipartite quantum source is assumed. Universal and asymptotically faithful protocols for quantum state merging with classical forward communication are developed, which achieve optimal entanglement as well as classical forward communication rates.
In case of a general arbitrarily varying quantum source, it turns out, that the so-called robustification approach which is generically sucessful for several other coding problems, turns out to be suboptimal.
As an application of the results from Chapter 3, one-way entanglement distillation protocols for bipartite compound memoryless quantum sources, which are optimal regarding their entanglement rates are derived.
For arbitrarily varying bipartite quantum sources, the aforementioned robustification approach is shown to be suitable. In this way, protocols which achieve optimal entanglement rates are developed also for general arbitrarily varying sources.
The derived protocols combined with proofs for the corresponding converse bounds result in full coding theorems for one-way entanglement distillation for both models of system uncertainty. As often observed in quantum information theory, a multi-letter characterization of the capacities is given rather than a description in terms of functions which can be evaluated on a single source output. However, the obtained capacity functions are shown to be continuous, which at least guarantees a certain stability of the capacities against perturbations of the uncertainty sets.
In Chapter 5, generation of secret keys from compound memoryles tripartite semiclassical quantum sources is considered. Besides two legitimate users of the system, also an eavesdropping third party has access to subsystems emitted by the source.

The legitimate users try to generate a perfectly correlated and equdistributed joint random variable from their samples. Moreover, it is demanded, that the eavesdropper gains asymptocally zero knowledge about the random variable from performing measurements on his/her subsystems. To quantify privacy of the key, a strong security criterion is established.
Assuming one of the legitimate parties receiving classical systems from the source and moreover being granted with the possibility of unlimited classical public communication, universal protocols for secret-key distillation are derived. If in addition a certain regularity condition is fulfilled for the generating uncertainty set of states, the achieved secret-key rates turn out to be optimal in the asymptotic limit of infinitely large numbers of source outputs.
To contrast the obtained results, secret-key distillation from compound memoryless semiclassical sources is considered also in situations, where the legitimate sender is equipped with perfect knowledge of the marginal source statistics on his/her systems deriving from the source state. A capacitiy formula for the forward secret-key distillation capacity is derived for this case, while the assumption of regularity is not necessary.
Interestingly, capacities are equal with and without sender marginal knowledge as long the source fulfills the regularity condition, while they may substiantially differ for non-regular sources.
An example is given to demonstrate, that sender's perfect marginal knowledge may lead to extreme advantage compared to the case without this knowledge being granted. While having marginal knowledge allows to obtain perfect secret-keys at positive capacity via simple protocols which provide perfectly secure keys even for finite blocklengths, lack of sender marginal knowledge forbids to distill secret-keys at any strictly positive rate.
The considerations on regularity are concluded by a review of the mentioned regularity condition. It turns out, that application of the general theory of continuity of set-valued maps together with properties of the defined performance and security measures allow even to weaken the regularity conditions. This yields a larger class of compound memorless quantum sources with a general capacity formula.

Danksagung

Verschiedene Menschen hatten - jeder auf seine eigene Art und Weise - großen Anteil am gelingen dieser Arbeit.

Zuerst möchte ich mich bei Holger Boche bedanken, der mir die Möglichkeit gegeben hat, am Lehrstuhl für Theoretische Informationstechnik unter seiner Betreuung diese Arbeit anzufertigen. Seine Geduld hat mir das ausgiebige Nachdenken ermöglicht und sein mitreissender wissenschaftlicher Enthusiasmus mich so einige Frustration überwinden lassen.
Besonderer Dank gilt Igor Bjelaković, von dem ich gerade am schwierigen Anfang unglaublich viel über wissenschaftliches Arbeiten ("Umwege erhöhen die Ortskenntnis") und Informationstheorie gelernt habe. Igor war mir in der Zeit unserer Zusammenarbeit immer eine Quelle wissenschaftlicher Inspiration und hat auch danach nie das Interesse an meiner Arbeit verloren.
Ein Dank geht auch an meine Kollegen am LTI - besonders an meine Schreibtisch-, Diskussions- und manchmal auch Leidensgenossen Moritz Wiese (Flaschenöffner), Andrea Grigorescu Vlass (Stehcafé), Ezra Tampubolon und Philipp Walk.
Herzlich bedanken möchte ich mich bei Prof. Michael Wolf für seine Bereitschaft, als Gutachter für diese Arbeit zu fungieren. Prof. Gerhard Kramer gilt mein Dank dafür, dass er als Vorsitzender der Prüfungskommission das Promotionsverfahren möglich macht.
Abschließend möchte ich meine Familie und Magdalena erwähnen - ich bin Euch auf unzählbar viele unterschiedliche Arten und Weisen dankbar!

Contents

1 Preliminaries **1**

2 Introduction **7**

3 Quantum state merging under source uncertainties **13**
3.1 Definitions and results 13
3.1.1 Compound memoryless quantum sources 14
3.1.2 Arbitrarily varying quantum sources 15
3.2 Resource cost region for state merging of compound quantum sources - Proofs . 16
3.2.1 Coding theorem 17
3.2.2 Converse theorem 25
3.3 On quantum state merging for arbitrarily varying quantum sources 30

4 One-way entanglement distillation under source uncertainties **35**
4.1 Definitions and results 35
4.1.1 Compound quantum sources 35
4.1.2 Arbitrarily varying quantum sources 37
4.2 Entanglement distillation from compound quantum sources - Proofs 37
4.2.1 Continuity of entanglement distillation capacities 38
4.2.2 Proof of the coding theorem for compound quantum sources 39
4.2.3 An application: Universal entanglement generation codes for compound quantum channels 42
4.3 Entanglement distillation from arbitrarily varying quantum sources 44
4.3.1 AVQS generated by finite sets 45
4.3.2 General AVQS 48

5 Secret-key distillation for compound classical-quantum-quantum sources **53**
5.1 Basic definitions and main Result 53
5.1.1 Source model 53
5.1.2 Secret key generation from compound cqq sources: Definitions and results 54
5.2 Secret-key distillation without state knowledge 58
5.3 Secret-key distillation with sender marginal information (SMI) 83
5.4 Discussion of regularity of compound cqq sources 86
5.4.1 Operational significance of regularity conditions 87
5.4.2 A weaker notion of regularity 89

5.5 Special case: Forward secret-key distillation capacity of a classical tripartite compound sources 92
5.6 Conclusion 93

A Appendix A **95**

B Appendix B **105**

Notation

$\mathbb{R}^+$	Set of nonnegative real numbers
$\mathcal{L}(\mathcal{H},\mathcal{K})$	Set of linear maps from $\mathcal{H}$ to $\mathcal{K}$
$\mathcal{L}(\mathcal{H})$	Set of linear maps from $\mathcal{H}$ to $\mathcal{H}$
$\mathcal{S}(\mathcal{H})$	Set of density matrices on $\mathcal{H}$
$\mathfrak{P}(\mathcal{X})$	Set of probability distributions of $\mathcal{X}$
$\mathcal{C}(\mathcal{H},\mathcal{K})$	Set of completely positive and trace preserving (c.p.t.p.) maps from $\mathcal{L}(\mathcal{H})$ to $\mathcal{L}(\mathcal{K})$ (we also use the term quantum channel)
$\mathcal{C}^{\downarrow}(\mathcal{H},\mathcal{K})$	Set of completely positive and trace nonincreasing maps (operations) from $\mathcal{L}(\mathcal{H})$ to $\mathcal{L}(\mathcal{K})$
$\mathcal{CQ}(\mathcal{X},\mathcal{H})$	Set of classical quantum channels $\mathcal{X} \to \mathcal{S}(\mathcal{H})$
$\mathrm{conv}(A)$	Convex hull of A
$\overline{A}$	Closure of A
A^c	Complement of A
$A \subset B$	A is a (not necessarily proper) subset of B
$\lfloor a \rfloor$	Largest integer not exceeding a
$\lceil a \rceil$	Smallest integer not less then a
$[N]$	The set $\{1,\ldots,N\}$, $N \in \mathbb{N}$
$\exp,\ \log$	Are understood to the base 2
$H(p)$	Shannon entropy of probability distribution p
$h(x)$	Shannon entropy of the binary probability distribution with values $(x, 1-x)$, $x \in [0,1]$
$S(\rho)$	Von Neumann entropy of density matrix ρ
$S(A\|B,\sigma_{AB})$	Conditional von Neumann entropy of bipartite density matrix σ_{AB}
$I(A;B,\sigma_{AB})$	Quantum mutual information of a bipartite state σ_{AB}
$I_c(\rho,\mathcal{N})$	Coherent information of a density matrix ρ and quantum channel $\mathcal{N}$
$I(A\rangle B,\sigma_{AB})$	Coherent information of the bipartite density matrix σ_{AB}
$D(p\|\|q)$	Kullback-Leibler divergence of probability distributions p, q
$D(\rho\|\|\sigma)$	Quantum relative entropy of density matrices ρ, σ
$\mathrm{sr}(\psi)$	Schmidt rank of the bipartite state vector ψ

1 Preliminaries

In this chapter, we fix notation and state some conventions which are freely used throughout this thesis. Many other results have become quite standard in classical and quantum information theory. If not referenced within the text, they can be easily found in the textbooks [CK11] and [Wil13].
All Hilbert spaces appearing in this work are considered to be finite dimensional complex vector spaces. $\mathcal{L}(\mathcal{H})$ is the set of linear maps and $\mathcal{S}(\mathcal{H})$ the set of states (density matrices) on a Hilbert space $\mathcal{H}$ in our notation. For a finite alphabet $\mathcal{X}$ and a Hilbert space $\mathcal{K}$, we denote by $\mathcal{CQ}(\mathcal{X},\mathcal{K})$ the set of classical-quantum channels, i.e. maps from $\mathcal{X}$ to $\mathcal{S}(\mathcal{K})$
Regarding states on multiparty systems, we make use of the following convention. For a system with subsystems belonging to parties X, Y, Z, for instance, we write $\mathcal{H}_{XYZ} \coloneqq \mathcal{H}_X \otimes \mathcal{H}_Y \otimes \mathcal{H}_Z$, and denote the marginals by the letters assigned to subsystems, i.e. $\sigma_{XZ} \coloneqq \mathrm{tr}_{\mathcal{H}_Y}(\sigma)$ for $\sigma \in \mathcal{S}(\mathcal{H}_{XYZ})$ and so on.
We denote the quantum fidelity of positive definite matrices $\rho, \sigma \in \mathcal{L}(\mathcal{H})$ by

$$F(\rho,\sigma) \coloneqq \left\| \rho^{\frac{1}{2}} \sigma^{\frac{1}{2}} \right\|_1^2 .$$

If $\rho,\ \sigma$ are density matrices, we frequently use the inequalities

$$1 - \sqrt{F(\rho,\sigma)} \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho,\sigma)} \tag{1.1}$$

relating fidelity to the trace norm distance of ρ and σ.
Informational quantities
The von Neumann entropy of a quantum state ρ is defined by

$$S(\rho) \coloneqq -\mathrm{tr}(\rho \log \rho),$$

where we denote by $\log(\cdot)$ and $\exp(\cdot)$ the base two logarithms and exponentials throughout this paper. Given a quantum state ρ on $\mathcal{H}_{XY}$, we denote the conditional von Neumann entropy of ρ given Y by

$$S(X|Y,\rho) \coloneqq S(\rho) - S(\rho_Y),$$

the quantum mutual information by

$$I(X;Y,\rho) \coloneqq S(\rho_X) + S(\rho_Y) - S(\rho).$$

A convenient way of representing systems which have quantum as well as classical subsets is to coherify the classical systems. The density matrix

$$\rho := \sum_{x\in\mathcal{X}} P_X(x)\,|x\rangle\langle x| \otimes \rho_x \in \mathcal{S}(\mathcal{H}_X \otimes \mathcal{K}_B) \tag{1.2}$$

represents a density matrix of a source, where the statistics of a subystem is driven by a classical random variable X with values in $\mathcal{X}$ and $\rho_x \in \mathcal{S}(\mathcal{K}_B)$ is a density matrix on $\mathcal{K}_B$ for each $x \in \mathcal{X}$. A quantum system with Hilbert space $\mathcal{H}_X := \mathbb{C}^{|\mathcal{X}|}$ was introduced where each $x \in \mathcal{X}$ corresponds to the element $|x\rangle$ of a once and for all fixed orthonormal basis (we may assume that this is for each system introduced the canonical basis). We set the convention to indicate the quantum systems belonging to coherified classical systems by the corresponding random variable. This convention extends to notation of entropic quantities. E.g.

$$I(X;B,\rho) = H(X) + S(\rho_B) - S(\rho). \tag{1.3}$$

corresponds to the quantum mutual information of the state ρ in (1.2). The conditional quantum mutual information of a density matrix σ_{ABX} is defined

$$I(A;B|X,\sigma) := S(\sigma_{AX}) + S(\sigma_{BX}) - S(\sigma_{ABX}) - S(\sigma_X).$$

If X belongs to a classical system i.e

$$\sigma = \sum_{x\in\mathcal{X}} P_X(x)\ |x\rangle\langle x| \otimes \sigma_{AB,x}$$

with $\rho_{AB,x}$ being a bipartite density matrix on the remaining systems Hilbert spaces, it holds

$$I(A;B|X,\sigma) = \sum_{x\in\mathcal{X}} P_X(x)\ I(A;B,\rho_{AB,x}).$$

Whenever informational quantities are evaluated on classical systems, we feel free to express them in terms of the corresponding classical informational quantities evaluated on the corresponding probability distributions resp. random variables where we completely adopt the notation and calculational rules as presented in [CK11] if no special reference is given.

Types, typical sequences/subspaces
From there, we also take the definition and properties of types and typical sequences. For given alphabet $\mathcal{X}$ and $n \in \mathbb{N}$ (which we always regard being finite) we denote the set of probability distributions on $\mathcal{X}$ as $\mathfrak{P}(\mathcal{X})$. We will use $[N]$ use as a shortcut for the set $\{1,\dots,N\}$ for each $N \in \mathbb{N}$. The set of types (i.e. empirical distributions) on $\mathcal{X}^n$ is denoted by $\mathfrak{T}(n,\mathcal{X})$, it holds

$$|\mathfrak{T}(n,\mathcal{X})| \le (n+1)^{|\mathcal{X}|}.$$

For given type $\lambda \in \mathcal{T}(n,\mathcal{X})$, we denote the set of λ-typical words in $\mathcal{X}^n$ by T^n_λ. For each $\delta > 0$, $p \in \mathfrak{P}(\mathcal{X})$, the set of δ-typical sequences for p in $\mathcal{X}^n$ is defined by

$$T^n_{p,\delta} := \left\{x^n \in \mathcal{X}^n : \forall a \in \mathcal{X} : |\tfrac{1}{n}N(a|x^n) - p(a)| \le \delta \ \wedge\ p(a) = 0 \Rightarrow N(a|x^n) = 0\right\}, \tag{1.4}$$

where $N(a|x^n)$ is the number of occurrences of a in x^n. Several kinds of bounds are known for these sets, we will explicitly employ the bound

$$p^n\left((T^n_{p,\delta})^c\right) \leq 2^{-nc\delta^2} \tag{1.5}$$

which holds with the universal constant $c := \frac{2}{\ln 2}$ for each $\delta > 0$ and large enough n.

Hausdorff distance
For any two nonempty sets $\mathcal{X}, \mathcal{Y}$ of a metric space (M, d), their Hausdorff distance is defined by

$$\begin{aligned} d_H(\mathcal{X}, \mathcal{Y}) :&= \max\left\{\sup_{x\in\mathcal{X}}\inf_{y\in\mathcal{Y}} d(x,y), \sup_{y\in\mathcal{Y}}\inf_{x\in\mathcal{X}} d(x,y)\right\} \\ &= \inf\{\epsilon > 0 : \ \mathcal{Y}' \subset \mathcal{X}_\epsilon \ \wedge \mathcal{X} \subset \mathcal{Y}_\epsilon\} \end{aligned}$$

where A_ϵ denotes the ϵ-blowup of A (with regard to d) for each set A. We use this concept mainly with the underlying spaces being the set of probability distributions on a finite set equipped with the metric induced by the variational distances or on the set of density matrices on a Hilbert space equipped with the trace distance. On the set of subsets of a bounded set, d_H has only finite values. Several properties of the Hausdorff distance are inherited from the trace norm. We will use the triangle inequality

$$d_H(A, C) \leq d_H(A, B) + d_H(B, C) \qquad (A, B, C \subset \mathcal{H}) \tag{1.6}$$

and monotonicity of the Hausdorff distance under c.p.t.p. maps, i.e. for each $\mathcal{N} \in \mathcal{C}(\mathcal{H}, \mathcal{K})$, $A, B \in \mathcal{S}(\mathcal{H})$, it holds

$$d_H(A, B) \geq d_H(\mathcal{N}(A), \mathcal{N}(B)), \tag{1.7}$$

where $\mathcal{N}(X)$ is the image of each set X under $\mathcal{N}$.

Local operations and classical communication (LOCC) channels
An important class of protocols for our considerations in Chapter 3 and Chapter 4 are local operations and classical communication (LOCC) channels In this section, we give a short account to the class of one-way LOCC channels which we use in our considerations. For further information, we recommend the survey article by Keyl [Key02] (and references therein). A more recent general treatment can be found in Ref. [Chi+14].
Crucial for the definition of LOCC channels is the concept of an instrument. Instruments (or operation valued measures[DL70]) were introduced to model the situation, where a measurement is made, and not only the measurement results but also the state transformations according to the measurement values are taken into account. To each measurement result i, there is assigned a positive trace non-increasing cp map $\mathcal{I}_i$ which transforms the input state. In this paper, we restrict ourselves to finite sets of possible measurement results.

Definition 1. *A (finite)* instrument $\mathcal{A}$ *is a map*

$$\begin{aligned} \mathcal{A} : I &\to \mathcal{C}^{\downarrow}(\mathcal{H}, \mathcal{K}) \\ i &\mapsto \mathcal{A}_i \qquad\qquad (i \in I) \end{aligned}$$

with a finite index set I and Hilbert spaces $\mathcal{H}$, $\mathcal{K}$, such that $\sum_{i\in I} \mathcal{A}_i$ is trace preserving. The instrument $\mathcal{A}$ is completely determined by the family $\{\mathcal{A}_i\}_{i\in I}$. We will sometimes write $\mathcal{A} = \{\mathcal{A}_i\}_{i\in I}$ to denote the instrument $\mathcal{A}$.

For bipartite systems, an instrument at, say, A's (the sender's) site can be combined with a parameter-dependent channel use, which is defined by a function

$$\begin{aligned} \mathcal{B} : I &\to \mathcal{C}(\mathcal{H}_B, \mathcal{K}_B) \\ i &\mapsto \mathcal{B}_i \qquad\qquad (i \in I), \end{aligned}$$

i.e. each $\mathcal{B}_i$ is a completely positive and trace preserving map. A one-way LOCC channel is then defined as a combination of an instrument and a parameter-dependent channel. This leads to the following definition.

Definition 2. *A channel $\mathcal{N} \in \mathcal{C}(\mathcal{H}_{AB}, \mathcal{K}_{AB})$ is called $A \to B$* one-way LOCC channel*, if it takes the form*

$$\mathcal{N}(\rho) = \sum_{i\in I} \mathcal{A}_i \otimes \mathcal{B}_i(\rho) \qquad (\rho \in \mathcal{S}(\mathcal{H}_{AB})), \tag{1.8}$$

where $\mathcal{A} = \{\mathcal{A}_i\}_{i\in I}$, $\mathcal{A}_i \in \mathcal{C}^{\downarrow}(\mathcal{H}_A, \mathcal{K}_A)$, is an instrument and $\{\mathcal{B}_i\}_{i\in I}$ is a parameter-dependent channel.

A one-way LOCC can also again be considered as a "one-way local" instrument[Chi+14] with members $\{\mathcal{A}_i \otimes \mathcal{B}_i\}_{i\in I}$. There is a convenient way of handling one-way LOCCs. One can equivalently write the instrument $\mathcal{A}$ used on A's site in channel form

$$\mathcal{A}(\rho) = \sum_{i\in I} \mathcal{A}_i(\rho) \otimes |e_i\rangle \langle e_i| \qquad (\rho \in \mathcal{S}(\mathcal{H}_A))$$

with an orthonormal basis $\{e_i\}_{i\in I} \subset \mathbb{C}^{|I|}$. If the basis is assigned to a system on B's site (which models a classical communication and coherent storage of the measurement results at the receiver's system), the parameter-dependent channel can be written in the form

$$\mathcal{B}(\rho) := \sum_{i\in I} |e_i\rangle \langle e_i| \otimes \mathcal{B}_i(\rho) \qquad (\rho \in \mathcal{S}(\mathcal{H}_B))$$

(this map may not not be trace-preserving). Then we have for $\rho \in \mathcal{S}(\mathcal{H}_{AB})$

$$\begin{aligned} \mathcal{N}(\rho) &= (id_{\mathcal{K}_A} \otimes \mathcal{B}) \circ (\mathcal{A} \otimes id_{\mathcal{H}_B})(\rho) \\ &= \sum_{j,i\in I} \mathcal{A}_i \otimes \mathcal{B}_j(\rho) \otimes |e_i\rangle \langle e_i| |e_j\rangle \langle e_j| \\ &= \sum_{i\in I} \mathcal{A}_i \otimes \mathcal{B}_i(\rho) \otimes |e_i\rangle \langle e_i|, \end{aligned}$$

where the second line includes a permutation of the tensor factors. Tracing out the classical information exchanged within the application of the map (i.e. the system with space $\mathbb{C}^{|I|}$) leads back to the form given in Eq. (1.8). The more general class of two-way LOCC channels exhibits a more intricate definition for which we refer to [HZ12].
Moreover, Def. 2 should not be confused with the definition of the class of separable channels. A channel $\mathcal{M} \in \mathcal{C}(\mathcal{H}_{AB}, \mathcal{K}_{AB})$ is called *separable*, if it takes the form

$$\mathcal{M}(\rho) = \sum_{i \in I} \mathcal{A}_i \otimes \mathcal{B}_i(\rho) \qquad (\rho \in \mathcal{S}(\mathcal{H}_{AB})) \tag{1.9}$$

where $\mathcal{A}_i \in \mathcal{C}^{\downarrow}(\mathcal{H}_A, \mathcal{K}_A)$ and $\mathcal{B}_i \in \mathcal{C}^{\downarrow}(\mathcal{H}_B, \mathcal{K}_B)$ for all $i \in I$. From eqns. (1.8) and (1.9), the difference between the one-way LOCC and separable channels can be observed. While separable channels allow general trace decreasing cp maps for both parties, the receiver party is restricted to usage of trace preserving cp maps (i.e. channels) in the one-way LOCC class of channels.

2 Introduction

System uncertainties are an unavoidable feature of real-world communication systems. On the theoretical side, it is therefore a major challenge to develop protocols and coding schemes which have the property of being robust regarding variations of the statistical parameters of communication systems. This thesis presents results of endeavours in this direction within the framework of quantum Shannon theory, where mainly communication systems including statistical uncertainties in the states of quantum sources are considered.
We consider the two following prominent models of source uncertainty, where the statistics of the source is in each of both cases described by a set $\mathfrak{X} := \{\rho_s\}_{s\in S}$ of density matrices on the system Hilbert space:

- **Compound memoryless quantum source.** A *compound memoryless quantum source* models the situation, where the source outputs have memoryless structure in each case, but the generating state can be any member of $\mathfrak{X}$. A block of n outputs of the source is described by a state

$$\underbrace{\rho_s \otimes \cdots \otimes \rho_s}_{n \text{ times}},$$

  where ρ_s can be any state from $\mathfrak{X}$.

- **Arbitrarily varying quantum source (AVQS).** If the communication parties are confronted with an *arbitrarily varying quantum source*, the users cannot even be sure, that the state generating the output statistics remains constant. A block of n outputs of the source are described by a sequence

$$\rho_{s_1} \otimes \cdots \otimes \rho_{s_n},$$

  where $(s_1, \ldots, s_n)$ can be any n-sequence of letters contained in the index set S.
  The AVQS model covers a highly pessimistic assumption on the scenario. It can be understood to include presence of an additional malicious party (a "jammer") which attacks the communication by freely choosing source states from $\mathfrak{X}$.

The above introduced models of source uncertainty have been studied extensively within the framework of classical Shannon theory since the sixties of the past century for source as well as channel coding problems (see the excellent resource [CK11] for an exhaustive overview of the results gained). When the research leading to this thesis was launched, the first results regarding compound and arbitrarily varying systems within quantum Shannon theory where just published or close to be published. Here we mention the full coding theorem for classical

message transmission over compound classical-quantum channels [BB09], the full coding theorem for the quantum capacities of compound quantum channels [BBN08], [BBN09], and the corresponding results [AB07], [Ahl+12] considering the capacities of arbitrarily varying classical-quantum and quantum channels. The mentioned results covered problems for systems under uncertainties of quantum channel resources. Their source counterparts introduced above did receive hardly any scientific attention up to then. Explicitly, only two major results where known, the first being a generalization of the Sanov Theorem [San57] well known in classical stochastics proven in [Bje+05], and a universal version of the quantum source compression theorem [Sch95], which appeared in [Joz+98].
The results presented here summarize an attempt, to remove a bit of this blind spot, and add some insights on this topic to the panorama of results in quantum Shannon theory.
In this work, three different communication tasks are considered, which are introduced in the following.

Forward quantum state merging (results partly published in [BBJ13], [BJ14b]).
Quantum state merging, introduced by Horodecki, Oppenheim, and Winter [HOW05] is regarded as one of the primitive protocols close to the root of the so-called "family of quantum protocols" [DHW04], [Abe+09]. In this document we use the term "(quantum) state merging" to label the corresponding communication task rather then the protocol construction to perform this task as used in [HOW05], [HOW07].
Within the quantum state merging task, two communication parties A and B have access on the outputs of a bipartite quantum source. They aim to transfer A's share of the source state to B such that after the process, B holds systems which are prepared according to the statistics of the complete source output. To accomplish this state transfer, they are allowed to perform a so-called one-way Local Operations and Classical Communications (LOCC) protocol, where they act locally on their systems while A is allowed to send classical messages to B over a noiseless classical channel. As an additional commmunication resource, the have access to additional shared maximally entangled states.
In [HOW07] the optimal asymptotic entanglement and classical communication costs where determined in a case of a memoryless quantum source with perfectly known state.
For a memoryless source with generic density matrix ρ_{AB}, the optimal entanglement cost was determined to be given by the conditional von Neumann entropy $S(A|B)$, which was shown to be achievable with optimal classical forward communication cost rate $I(A;E)$, which is the quantum mutual information between A and a system E purifying the source.
The result from [HOW07] where appealing for several reasons. On one hand, it provided the negative values of the conditional von Neumann entropy a clear operational meaning. Sometimes state merging can be accomplished with a gain rather than consumption of maximally entangled systems (negative entanglement rates), which in turn can be used as a resource in further communication tasks.
The protocol introduced in [HOW07] on the other hand, has the property of being of primitive nature in the sense, that optimal protocols for several other communication tasks can be derived from it. Examples for such tasks are one-way entanglement distillation [Dev05], distributed compression of quantum sources (which is regarded as a fully quantum version of the classical

Slepian Wolf protocol), and entanglement generation of quantum multiple access channels [YHD08].
In light of these facts, it seems highly desirable, to develop protocols for quantum state merging being universal in the sense, that they are robust against imperfections in the user's knowledge of the source density matrix.
Chapter 3 is devoted to this goal. Assuming the bipartite source subject to state merging to be a compound memoryless quantum source, we continue research begun in the author's diploma thesis [Jan10]. We develop protocols, which are optimal regarding their entanglement as well as classical forward communication rates. Moreover we provide a new, stronger, lower bound to the forward classical communication rate demanded to accomplish asymptotically perfect merging of memoryless quantum sources. It turns out, that the needed classical communication rate cannot substantially be decreased by allowing suboptimal entanglement rates. This enables us, to give a full single-letter characterization of the classical forward communication and entanglement rate resource tradeoff region of quantum state merging in case of an arbitrary (not necessarily finite or countable) compound memoryless quantum source.
Assuming presence of an arbitrarily varying quantum source, we point out, that so-called robustification and elimination techniques developed by Ahlswede [Ahl78] in the regime of classical channel coding do not lead to optimal entanglement consumption rates. We give a counterexample, where strictly higher entanglement rates are possible for an arbitrarily varying quantum source generated by a finite set of quantum states, then achieved by application of the robustification technique.

Forward entanglement distillation (results partly published in [BBJ13], [BJ14b]).
Entanglement is known as a powerful and versatile communication resource in quantum information theory. An early striking example of this fact is delivered by the quantum teleportation protocol [Ben+93], where shared maximal entanglement and the possibility of classical forward communication enables two communication parties to simulate a noiseless quantum channel. A further development of this idea seems to open a way to future long-distance transmission lines within a future "quantum" internet. It is impossible to coherently implement repeater stations in the manner established in present technology because the no-cloning theorem of quantum mechanics prevents universal read-and-repeat procedures. Therefore, genuinely "quantum" repeater protocols [Dür+99] built from nested teleportation and entanglement swapping procedures are a promising alternative. To empower such quantum repeaters, it will be necessary to distill nearly maximal entangled pairs from noisy entanglement generated by systems under uncertainty.
In case of a memoryless bipartite quantum source with the generating density matrix being perfectly known to the communication parties, optimal asymptotic entanglement distillation rates of LOCC protocols with uni- as well as bidirectional classical communication where determined in [Dev05]. Therein, a correspondence between protocols for distillation of classical secret keys and quantum entanglement where exploited.
As a first result, we determine the optimal entanglement rates for one-way LOCC protocols for compound memoryless bipartite quantum source. Rather than utilizing the connection between entanglement distillation and classical secret-key distillation mentioned above, we derive en-

tanglement distillation protocols from quantum state merging protocols for compound quantum sources. We characterize the one way entanglement distillation capacity by a multi-letter formula which generalizes the one derived in [Dev05] in this direction.
We apply the obtained results, to give another proof of the coding theorem for the entanglement generation capacity of compound quantum channels originally proven in [BBN09]. Our approach allows a conceptually simple proof, by deriving entanglement generation codes for compound channels from protocols for entanglement distillation from effective joint states obtained between sender and receiver by channel transmission.
Opposed to our observations made for quantum state merging protocols, the robustification and elimination techniques lead to optimal entanglement distillation protocols for arbitrarily varying quantum sources. By following this strategy, we also to determine the optimal rates for one-way entanglement distillation also in case of arbitrarily varying quantum sources. Unfortunately, the characterizations of the one-way entanglement distillation capacities for both models of source uncertainty are given in terms multi-letter formulae which are not very convenient for calculation of numerical values of the entanglement distillation capacities.
However, since the capacity formulas derived in the compound as well as arbitrarily varying case are continuous functions of the uncertainty sets, we are able to prove stability of the problem. If two sets of generating density matrices are close in the Hausdorff distance, their one-way entanglement capacities are close for the corresponding compound and arbitrarily varying quantum sources.

Forward secret key distillation.
Common randomness shared by cooperating communication parties which additionally has the property of being uncorrelated to a third, eavesdropping party is well-known as a valuable resource in classical as well as quantum information theory. This fact becomes apparent e.g. when legitimate parties use a one-time pad coding [Ver26] procedure to securely randomize codewords to enhance the security of messages sent over an insecure transmission line.
A possible way to generate this resource is, to distill it from potentially noisy and insecure correlations pre-shared by the parties. Development of methods to obtain such secret-keys was for long a classic domain cryptographic research [DH76]. The cryptographic approach is usually, to exploit assumed limited computational capabilities of eavesdropping parties which lead to substantial advantages deriving from cooperation of legitimate parties when they apply high-complexity protocols. We follow the more recent so-called information-theoretic approach, where rather the principal limitations of the eavesdropping parties are utilized to obtain security. Initiated by works of Ahlswede and Csiszar [AC93] and Maurer [Mau93], this direction was intensively studied in the past decades, such that integrating security on the physical layer of communication systems more and more heads towards technological application [WTS07].
The possibilities of distilling secret keys by information theoretic methods was also studied for quantum systems in [Dev05], where the secret-key distillation of classical-quantum-quantum sources and completely quantum sources where studied. However, to obtain these results, the sources where assumed to be memoryless with statistical properties (i.e. the generating density matrix) perfectly known to the legitimate parties.
In this thesis, we consider presence of a compound memoryless classical-quantum-quantum (cqq) source where one receiver is assumed to receive outputs of a classical source, while the re-

maining communication parties receive quantum systems. The classical systems-receiving party is also allowed to broadcast classical messages to both of the other parties to support processing. It turns out, that some compound cqq sources are notoriously hard to approximate. This fact leads us to introducing a regularity condition on sets of cqq density matrices, where we demand, the possible sets of marginal states of the sender-legitimate receiver and sender-eavesdropper systems only to differ in a controllable amount when the derived marginals on the sender system alone do not differ much. For the class of density matrices fulfilling this property, the approximation methods we develop lead to a proof of achievability, whose optimal rates are also optimal in general, i.e. we obtain a full characterization of the forward secret-key capacity.
We also consider the case, where the sending party is equipped with perfect knowledge of the marginal distribution on his/her systems derived from the source. The capacities are shown to be equal for the cases with and without this kind of sender information. Moreover, the formula derived is shown to be also valid for all irregular sources.
The reader may ask for a general proof of validity of the mentioned capacity formula also for the case of no sender state information. Regarding this question we prove a disappointing negative result. The forward secret-key distillation capacities with and without sender state information differ substantially for some compound cqq sources.
Things get even harder. A counterexample we introduce shows, that the legitimate parties can achieve positive capacity with zero error and zero correlation of the key in case of sender-state knowledge, while they are unable to achieve any positive rate without sender state knowledge. This sheds some light on the structure of compound cqq sources. Even if there may be weaker regularity conditions than the one presented here which lead to a general capacity formula, the notion of regularity bears an operational core. While perfect knowledge of the sender's marginal state does not help to achieve higher forward secret-key distillation rates for regular sources, irregularity of the source can split both capacities.

3 Quantum state merging under source uncertainties

3.1 Definitions and results

In this section, we provide precise definitions regarding one-way quantum state merging for compound and arbitrarily varying quantum sources. We first describe the general type of protocols we admit for quantum state merging where we are interested in the entanglement as well as classical resource costs of quantum state merging.
A quantum channel $\mathcal{M}$ is an (l, k_l, D_l) $A \to B$ *merging* for bipartite sources on $\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B$, if it is an $A \to B$ LOCC channel (according to Definition 2 in Chapter 1)

$$\mathcal{M} : \mathcal{L}(\mathcal{K}^l_{0,AB} \otimes \mathcal{H}^{\otimes l}_{AB}) \to \mathcal{L}(\mathcal{K}^l_{1,AB} \otimes \mathcal{H}^{\otimes l}_{B'B}), \tag{3.1}$$

with $k_l := \dim \mathcal{K}^l_{A,0} / \dim \mathcal{K}^l_{A,1}$, where we assume $\mathcal{K}_{A,i} \simeq \mathcal{K}_{B,i}$ $(i = 1, 2)$, and

$$\mathcal{M}(x) = \sum_{k=1}^{D_l} \mathcal{A}_k \otimes \mathcal{B}_k(x). \qquad (x \in \mathcal{L}(\mathcal{K}^l_{0,AB} \otimes \mathcal{H}^{\otimes l}_{AB})) \tag{3.2}$$

where $\{\mathcal{A}_k\}_{k=1}^{D_l} \subset \mathcal{C}^{\downarrow}(\mathcal{K}^l_{0,A} \otimes \mathcal{H}^{\otimes l}_A, \mathcal{K}^l_{1,A})$ constitutes an instrument and $\{\mathcal{B}_k\}_{k=1}^{D_l} \subset \mathcal{C}(\mathcal{K}^l_{B,0} \otimes \mathcal{H}^{\otimes l}_B, \mathcal{K}^l_{B,1} \otimes \mathcal{H}^{\otimes l}_{B'B})$ is a family of channels depending on the index $k \in [D_l]$. The spaces $\mathcal{K}^l_{AB,0}, \mathcal{K}^l_{AB,1}$ are understood to represent bipartite systems shared by A and B, which carry the input and output entanglement resources used in the process. As a convention, we will incorporate the maximally entangled states $\phi^l_i \in \mathcal{S}(\mathcal{K}^l_{AB,i})$, $i = 0, 1$ into the definition of the protocol, we set

$$k_l := \frac{\dim \mathcal{K}^l_{0,A}}{\dim \mathcal{K}^l_{1,A}} = \frac{\dim \mathcal{K}^l_{0,B}}{\dim \mathcal{K}^l_{1,B}} = \frac{\mathrm{sr}(\phi^l_0)}{\mathrm{sr}(\phi^l_1)}. \tag{3.3}$$

As a measure of success for merging procedures, we define the *merging fidelity* of $\mathcal{M}_l$ given a state $\rho \in \mathcal{S}(\mathcal{H}^{\otimes l}_{AB})$ by

$$F_m(\rho, \mathcal{M}_l) := F\left(\mathcal{M}_l \otimes \mathrm{id}_{\mathcal{H}^l_E}(\phi^l_0 \otimes \psi), \phi^l_1 \otimes \psi'\right). \tag{3.4}$$

In (3.4), ψ is a purification of ρ with an environmental system described on an additional Hilbert space $\mathcal{H}^l_E$ (usually $\mathcal{H}^l_E = \mathcal{H}^{\otimes l}_E$ with some space $\mathcal{H}_E$), and ψ' is a state identical to ψ but defined on $\mathcal{H}^{\otimes l}_{B'B}$ completely under control of B. In [Jan10], a representation of the merging fidelity was derived, which implies the following properties of F_m, we will use frequently within the next chapters.

Lemma 3 ([Jan10], cf. [BBJ13])**.** *For a state $\rho \in \mathcal{S}(\mathcal{H}_{AB})$ and a cptp map $\mathcal{M}$ as introduced above, it holds*

1. *$F_m(\rho, \mathcal{M})$ does not depend on the chosen purification of ρ*
2. *$F_m(\cdot, \mathcal{M})$ is a convex function.*
3. *$F_m(\rho, \cdot)$ is a linear function.*

Next, we give precise definitions of achievable rate pairs and cost regions for one-way quantum state merging in case of compound and arbitrarily varying bipartite quantum sources. For the next subsections, we will always assume $\mathfrak{X}$ to be any set of bipartite quantum states $\mathfrak{X} := \{\rho_s\}_{s\in\mathbf{S}} \subset \mathcal{S}(\mathcal{H}_{AB})$. The expressions "compound source $\mathfrak{X}$" and "AVQS $\mathfrak{X}$" are shortcuts to the respective compound and arbitrarily varying sources generated by $\mathfrak{X}$.

3.1.1 Compound memoryless quantum sources

In this section, we define achievable rate pairs for one-way quantum state merging of the compound source $\mathfrak{X}$. A compound quantum source models a situation, where A and B receive for each blocklength n outputs with density matrix $\rho_s^{\otimes n}$ where s can be any index from the index set S. Consequently, they have to apply protocols which are universal in the sense, that the merging fidelity is suitably lower-bounded for *each* of the possible statistics generated by the density matrices from $\mathfrak{X}$.

Definition 4. *A pair $(R_q, R_c) \in \mathbb{R} \times \mathbb{R}^+$ is called an* achievable rate pair *for $A \to B$ merging of the compound source $\mathfrak{X}$, if there exists a sequence $\{\mathcal{M}_l\}_{l\in\mathbb{N}}$ of (l, k_l, D_l) $A \to B$ mergings, such that the conditions*

1. $\lim\limits_{l\to\infty} \inf\limits_{\rho\in\mathfrak{X}} F_m(\rho^{\otimes l}, \mathcal{M}_l) = 1$
2. $\limsup\limits_{l\to\infty} \frac{1}{l} \log k_l \leq R_q$
3. $\limsup\limits_{l\to\infty} \frac{1}{l} \log D_l \leq R_c$

are satisfied.

In [Jan10], a principal formula for the optimal possible *entanglement* cost of quantum state merging for compound quantum sources was derived for a situation, where the communication parties are provided with classical forward communication of arbitrary rate. If we denote the *merging cost* of the compound source $\mathfrak{X}$, i.e. optimal achievable entanglement rate with free choice of the classical forward communication rate by $C_{m,\to}(\mathfrak{X})$, the result can be stated as follows.

Theorem 5 ([Jan10], cf. [BBJ13])**.**

$$C_{m,\rightarrow}(\mathfrak{X}) = \sup_{\rho\in\mathfrak{X}} S(A|B,\rho). \tag{3.5}$$

However, the protocol class which was introduced in [Jan10] to prove achievability in the above theorem is suboptimal regarding its classical communication demands in general. In this chapter, we give a description of the full resource tradeoff region for quantum state merging of compound memoryless bipartite quantum sources. We define

Definition 6. *Let $\mathfrak{X} \subset \mathcal{S}(\mathcal{H}_{AB})$ be a set of density matrices. The* resource cost region for $A \rightarrow B$ quantum state merging of $\mathfrak{X}$ *is defined by*

$$\mathfrak{M}_{\rightarrow}(\mathfrak{X}) := \{(R_q, R_c) \in \mathbb{R}\times\mathbb{R}^+ : \ (R_q, R_c) \text{ achievable rate pair for } A \rightarrow B - \text{merging of } \mathfrak{X}\}$$

Theorem 7.

$$\mathfrak{M}_{\rightarrow}(\mathfrak{X}) = \left\{(R_q, R_c) : \ R_q \geq \sup_{\rho\in\mathfrak{X}} S(A|B,\rho) \ \wedge \ R_c \geq \sup_{\rho\in\mathfrak{X}} \ I(A;E,\psi)\right\}. \tag{3.6}$$

The quantum mutual information in (3.6) is evaluated on any purification $\psi \in \mathcal{H}_{ABE}$ of ρ for each $\rho \in \mathfrak{X}$.

It follows immediately from Theorem 7, that the resource cost region of one-way quantum state merging for compound quantum sources is stable under perturbations of the set of density matrices generating the source.

3.1.2 Arbitrarily varying quantum sources

Regarding the task of quantum state merging for arbitrarily varying quantum sources, our cosiderations lead to an interesting negative result.It turns out, that the famous robustification and elimination technique developed by Ahlswede [Ahl78] for deriving the coding theorem for message transmission over an arbitrarily varying classical channel produces protocols being suboptimal in general. We give an example for this fact. Therefore in the following definitions, we concentrate on the entanglement cost of quantum state merging protocols while allowing free classical forward communication.

Definition 8. *A number $R_q \in \mathbb{R}$ is called an* achievable entanglement rate for $A \rightarrow B$ merging of the AVQS $\mathfrak{X}$ *if there exists a finite number R_c and a sequence $\{\mathcal{M}_l\}_{l\in\mathbb{N}}$ of (l, k_l, D_l) $A \rightarrow B$ mergings satisfying*

1. $\lim\limits_{l\rightarrow\infty} \inf\limits_{s^l\in\mathbf{S}^l} F_m(\rho_{s^l}, \mathcal{M}_l) = 1$,
2. $\limsup\limits_{l\rightarrow\infty} \frac{1}{l}\log k_l \leq R_q$, *and*

3. $\limsup_{l\to\infty} \frac{1}{l} \log D_l \leq R_c$.

Definition 9. *The $A \to B$ merging cost $C^{AV}_{m,\to}(\mathfrak{X})$ of the AVQS $\mathfrak{X}$ is defined by*

$$C^{AV}_{m,\to}(\mathfrak{X}) := \inf \left\{ R_q \in \mathbb{R} : \begin{array}{c} R_q \text{ is an achievable entanglement rate for } A \to B \text{ merging} \\ \text{of the AVQS } \mathfrak{X} \text{ with some classical communication rate } R_c \end{array} \right\} \tag{3.7}$$

We will, by giving a suitable example, prove the following claim.

Example 10. *There exists a set $\tilde{\mathfrak{X}}$ of bipartite density matrices, such that*

$$C^{AV}_{m,\to}(\tilde{\mathfrak{X}}) < \sup_{\rho \in \mathrm{conv}(\tilde{\mathfrak{X}})} S(A|B,\rho) \tag{3.8}$$

holds.

3.2 Resource cost region for state merging of compound quantum sources - Proofs

In this section, we prove Theorem 7. We first show achievability in Section 3.2.1. The converse statement is shown in the subsequent Section 3.2.2. Before we present the proofs, we introduce an important class of protocols for performing quantum state merging which where initially introduced in [HOW07] and also considered in [Jan10].
Let d_A be the dimension of a Hilbert space $\mathcal{H}_A$. For an integer $0 < L \leq d_A$ we use the term *L-merging* if we speak of a channel

$$\mathcal{M} : \mathcal{L}(\mathcal{H}_{AB}) \to \mathcal{L}(\mathcal{K}_{AB}) \otimes \mathcal{L}(\mathcal{H}_{B'B})$$

which is of the form

$$\mathcal{M}(\rho) = \sum_{k=0}^{D} a_k \otimes u_k(\rho) a_k^* \otimes u_k^*, \tag{3.9}$$

for every $\rho \in \mathcal{S}(\mathcal{H}_{AB})$. and has the following properties. D is defined $D := \lfloor \frac{d_A}{L} \rfloor$ and $\mathcal{K}_A$ and $\mathcal{K}_B$ are Hilbert spaces with $\dim \mathcal{K}_A = \dim \mathcal{K}_B = L$ and $\mathcal{K}_A \subseteq \mathcal{H}_A$ is a subspace of $\mathcal{H}_A$, where

- $\{a_k\}_{k=0}^D \subset \mathcal{L}(\mathcal{H}_A, \mathcal{K}_A)$ is a set of rank L partial isometries (except a_0 which has rank $d_A - L \cdot D < L$) with pairwise orthogonal initial subspaces (in the following, we call such channels *L-instrument* for short).
- $\{u_k\}_{k=0}^D \subset \mathcal{L}(\mathcal{H}_B, \mathcal{K}_B \otimes \mathcal{H}_{B'B})$ is a family of isometries.

Remark 11. *Note, that we introduced L-mergings as protocols which do not consume additional shared maximal entanglement. If a shared maximal entangled state $|\phi\rangle\langle\phi|$ has to be consumed by a protocol for merging a state ρ_{AB}, an L-merging can be applied on the state*

$$\rho_{AB} \otimes |\phi\rangle\langle\phi| \tag{3.10}$$

instead.

The following proposition states a universal one-shot result for the above introduced class of merging schemes and finite set of possible source density matrices. It will serve as a starting point for our considerations. For a complete proof the reader may confer [BBJ13].

Theorem 12 ([BBJ13],Corollary 2)**.** *Let $N \in \mathbb{N}$, and $\{\rho_{AB,i}\}_{i=1}^N \subset \mathcal{S}(\mathcal{H}_{AB})$ be a set of density matrices. For each $L \in \mathbb{N}$, there exists an L-merging $\mathcal{M}$ such that*

$$F_m(\overline{\rho}_{AB}, \mathcal{M}) \geq 1 - 2\left(\frac{L}{d_A} + 2\sum_{i=1}^{N}\sqrt{L \cdot \operatorname{rank}(\rho_{AB,i})\|\rho_{B,i}\|_2^2}\right), \tag{3.11}$$

with $\overline{\rho}_{AB} := \frac{1}{N}\sum_{n=1}^{N}\rho_{AB,i}$. Since $F_m(\cdot, \mathcal{M})$ is a convex function, it also holds

$$\min_{i\in[N]} F_m(\rho_{AB,i}, \mathcal{M}) \geq 1 - 2N\left(\frac{L}{d_A} + 2\sum_{i=1}^{N}\sqrt{L \cdot \operatorname{rank}(\rho_{AB,i})\|\rho_{B,i}\|_2^2}\right).$$

3.2.1 Coding theorem

In this section we prove the achievability part of Theorem 7, which is the following statement.

Theorem 13. *Let $\mathfrak{X} \subset \mathcal{S}(\mathcal{H}_{AB})$ be a set of density matrices. It holds*

$$\left\{R_q \geq \sup_{\rho\in\mathfrak{X}} S(A|B,\rho) \wedge R_c \geq \sup_{\rho\in\mathfrak{X}} I(A;E,\psi)\right\} \subset \mathfrak{M}(\mathfrak{X}) \tag{3.12}$$

where $I(A;E,\psi)$ is the quantum mutual information between the A and E systems in a purification ψ of ρ for each $\rho \in \mathfrak{X}$.

The preliminary Proposition 14 below is a slight generalization of Theorem 6 in [BBJ13]. For further application in the next chapter on entanglement distillation, we have to ascertain existence of protocols with merging fidelity going to one asymptotically with *exponentially decreasing* tradeoffs, which was not explicit in [BBJ13]. Proposition 14 states existence of protocols achieving the optimal entanglement rate with the mentioned performance, but with generally suboptimal classical communication demands. These protocols in turn, will be utilized to derive protocols suitable for the proof of Proposition 17.

Proposition 14. *Let $\mathfrak{X} \subset \mathcal{S}(\mathcal{H}_{AB})$ be a set of states on $\mathcal{H}_{AB}$. For each $R > \sup_{\rho\in\mathfrak{X}} S(A|B,\rho)$, there is a number $l_0(\mathfrak{X}, R, \delta)$, such that for each $l > l_0$ we find an (l, k_l, D_l)-$A \to B$ merging*

$$\mathcal{M}(\cdot) := \sum_{k=1}^{D_l} \mathcal{A}_k \otimes \mathcal{U}_k(\cdot) \tag{3.13}$$

$$\mathcal{A}_k := A_k(\cdot)A_k^*, \text{ and } \mathcal{U}_k := U_k(\cdot)U_k^* \qquad (k \in [D_l]) \tag{3.14}$$

with $\{A_k^* A_k\}_{k=1}^{D_l}$ being a projection valued measure and $\{U_k\}_{k=1}^{D_l}$ being isometries, such that

$$\inf_{\rho\in\mathfrak{X}} F_m(\rho^{\otimes l}, \mathcal{M}) \geq 1 - 2^{-lc_1} \tag{3.15}$$

holds with a constant $c_1(\mathfrak{X}, R, \delta) > 0$,

$$k_l = \lfloor \exp(nR) \rfloor, \tag{3.16}$$

and

$$D_l = \lceil \exp(\sup_{\rho\in\mathfrak{X}} I(A; E, \psi) \rceil, \tag{3.17}$$

where the quantum mutual information in (3.17) is evaluated on the AE marginal state of any purification ψ of ρ, where E labels the additional systems for purification.

Proof. The assertion to prove includes both, a strengthening of the fidelity convergence rates in Ref. [Jan10], Theorem 4 to exponentially decreasing trade-offs, and a generalization of Theorem 6 in Ref. [BBJ13] to arbitrary (not necessary finite or countable) sets of states.
Write $R = \sup_{\rho\in\mathfrak{X}} S(A|B,\rho) + \delta$. Approximating $\mathfrak{X}$ by a τ_l-net $\mathfrak{X}_{\tau_l} := \{\rho_i\}_{i=1}^{N_{\tau_l}} \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ for each blocklength l (see Ref. [BBJ13] for details) and using the result for finite sets, we infer by careful observation of the merging fidelities in Ref. [BBJ13] (see eqns. (36), (37), and (58) therein), that for given $\delta > 0$ and large enough blocklength l, there exists a (l, k_l, D_l) $A \to B$ merging $\mathcal{M}_l$ with the properties stated in (3.13) and (3.14) , where

$$\inf_{\rho\in\mathfrak{X}} F_m(\rho^{\otimes l}, \mathcal{M}_l) \geq 1 - N_{\tau_l}^2 \cdot 2^{-l\theta} - 4\sqrt{l \cdot \tau_l} \tag{3.18}$$

is valid for the merging fidelities with a constant $\theta = \theta(\delta) > 0$, and

$$\frac{1}{l} \log k_l \leq \sup_{\rho\in\mathfrak{X}} S(A|B,\rho) + \frac{\delta}{2}. \tag{3.19}$$

(see (57) in Ref. [BBJ13]). Moreover, we can bound the number of messages for the classical $A \to B$-communication (see (101) in Ref. [BBJ13]) by,

$$\frac{1}{l} \log D_l \leq \max_{1\leq i\leq N_{\tau_l}} S(\rho_{A,i}) + \max_{1\leq i\leq N_{\tau_l}} S(A|B,\rho_i) + \frac{\delta}{2} + \frac{1}{l} \log N_{\tau_l} \tag{3.20}$$

$$\leq \sup_{\rho\in\mathfrak{X}} S(\rho_A) + \sup_{\rho\in\mathfrak{X}} S(A|B,\rho) + \overline{\nu}(\tau_l) + \frac{\delta}{2} + \frac{1}{l} \log N_{\tau_l}, \tag{3.21}$$

where the summand $\overline{\nu}(\tau_l) := 3\tau_l \log \frac{\dim \mathcal{H}_{AB}}{\tau_l}$ follows from threefold application of Fannes' inequality [Fan73], i.e.

$$\left| \max_{1\leq i\leq N_{\tau_l}} S(\rho_{A,i}) + \max_{1\leq i\leq N_{\tau_l}} S(A|B,\rho_i) - \sup_{\rho\in\mathfrak{X}} S(\rho_A) + \sup_{\rho\in\mathfrak{X}} S(A|B,\rho) \right| \leq \overline{\nu}(\tau_l). \tag{3.22}$$

Due to the bound given in Ref. [BBJ13], Lemma 9, it is known, that the nets can be chosen with cardinality bounded by

$$N_{\tau_l} \leq \left(\frac{3}{\tau_l}\right)^{2(\dim \mathcal{H}_{AB})^2} \tag{3.23}$$

for each $l \in \mathbb{N}$. Choosing net parameter $\tau_l = 2^{-l\theta'}$ with $\theta' := \min\{\theta/8(\dim \mathcal{H}_{AB})^2, \delta/4\}$ for each l, we infer

$$\inf_{\rho\in\mathfrak{X}} F_m(\rho^{\otimes l}, \mathcal{M}_l) \geq 1 - 2^{-l\frac{\theta}{2}} - 2^{-l\frac{\theta'}{4}} \geq 1 - 2^{-lc_1} \tag{3.24}$$

with a constant $c_1 = c_1(\delta) > 0$, and

$$\frac{1}{l} \log D_l \leq \sup_{\rho\in\mathfrak{X}} S(\rho_A) + \sup_{\rho\in\mathfrak{X}} S(A|B,\rho) + \delta \tag{3.25}$$

from (3.21) if l is large enough, to satisfy $\overline{\nu}(\tau_l) \leq \frac{\delta}{4}$. Collecting the bounds in (3.19), (3.24), and (3.25), we are done. □

The assertion of Proposition 17 below states existence of protocols for each large enough block-length, which are aproximately optimal regarding their entanglement as well as classical communication demands. The proof utilizes the protocols derived to prove Propostion 14 above together with a suitable estimation of the von Neumann entropy of the source state on the A systems. As a prerequisite, we collect some results from representation theory of the symmetric groups, where we recommend [Sim96] as reference. We denote by $YF_{d,l}$ the set of young frames with at most d rows and l boxes for $d, l \in \mathbb{N}$. A young frame $\lambda \in YF_{d,l}$ is determined by a tuple $(\lambda_1, ..., \lambda_d)$ of nonnegative integers summing to l. The box-lengths $\lambda_1, ..., \lambda_d$ of λ define a probability distribution $\overline{\lambda}$ on $[d]$ in a natural way via the definition $\overline{\lambda}(i) := \frac{1}{l}\lambda_i$ for each $1 \leq i \leq d$. To each Young frame $\lambda \in YF_{d,l}$, there is an invariant subspace of $(\mathbb{C}^d)^{\otimes l}$, and we denote by $P_{\lambda,l}$ the projector onto the subspace belonging to λ.
Theorem 15 below allows, to asymptotically estimate the spectrum of a density operator ρ by projection valued measurements on i.i.d. sequences of the form $\rho^{\otimes l}$, and is an important ingredient of our proof of Proposition 17. A variant of the first statement of the theorem was first proven in by Keyl and Werner[KW01]. The bounds stated below are from Ref. [CM06], while the remaining statements of the theorem are well-known facts in group representation theory (Ref. [CM06] and references therein are recommended for further information).

Theorem 15 (cf. Refs. [KW01] and [CM06])**.** *The following assertions are valid for each* $d, l \in \mathbb{N}$.

1. *For $\lambda \in YF_{d,l}$ and $\rho \in \mathcal{S}(\mathbb{C}^d)$, it holds*

$$\mathrm{tr}(P_{\lambda,l}\rho^{\otimes l}) \leq (l+1)^{d(d-1)/2} \exp(-lD(\overline{\lambda}||r)) \tag{3.26}$$

where $\overline{\lambda} \in \mathfrak{P}([d])$ is the probability distribution given by the normalized box-lengths of λ, and r is the probability distribution on $[d]$ induced by the decreasingly ordered spectrum of ρ (with multiplicities of eigenvalues counted), and $D(\overline{\lambda}||r)$ denotes the Kullback-Leibler divergence of $(\overline{\lambda}, r)$.

2. *$|YF_{d,l}| \leq (l+1)^d$.*

3. *For $\lambda, \lambda' \in YF_{d,l}$, it holds $P_{\lambda,l}P_{\lambda',l} = 0$ if $\lambda \neq \lambda'$.*

Lemma 16 (Refs. [Win99], [ON07])**.** *Let τ, X be matrices with $\tau \geq 0$, $\mathrm{tr}(\tau) \leq 1$, and $0 \leq X \leq 1$, $\epsilon \in (0,1)$. If $\mathrm{tr}(\rho X) \geq 1 - \epsilon$, it holds*

$$\|\sqrt{X}\rho\sqrt{X} - \rho\|_1 \leq 2\sqrt{\epsilon}. \tag{3.27}$$

The following proposition is the main result of this section.

Proposition 17. *Let $\mathfrak{X} \subset \mathcal{S}(\mathcal{H}_{AB})$ be a set of states on $\mathcal{H}_{AB}$. For each $\delta > 0$, there exists a number $l_0 = l_0(\delta)$, such that for each $l > l_0$ there is an (l, k_l, D_l) $A \to B$ merging $\mathcal{M}_l$ with*

$$\inf_{\rho \in \mathfrak{X}} F_m(\rho^{\otimes l}, \mathcal{M}_l) \geq 1 - 2^{-lc_2} \tag{3.28}$$

with a constant $c_2 = c_2(\mathfrak{X}, \delta) > 0$,

$$\frac{1}{l}\log k_l \leq \sup_{\rho \in \mathfrak{X}} S(A|B,\rho) + \delta, \tag{3.29}$$

and

$$\frac{1}{l}\log D_l \leq \sup_{\rho \in \mathfrak{X}} I(A;E,\psi) + \delta, \tag{3.30}$$

where the quantum mutual information in (3.30) is evaluated on the AE marginal state of any purification ψ of ρ.

Proof of Proposition 17. The strategy of proof will be as follows. We decompose $\mathfrak{X}$ into a finite number of disjoint subsets $\mathfrak{X}_1, ..., \mathfrak{X}_N$, each containing only states with approximately equal entropy on the A-marginal system and combine an entropy estimating instrument on the A-system with a suitable merging scheme for each set $\mathfrak{X}_i$ according to Proposition 14. We fix $\delta > 0$, and assume, to simplify the argument, that

$$\tilde{s} := \sup_{\rho \in \mathfrak{X}} S(A|B,\rho) < 0 \tag{3.31}$$

holds (i.e. merging is possible without input entanglement resources for large enough blocklengths). Otherwise the argument below can be carried out using further input entanglement and wasting it before action of the protocol. We define $d := \dim \mathcal{H}_A$ and fix $\eta \in (0,1]$ to be determined later. Consider the sequence

$$s_0 := 0 < s_1 < ... < s_N := \log d, \quad s_i := s_{i-1} + \eta \text{ for each } 1 \leq i < N.$$

Define Intervals $I_1 := [s_0, s_1]$ and $I_i := (s_{i-1}, s_i]$ for $i = 2, ..., N$, which generate a decomposition of $\mathfrak{X}$ into disjoint sets $\mathfrak{X}_1, ..., \mathfrak{X}_N$ by definitions

$$\mathfrak{X}_i := \{\rho \in \mathfrak{X} : S(\rho_A) \in I_i\} \qquad (i \in [N]), \tag{3.32}$$

and set

$$\tilde{\mathfrak{X}}_i := \bigcup_{j \in n(i)} \mathfrak{X}_i (i \in [N])$$

where $n(i)$ is defined $n(i) := \{j \in [N] : |j - i| \leq 1\}$ for all i. In order to construct an entropy estimating instrument in the A marginal systems, we define an operation $\mathcal{P}_l^{(i)} \in \mathcal{C}^{\downarrow}(\mathcal{H}_{AB}^{\otimes l}, \mathcal{H}_{AB}^{\otimes l})$ by

$$\mathcal{P}_l^{(i)}(\cdot) := p_{i,l} \otimes \mathbb{1}_{\mathcal{H}_B^{\otimes l}}(\cdot) p_{i,l}^* \otimes \mathbb{1}_{\mathcal{H}_B^{\otimes l}} \text{with} \qquad p_{i,l} := \sum_{\substack{\lambda \in YF_{d,l}: \\ H(\overline{\lambda}) \in I_i}} P_{\lambda,l}$$

for each $i \in [N]$ using the notation from Theorem 15. Notice, that $p_1, ..., p_N$ form a projection valued measure on $\mathcal{H}_A^{\otimes l}$ due to Theorem 15.3. By construction, we have for each state $i \in [N]$, $\rho \in \mathfrak{X}_i$,

$$\sum_{j \in [N] \setminus n(i)} \operatorname{tr}(\mathcal{P}_l^{(j)}(\rho^{\otimes l})) = \sum_{j \in [N] \setminus n(i)} \operatorname{tr}(p_{i,l} \rho_A^{\otimes l}) \tag{3.33}$$

$$= \sum_{\substack{\lambda \in YF_{d,l}: \\ |H(\lambda) - S(\rho_A)| \geq \eta}} \operatorname{tr}(P_{\lambda,l} \rho_A^{\otimes l}) \tag{3.34}$$

$$\leq |YF_{d,l}| \cdot (l+1)^{d(d-1)/2}$$

$$\times \exp\left(-l \left(\min_{r: H(r) \in I_i} \min_{\substack{\lambda \in YF_{d,l}: \\ |H(\overline{\lambda}) - H(r)| \geq \eta}} D(\overline{\lambda} || r) \right) \right), \tag{3.35}$$

where (3.33) and (3.34) are valid due to construction and (3.35) follows from Theorem 15.1. Since the relative entropy term in the exponent on the r.h.s. of (3.35) is bounded away from zero for each fixed number $\eta > 0$ (a proof of this fact can be found in Appendix B), i.e.

$$\min_{r: H(r) \in I_i} \min_{\substack{\lambda \in YF_{d,l}: \\ |H(\overline{\lambda}) - H(r)| \geq \eta}} D(\overline{\lambda} || r) \geq 2c_3 \qquad (i \in [N]) \tag{3.36}$$

with a constant $c_3 = c_3(\eta) > 0$, and the functions outside the exponential term are growing polynomially for $l \to \infty$ (see Theorem 15.2) , we infer

$$\sum_{j\in[N]\setminus n(i)} \mathrm{tr}(\mathcal{P}_l^{(j)}(\rho^{\otimes l})) \geq 2^{-lc_3} \qquad (i \in [N]) \tag{3.37}$$

provided that l is large enough.
Define index sets $J := \{i : \mathfrak{X}_i \neq \varnothing\}$ and $\tilde{J} := \{i : \tilde{\mathfrak{X}}_i \neq \varnothing\}$. We know from Proposition 14, that for each sufficiently large l, we find an $(l, k_l, D_l^{(i)})$ $A \to B$ merging $\widetilde{\mathcal{M}}_l^{(i)}$ for each $i \in \tilde{J}$ such that

$$\inf_{\rho\in\tilde{\mathfrak{X}}_i} F_m(\rho^{\otimes l}, \widetilde{\mathcal{M}}_l^{(i)}) \geq 1 - 2^{-l\tilde{c}_i} \tag{3.38}$$

holds with a constant $\tilde{c}_i > 0$,

$$-\frac{1}{l}\log k_l \leq \sup_{\rho\in\tilde{\mathfrak{X}}_i} S(A|B,\rho) + \frac{\delta}{2} \tag{3.39}$$

and

$$\frac{1}{l}\log D_l^{(i)} \leq \sup_{\rho\in\tilde{\mathfrak{X}}_i} S(\rho_A) + \sup_{\rho\in\tilde{\mathfrak{X}}_i} S(A|B,\rho) + \frac{\delta}{2} \tag{3.40}$$

for the classical $A \to B$ communication rate. By construction of the sets $\tilde{\mathfrak{X}}_i$, $i \in \tilde{J}$, it also holds

$$I(A;E,\psi) \;=\; S(\rho_A) + S(A|B,\rho) \;\geq\; \sup_{\rho\in\tilde{\mathfrak{X}}_i} S(\rho_A) - 3\eta + S(A|B,\rho) \tag{3.41}$$

for each $\rho \in \tilde{\mathfrak{X}}_i$. Taking suprema over the set $\tilde{\mathfrak{X}}_i$ on both sides of the above inequality in combination with (3.40) leads us to the estimate

$$\frac{1}{l}\log D_l^{(i)} \;\leq\; \sup_{\rho\in\tilde{\mathfrak{X}}_i} I(A;E,\rho) + \frac{\delta}{2} + 3\eta \;\leq\; \sup_{\rho\in\mathfrak{X}} I(A;E,\rho) + \frac{\delta}{2} + 3\eta \tag{3.42}$$

for each $i \in [\tilde{J}]$. Combining the entropy estimating instrument $\{\mathcal{P}_l^{(j)}\}_{j=1}^N$ with the corresponding merging protocols, we define

$$\mathcal{M}_l(\cdot) := \sum_{i=1}^{N} \widetilde{\mathcal{M}}_l^{(i)} \circ \mathcal{P}_l^{(i)}(\cdot). \tag{3.43}$$

The maps $\widetilde{\mathcal{M}}_l^{(i)}$ are yet undefined for all numbers $i \in [N] \setminus \widetilde{J}$. Since they will not be relevant for the fidelity, they may be defined by any trivial local operations, with $D_l^{(i)} = 1$ for $i \in [N] \setminus \widetilde{J}$. Moreover, we assume, that the merging rate of $\mathcal{M}_l^{(i)}$ for each i is stuck to the the worst and each $\mathcal{M}_l^{(i)}$ outputs approximately the same maximally entangled resource output state ϕ_l. We can always achieve this by partial tracing and local unitaries, which do not further affect the

classical communication rates.
By inspection of the definition in (3.43) one readily verifies, that $\mathcal{M}_l$ is, in fact, an (l, k_l, D_l) $A \to B$ merging, with

$$D_l = \sum_{i \in \tilde{J}} D_l^{(i)} + |N - \widetilde{J}|, \tag{3.44}$$

and therefore, classical communication rate bounded by

$$\frac{1}{l} \log D_l = \frac{1}{l} \log \left(\sum_{i \in \tilde{J}} D_l^{(i)} + |N - \widetilde{J}| \right) \tag{3.45}$$

$$\leq \frac{1}{l} \log \left(N \cdot \max_{i \in [N]} D_l^{(i)} \right) \tag{3.46}$$

$$\leq \sup_{\rho \in \mathfrak{X}} I(A; E, \psi) + \frac{\delta}{2} + 3\eta + \frac{\log N}{l}. \tag{3.47}$$

It remains to show, that we achieve achieve merging fidelity one with $\{\mathcal{M}_l\}_{l \in \mathbb{N}}$ for each $\rho \in \mathfrak{X}$ with exponentially decreasing trade-offs for large enough blocklengths. Assume ρ is a member of $\mathfrak{X}_i$ for any index $i \in J$. Then, it holds

$$F_m(\rho^{\otimes l}, \mathcal{M}_l) \geq \sum_{j \in n(i)} F_m(\rho^{\otimes l}, \widetilde{\mathcal{M}}_l^{(j)} \circ \mathcal{P}_l^{(j)}) \tag{3.48}$$

$$= \sum_{j \in n(i)} F_m(\rho^{\otimes l}, \widetilde{\mathcal{M}}_l^{(j)} \circ \widetilde{\mathcal{P}}_l^{(i)}) - \sum_{j \in n(i)} \sum_{\substack{k \in n(i) \\ k \neq j}} F_m(\rho^{\otimes l}, \widetilde{\mathcal{M}}_l^{(j)} \circ \mathcal{P}_l^{(k)}). \tag{3.49}$$

The inequality above holds, because the merging fidelity is linear in the operation and all summands are nonnegative together with the definition of $\mathcal{M}_l$. The equality is by some zero-adding of terms an using the definition $\widetilde{\mathcal{P}}_l^{(i)} := \sum_{j \in n(i)} \mathcal{P}_l^{(j)}$ together with linearity of the merging fidelity in the operation again. We bound the terms in (3.49) separately. Beginning with the second term, we notice, that the fidelity is homogeneous in its inputs and bounded by one for states, it holds

$$F(\widetilde{\mathcal{M}}_l^{(j)} \circ \mathcal{P}_l^{(k)} \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes n}}(\psi_l), \phi_l \otimes \psi_l') \ \leq \ \mathrm{tr}(\mathcal{P}_l^{(k)}(\rho_A^{\otimes n})). \tag{3.50}$$

Summing up the bounds in (3.50), rearranging the summands and using the definition of $\widetilde{\mathcal{P}}_l^{(i)}$, we obtain the bound

$$\sum_{j \in n(i)} \sum_{\substack{k \in n(i) \\ k \neq j}} F_m(\rho^{\otimes l}, \widetilde{\mathcal{M}}_l^{(j)} \circ \mathcal{P}_l^{(k)}) \leq \sum_{j \in n(i)} \sum_{\substack{k \in n(i) \\ k \neq j}} \mathrm{tr}(\mathcal{P}_l^{(k)}(\rho^{\otimes n})) \tag{3.51}$$

$$= (|n(i)| - 1) \, \mathrm{tr}(\widetilde{\mathcal{P}}_l^{(i)}(\rho^{\otimes l})) \tag{3.52}$$

$$\leq |n(i)| - 1. \tag{3.53}$$

To bound the first terms in (3.49), we use the well-known relation

$$F(a, b) \geq \mathrm{tr}(a) - \|a - b\|_1 \tag{3.54}$$

between fidelity and trace norm holding for any two matrices $a, b \geq 0$, $\mathrm{tr}(b) = 1$ on a Hilbert space. It then holds, for each $j \in n(i)$,

$$F_m(\rho^{\otimes l}, \widetilde{\mathcal{M}}_l^{(j)} \circ \widetilde{\mathcal{P}}_l^{(i)}) \;\geq\; \mathrm{tr}(\widetilde{\mathcal{P}}_l^{(i)}(\rho^{\otimes l})) - \|\widetilde{\mathcal{M}}_l^{(j)} \circ \widetilde{\mathcal{P}}_l^{(i)} \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_l) - \phi_l \otimes \psi_l'\|_1. \tag{3.55}$$

For the second term in (3.55) it holds by zero adding, triangle inequality and monotonicity of the trace norm under action of partial traces

$$\begin{aligned}\|\widetilde{\mathcal{M}}_l^{(j)} \circ \widetilde{\mathcal{P}}_l^{(i)} \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_l) - \phi_l \otimes \psi_l'\|_1 \;\leq\; & \|\widetilde{\mathcal{M}}_l^{(j)} \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_l), \phi_l \otimes (\psi')^{\otimes l}\|_1 \\ & + \|\widetilde{\mathcal{P}}_l^{(i)}(\rho^{\otimes l}) - \rho^{\otimes l}\|_1. \end{aligned} \tag{3.56}$$

We further yield the bound

$$\|\widetilde{\mathcal{M}}_l^{(j)} \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_l) - \phi_l \otimes \psi_l'\|_1 \leq 2\left(1 - F(\widetilde{\mathcal{M}}_l^{(j)} \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_l), \phi_l \otimes \psi_l')\right)^{\frac{1}{2}} \tag{3.57}$$

$$\leq 2 \cdot 2^{-l\frac{\tilde{c}_i}{2}} \tag{3.58}$$

by (1.1) together with (3.38), and

$$\|\widetilde{\mathcal{P}}_l^{(i)}(\rho^{\otimes l}) - \rho^{\otimes l}\|_1 \;\leq\; 2\sqrt{1 - \mathrm{tr}(\widetilde{\mathcal{P}}_l^{(i)}(\rho^{\otimes l}))} \;\leq\; 2 \cdot 2^{-l\frac{c_3}{2}}, \tag{3.59}$$

where the first inequality is by Lemma 16, and the second inequality is valid due to the bound in (3.37) along with the fact, that (because $p_{1,l}, ..., p_{N,l}$ is a resolution of the identity into pairwise orthogonal projections)

$$1 - \mathrm{tr}(\widetilde{P}^{(i)}(\rho^{\otimes l})) \;=\; \mathrm{tr}\left(\left(\mathrm{id}_{\mathcal{H}_{AB}^{\otimes l}} - \widetilde{\mathcal{P}}_l^{(i)}\right)(\rho^{\otimes l})\right) \;=\; \sum_{j \in [N] \setminus n(i)} \mathrm{tr}(\mathcal{P}_l^{(j)}(\rho^{\otimes l}))$$

holds. We define the constant c_4 by $c_4 := \min\{\tilde{c}_1, ..., \tilde{c}_N, c_3\}$. Combining (3.55) with (3.56)-(3.59) leads us to the estimate

$$F_m(\rho^{\otimes l} \widetilde{\mathcal{M}}_l^{(j)} \circ \widetilde{\mathcal{P}}_l^{(i)}) \geq \mathrm{tr}(\widetilde{\mathcal{P}}_l^{(i)}(\rho^{\otimes l})) - 4 \cdot 2^{-l\frac{c_4}{2}} \tag{3.60}$$

$$\geq 1 - 5 \cdot 2^{-l\frac{c_4}{2}} \tag{3.61}$$

for each $j \in n(i)$, where the last of the above inequalities, again is by the bound in (3.37). By inserting the bounds given in (3.53) and (3.61) into (3.49), we yield

$$F_m(\rho^{\otimes l}, \mathcal{M}_l) \geq |n(i)|(1 - 5 \cdot 2^{-l\frac{c_4}{2}}) - (|n(i)| - 1) \tag{3.62}$$

$$\geq 1 - 5|n(i)| \cdot 2^{-l\frac{c_4}{2}} \tag{3.63}$$

$$\geq 1 - 15 \cdot 2^{-l\frac{c_4}{2}}. \tag{3.64}$$

If we now choose η small enough and assume l_0 large enough, to suffice

$$3\eta + \frac{\log N}{l_0} \leq \delta, \tag{3.65}$$

(3.39), (3.47), and (3.64) show, that $\mathcal{M}_l$ has the desired properties for each $l > l_0$.
The assertion can be proven for the remaining case $\tilde{s} \geq 0$ by considering a compound set $\{\rho \otimes \phi_0 : \rho \in \mathfrak{X}\}$ with a maximally entangled state ϕ_0 having Schmidt rank large enough to ensure $\sup_{\rho \in \mathfrak{X}} S(A|B, \rho \otimes \phi_0) < 0$ and repeat the argument given above for the first case (note, that $I(A; E, \rho \otimes \phi_0) = I(A; E, \rho)$ holds for each state $\rho \in \mathfrak{X}$). □

3.2.2 Converse theorem

To complete the proof of Theorem 7, the following inclusion relation has to be shown.

Theorem 18. *Let $\mathfrak{X} \subset \mathcal{S}(\mathcal{H}_{AB})$ be a set of density matrices. It holds*

$$\mathfrak{M}(\mathfrak{X}) \subset \left\{ R_q \geq \sup_{\rho \in \mathfrak{X}} S(A|B,\rho) \wedge R_c \geq \sup_{\rho \in \mathfrak{X}} I(A;E,\psi) \right\} \tag{3.66}$$

where $I(A;E,\psi)$ is the quantum mutual information between the A and E systems in a purification of ρ for each $\rho \in \mathfrak{X}$.

Notice, that the inclusion

$$\mathfrak{M}_{\rightarrow}(\mathfrak{X}) \subset \bigcap_{\rho \in \mathfrak{X}} \mathfrak{M}_{\rightarrow}(\{\rho\}) \tag{3.67}$$

is obvious from the definition of $\mathfrak{M}_{\rightarrow}$. In [HOW07] was given a full characterization of the set $\mathfrak{M}_{\rightarrow}(\{\rho\})$ in each case. Taking the proof for the converse regarding the entanglement consumption given therein for granted, we provide a new upper bound for the classical forward communication cost of quantum state merging in case of a memoryless quantum source with perfectly known generating state. The proof given in [HOW07] to lower-bound the demanded classical communication rate for sucessful quantum state merging was formulated using assertions formulated on the meta-level of the so-called resource framework for quantum Shannon theory, which was introduced in [DHW08].
In Proposition 19 below, we provide a more elementary proof of this result with a bound which seems somewhat stronger as the one given in [HOW07].

Proposition 19 (cf. Ref. [HOW07], Theorem 8)**.** *Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$ be a bipartite state with purification ψ_{ABE} on a space $\mathcal{H}_{ABE}$ and $\epsilon \in (0,1)$. If $\mathcal{M}(\cdot) := \sum_{k=1}^{D} \mathcal{A}_k \otimes \mathcal{B}_k(\cdot)$ is an $A \rightarrow B$ one-way LOCC such that*

$$F(\mathcal{M} \otimes id_{\mathcal{H}_E^{\otimes l}}(\phi_K \otimes \psi_{ABE}^{\otimes l}), \phi_L \otimes \psi_{B'BE}^{\otimes l}) \geq 1 - \epsilon \tag{3.68}$$

holds with maximally entangled states ϕ_K, ϕ_L of Schmidt rank K resp. L, then

$$\frac{1}{l}\log(D) \geq I(A;E,\rho_{AE}) - 6\sqrt{\epsilon}\left(\frac{1}{l}\log(KL) + \log\dim\mathcal{H}_{AB}\right) - 3\eta(2\sqrt{\epsilon}) \tag{3.69}$$

holds, where the function η is defined on $[0,1]$ by

$$\eta(x) := \begin{cases} -x\log x & 0 < x \leq \frac{1}{e} \\ \frac{\log e}{e} & \frac{1}{e} < x \leq 1 \end{cases} \tag{3.70}$$

and $\eta(0) := 0$.

Proof. The proof is inspired by ideas from Ref. [GPW05]. Fix $\epsilon \in (0,1)$ and $l \in \mathbb{N}$. Let $\phi_K \in \mathcal{K}^0_{AB}$ and $\phi_L \in \mathcal{K}^1_{AB}$ maximally entangled input resp. output states of the protocol such that with notations

$$\psi_0 := \phi_K \otimes \psi^{\otimes l}_{ABE}, \text{ and } \quad \psi_1 := \phi_L \otimes \psi^{\otimes l}_{B'BE}$$

eq. (3.68) reads

$$F(\mathcal{M} \otimes id_{\mathcal{H}^{\otimes l}_E}(\psi_0), \psi_1) \geq 1 - \epsilon. \tag{3.71}$$

We use the abbreviations $\mathcal{H}^0_{BE} := \mathcal{K}^0_B \otimes \mathcal{H}^{\otimes l}_{BE}$, $p_k := \mathrm{tr}(\mathcal{A}_k \otimes id_{\mathcal{H}^0_{BE}}(\psi_0))$ for $k \in [D]$, and $T = \{k \in [D] : p_k \neq 0\}$. It is well known, that the von Neumann entropy is an almost convex function, i.e. for a state $\overline{\rho}$ defined as a mixture $\overline{\rho} := \sum_{i=1}^N p_i \rho_i$ of quantum states,

$$S(\overline{\rho}) \leq H(p_1, ..., p_N) + \sum_{i=1}^N p_i S(\rho_i)$$

holds, where $H(p_1, ..., p_N)$ is the Shannon entropy of the probability distribution on $[N]$ given by $p_1, ..., p_N$ (for a proof, see [NC00], Theorem 11.10). Using this fact, we obtain the lower bound

$$\begin{aligned} \log D &\geq H(p_1, ..., p_D) \\ &\geq S\left(\sum_{k\in T} \mathcal{A}_k \otimes id_{\mathcal{H}^0_{BE}}(\psi_0)\right) - \sum_{k\in T} p_k S\left(\frac{1}{p_k}\mathcal{A}_k \otimes id_{\mathcal{H}^0_{BE}}(\psi_0)\right) \end{aligned} \tag{3.72}$$

on $\log D$. We separately bound the terms on the r.h.s. of eq. (3.72). With definitions $\pi_{K,A} := \mathrm{tr}_{\mathcal{K}^0_B}(\phi_K)$, $\pi_{K,B} := \mathrm{tr}_{\mathcal{K}^0_A}(\phi_K)$ and $\pi_{L,A} := \mathrm{tr}_{\mathcal{K}^1_B}(\phi_L)$ (these are maximally mixed states of rank K resp. L) and $\mathcal{A}(\cdot) := \sum_{k\in T} \mathcal{A}_k(\cdot)$, we obtain

$$S\left(\sum_{k\in T} \mathcal{A}_k \otimes id_{\mathcal{H}^0_{BE}}(\psi_0)\right) \geq S(\pi_{K,B} \otimes \rho^{\otimes l}_{BE}) - S(\mathcal{A}(\pi_{K,A} \otimes \rho^{\otimes l}_A)) \tag{3.73}$$

$$\geq \log K + lS(\rho_{BE}) - \log L - \Delta_1(\epsilon) \tag{3.74}$$

$$= \log \frac{K}{L} - lS(\rho_A) - \Delta_1(\epsilon) \tag{3.75}$$

where $\Delta_1(\cdot) := 2\sqrt{\cdot}\log(L) + \eta(2\sqrt{\cdot})$. Here eq. (3.73) is by the Araki-Lieb inequality [AL70], and eq. (3.75) is due to the fact that $S(\rho_A) = S(\rho_{BE})$ holds. Eq. (3.74) is justified as follows. Using (1.1) together with monotonicity of the fidelity under partial traces, (3.71) implies

$$\|\mathcal{A}(\pi_{K,A} \otimes \rho^{\otimes l}_A) - \pi_{L,A}\|_1 \leq 2\sqrt{\epsilon}. \tag{3.76}$$

This, via application of Fannes' inequality leads us to

$$S(\mathcal{A}(\pi_{K,A} \otimes \rho^{\otimes l}_A)) \leq S(\pi_{L,A}) - 2\sqrt{\epsilon}\log L - \eta(2\sqrt{\epsilon}), \tag{3.77}$$

where η is the function defined in (3.70). To bound the second term on the r.h.s. of (3.72), we use Stinespring extensions of the individual trace decreasing c.p. maps which constitute $\mathcal{M}$. Let for each $k \in [D]$,

$$v_k : \mathcal{K}_A^0 \otimes \mathcal{H}_A^{\otimes l} \to \mathcal{K}_A^1 \otimes \mathcal{H}_{C'}$$

be a Stinespring extension of $\mathcal{A}_k$ and

$$u_k : \mathcal{K}_B^0 \otimes \mathcal{H}_B^{\otimes l} \to \mathcal{K}_B^1 \otimes \mathcal{H}_{B'B}^{\otimes l} \otimes \mathcal{H}_{C''} \tag{3.78}$$

be a Stinespring extension of $\mathcal{B}_k$. Here $\mathcal{H}_{C'}$ is a Hilbert space associated to A and $\mathcal{H}_{C''}$ belongs to B. We fix notations $\mathcal{V}_k(\cdot) := v_k(\cdot)v_k^*$ and $\mathcal{U}_k := u_k(\cdot)u_k^*$ and denote the normalized outputs of these extensions by

$$\gamma_k := \frac{1}{p_k}\mathcal{V}_k \otimes \mathcal{U}_k \otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_0) \tag{3.79}$$

for every $k \in T$. Note that $\mathcal{V}_1, ..., \mathcal{V}_D$ are trace decreasing, while $\mathcal{U}_1, ..., \mathcal{U}_D$ are channels. For every $k \in T$, we have

$$\begin{aligned} S\left(\tfrac{1}{p_k}\mathcal{A}_k \otimes id_{\mathcal{H}_{BE}^0}(\psi_0)\right) &= S\left(\tfrac{1}{p_k}\mathrm{tr}_{\mathcal{H}_{C'}}\mathcal{V}_k \otimes id_{\mathcal{H}_{BE}^0}(\psi_0)\right) \\ &= S\left(\tfrac{1}{p_k}\mathrm{tr}_{\mathcal{H}_{C'}}\mathcal{V}_k \otimes \mathcal{U}_k \otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_0)\right) \\ &= S(\mathrm{tr}_{\mathcal{H}_{C'}}\gamma_k), \end{aligned} \tag{3.80}$$

where the second equality is by the fact that u_k is an isometry and consequently the action of $\mathcal{U}_k$ does does not change the entropy. Note, that (3.71) implies, because fidelity is linear in the first input here, existence of a positive number c_k for every $k \in T$, such that

$$F\left(\frac{1}{p_k}\mathcal{A}_k \otimes \mathcal{B}_k \otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_0), \psi_1\right) = 1 - c_k \tag{3.81}$$

and $\sum_{k\in T} p_k c_k \leq \epsilon$ hold. Because γ_k is a purification of $\frac{1}{p_k}\mathcal{A}_k \otimes \mathcal{B}_k \otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_0)$ and ψ_1 is already pure, Uhlmann's theorem [Uhl76] (cf. [Joz94] for a finite dimensional version) ensures existence of a pure state φ_k on $\mathcal{H}_{C'} \otimes \mathcal{H}_{C''}$ with

$$\begin{aligned} F(\gamma_k, \psi_1 \otimes \varphi_k) &= \max\{|\langle \gamma_k, \sigma\rangle|^2 : \sigma \text{ purification of } \psi_0 \text{ on } \mathcal{K}_{AB}^1 \otimes \mathcal{H}_{B'BE}^{\otimes l} \otimes \mathcal{H}_{C'} \otimes \mathcal{H}_{C''}\} \\ &= F\left(\frac{1}{p_k}\mathcal{A}_k \otimes \mathcal{B}_k \otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_0), \psi_1\right) \end{aligned} \tag{3.82}$$

for every $k \in T$. From eqns. (3.81) and (3.82) we conclude, using (1.1),

$$\|\gamma_k - \psi_1 \otimes \varphi_k\|_1 \leq 2\sqrt{c_k}, \tag{3.83}$$

which implies, again via Fannes' inequality and monotonicity of the trace distance under partial tracing

$$\begin{aligned} S(\mathrm{tr}_{\mathcal{H}_{C'}}\gamma_k) &\leq S(\psi_1 \otimes \mathrm{tr}_{\mathcal{H}_{C'}}\varphi_k) + \Delta_2(c_k) \\ &\leq S(\mathrm{tr}_{\mathcal{H}_{C'}}\varphi_k) + \Delta_2(c_k) \end{aligned} \tag{3.84}$$

where $\Delta_2(\cdot) = 2\sqrt{\cdot}\log(\dim\mathcal{H}_{AB}^2 \dim\mathcal{H}_{C''}) + \eta(2\sqrt{\cdot})$. Consequently, we have

$$\sum_{k\in T} p_k S\left(\frac{1}{p_k}\mathcal{A}_k \otimes id_{\mathcal{H}_{BE}^0}(\psi_0)\right) = \sum_{k\in T} p_k S(\mathrm{tr}_{\mathcal{H}_{C'}}\gamma_k) \leq \sum_{k\in T} p_k S(\mathrm{tr}_{\mathcal{H}_{C'}}\varphi_k) + \Delta_2(\epsilon). \tag{3.85}$$

The above equality is by (3.80), the inequality follows by (3.84) together with the fact, that that Δ_2 is monotone and concave (see the definition of η in (3.70)). It remains to bound $\sum_{k\in T} p_k S(\mathrm{tr}_{\mathcal{H}_{C'}}\varphi_k)$. Abbreviating $\mathcal{H}_{AE}^1 := \mathcal{K}_A^1 \otimes \mathcal{H}_E^{\otimes l} \otimes \mathcal{H}_{C'}$, an argument very similar to the one above gives (again via (3.83) and an application of Fannes' inequality) the bound

$$\begin{aligned} S(\mathrm{tr}_{\mathcal{H}_{AE}^1}(\gamma_k)) &\geq S(\mathrm{tr}_{\mathcal{H}_{AE}^1}(\psi_1 \otimes \varphi_k)) - \Delta_3(c_k) \\ &= S(\pi_{L,B} \otimes \rho_{B'B}^{\otimes l} \otimes \mathrm{tr}_{\mathcal{H}_{C'}}\varphi_k) - \Delta_3(c_k) \end{aligned} \tag{3.86}$$

with the function $\Delta_3(\cdot) := 2\sqrt{\cdot}(\log(K) + l\log(\dim\mathcal{H}_{AB}\cdot\dim\mathcal{H}_{C''})) + 2\eta(\sqrt{\cdot})$. Using monotonicity and concavity of Δ_3 together with (3.86), we obtain

$$\sum_{k\in T} p_k S(\mathrm{tr}_{\mathcal{H}_{AE}^1}(\gamma_k)) \geq \log(L) + lS(\rho_{AB}) + \sum_{k\in T} p_k S(\mathrm{tr}_{\mathcal{H}_{C'}}\varphi_k) - \Delta_3(\epsilon). \tag{3.87}$$

If we now look at $\sum_{k=1}^D \mathcal{V}_k \otimes \mathcal{U}_k \otimes id_{\mathcal{H}_E^{\otimes l}}(\cdot)$ as an one-way LOCC-channel with local operations on systems belonging to A and E on one side and B on the other side which 3 the pure input state ψ_0 to the state described by the pure state mixture $\sum_{k\in T} p_k\gamma_k$, we have

$$\begin{aligned} S(\pi_K \otimes \rho_B^{\otimes l}) &= S(\mathrm{tr}_{\mathcal{K}_A^0\otimes\mathcal{H}_{AE}^{\otimes l}}\psi_0) \\ &= S\left(\mathrm{tr}_{\mathcal{H}_{AE}^1}\left(\sum_{k=1}^D \mathcal{V}_k \otimes id_{\mathcal{H}_{BE}^0}(\psi_0)\right)\right) \\ &\geq \sum_{k\in T} p_k\, S\left(\frac{1}{p_k}\mathrm{tr}_{\mathcal{H}_{AE}^1}\mathcal{V}_k \otimes id_{\mathcal{H}_{BE}^0}(\psi_0)\right) \\ &= \sum_{k\in T} p_k\, S\left(\frac{1}{p_k}\mathrm{tr}_{\mathcal{H}_{AE}^1}\mathcal{V}_k \otimes \mathcal{U}_k \otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_0)\right) \qquad (3.88) \\ &= \sum_{k\in T} p_k\, S\left(\mathrm{tr}_{\mathcal{H}_{AE}^1}\gamma_k\right). \qquad (3.89) \end{aligned}$$

The second of the above equalities is due to the fact, that $\sum_{k=1}^D \mathcal{V}_k(\cdot)$ is trace preserving, the inequality is by concavity of the von Neumann entropy. Eq. (3.88) is because the von Neumann entropy is not changed by application of unitary channels in the input. The last equality is by the definitions introduced in (3.79). With (3.87), (3.89) and the equality $S(\rho_{AB}) = S(\rho_E)$, we obtain

$$S(\pi_K \otimes \rho_B^{\otimes l}) \geq \log(L) + lS(\rho_E) + \sum_{k\in T} p_k S(\mathrm{tr}_{\mathcal{H}_{C'}}\varphi_k) - \Delta_3(\epsilon). \tag{3.90}$$

Rearranging the terms in inequality (3.90) and using (3.85) leads to the bound

$$\sum_{k\in T} p_k S\left(\tfrac{1}{p_k}\mathcal{A}_k \otimes id_{\mathcal{H}^0_{BE}}(\psi_0)\right) \leq \log\frac{K}{L} + l(S(\rho_{AE}) - S(\rho_E)) + \Delta_2(\epsilon) + \Delta_3(\epsilon). \tag{3.91}$$

Here, we additionally used the fact, that $S(\rho_B) = S(\rho_{AE})$ holds. Combining the bounds from (3.74) and (3.91) with (3.72), we arrive at

$$\frac{1}{l}\log D \geq I(A;E,\rho_{AE}) - \frac{1}{l}(\Delta_1(\epsilon) + \Delta_2(\epsilon) + \Delta_3(\epsilon)). \tag{3.92}$$

In fact, we find Stinespring extensions on spaces $\mathcal{H}_{C'}$ and $\mathcal{H}_{C''}$ with

$$\dim\mathcal{H}_{C'} = K\cdot L\cdot\dim\mathcal{H}_A^l \tag{3.93}$$

$$\dim\mathcal{H}_{C''} = K\cdot L\cdot\dim\mathcal{H}_B^{2l}\dim\mathcal{H}_A^l. \tag{3.94}$$

Using the definition of Δ_1, Δ_2 and Δ_3 with the above dimensions, we conclude

$$\frac{1}{l}\log D \geq I(A;E,\rho_{AE}) - 6\sqrt{\epsilon}\left(\frac{\log KL}{l} + \log\dim\mathcal{H}_{AB}\right) - 3\eta(2\sqrt{\epsilon}), \tag{3.95}$$

which we aimed to prove. □

It is worth noticing, that the above proposition gives evidence to the claim, that maximal entanglement and forward classical communication are orthogonal communication resources for quantum state merging. The above lower bound on the classical forward communication rate is only weakly dependent on the entanglement rate. Therefore the quantum mutual information between the A and purifying E systems is the optimal classical communication rate for asymptotically perfect state merging procedures *regardless*, even when protocols with any suboptimal entanglement rates are applied.

Proof of Theorem 18. We define for each state $\rho\in\mathfrak{X}$ the sets

$$\tilde{\mathfrak{M}}_{\to,q}(\rho) := \{(R_q,R_c)\in\mathbb{R}\times\mathbb{R}^+ :\ R_q \geq S(A|B,\rho)\},\ \text{ and} \tag{3.96}$$

$$\tilde{\mathfrak{M}}_{\to,c}(\rho) := \{(R_q,R_c)\in\mathbb{R}\times\mathbb{R}^+ :\ R_c \geq I(A;E,\rho)\}, \tag{3.97}$$

we infer from the converse to the entanglement rate of state merging for a perfectly known memoryless quantum source from [HOW07], that

$$\mathfrak{M}_{\to}(\{\rho\}) \subset \tilde{\mathfrak{M}}_{\to,q}(\rho) \tag{3.98}$$

holds. Applying the bound from Proposition 19 above, for each sequence $\{\mathcal{M}_l\}$ of (l,k_l,D_l) mergings for ρ with

$$\lim_{l\to\infty} F_m(\rho^{\otimes l},\mathcal{M}_l) = 1 \tag{3.99}$$

the bound

$$\liminf_{l\to\infty} \frac{1}{l}\log D_l \geq I(A;E,\psi) \tag{3.100}$$

is valid. Consequently, we have

$$\mathfrak{M}_{\to}(\{\rho\}) \subset \tilde{\mathfrak{M}}_{\to,c}(\rho), \tag{3.101}$$

which implies

$$\mathfrak{M}_{\to}(\mathfrak{X}) \subset \bigcap_{\rho\in\mathfrak{X}} \mathfrak{M}_{\to}(\{\rho\}) \subset \bigcap_{\rho\in\mathfrak{X}} \left(\tilde{\mathfrak{M}}_{\to,q}(\rho) \cap \tilde{\mathfrak{M}}_{\to,c}(\rho)\right) \tag{3.102}$$

□

3.3 On quantum state merging for arbitrarily varying quantum sources

In this Section, we consider quantum state merging in case that the bipartite source A and B have to merge is an arbitrarily varying quantum source (AVQS). In the preceding section, we have determined the optimal entanglement as well as classical communication cost in case of a compound quantum source, and achieved these rates by protocols with merging fidelity going to one exponentially. One would expect, that applying Ahlswede's [Ahl+12] ingeniuous robustification technique, to state merging protocols for the compound source generated by $\overline{\mathrm{conv}(\mathfrak{X})}$ eventually leads to optimal protocols for merging the AVQS generated by a set $\mathfrak{X}$, to prove

$$C^{AV}_{m,\to}(\mathfrak{X}) = C_{m,\to}(\mathrm{conv}(\mathfrak{X})) = \sup_{\rho\in\mathrm{conv}(\mathfrak{X})} S(A|B,\rho). \tag{3.103}$$

Indeed, it seems possible, to prove the relation

$$C^{AV}_{m,\to}(\mathfrak{X}) \leq C_{m,\to}(\mathrm{conv}(\mathfrak{X})) \tag{3.104}$$

using Ahlswede's elimination and derandomization techniques (at least if the AVQS is generated by a finite set of states). We do not carry out the argument here. Instead, we provide a simple counterexample to the relation in (3.103).
Consider a finite set $\hat{\mathfrak{X}} := \{\rho_s\}_{s=1}^N$ of bipartite states on a Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, which is generated by unitaries in the following sense. Let $\rho_1 \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, where we assume $S(A|B,\rho_1) < 0$ and $\dim \mathcal{H}_A \geq N \cdot \dim \mathrm{supp}(\rho_{A,1})$, $U_1 = \mathbb{1}_{\mathcal{H}_A}$ and $U_2, ..., U_N$ unitaries on $\mathcal{H}_A$ such that with the definitions

$$\rho_s := U_s \otimes \mathbb{1}_{\mathcal{H}_B}(\rho_1) U_s^* \otimes \mathbb{1}_{\mathcal{H}_B} \qquad (s \in [N]) \tag{3.105}$$

the supports of the A-marginals are pairwise orthogonal, i.e.

$$\mathrm{supp}(\rho_{A,s}) \perp \mathrm{supp}(\rho_{A,s'}) \qquad (s, s' \in [N], s' \neq s). \tag{3.106}$$

Note, that our definitions also imply the relations $\rho_{B,s} = \rho_{B,1}$ $(s \in [N])$ and

$$\operatorname{supp}(\rho_s) \perp \operatorname{supp}(\rho_{s'}) \qquad (s, s' \in [N], s \neq s'). \tag{3.107}$$

In the following we show, that sets constructed in the above described manner are counterexamples to (3.103) if $N > 1$.

Example 20. *For the AVQS generated by $\hat{\mathfrak{X}} := \{\rho_s\}_{s=1}^N$, it holds*

$$C_{m,\to}^{AV}(\hat{\mathfrak{X}}) = C_{m\to}\left(\operatorname{conv}(\hat{\mathfrak{X}})\right) - \log N. \tag{3.108}$$

The classical $A \to B$ communication cost for merging of the AVQS $\hat{\mathfrak{X}}$ is upper bounded by

$$\sup_{\sigma \in \operatorname{conv}(\hat{\mathfrak{X}})} I(A; E, \sigma) - \log N, \tag{3.109}$$

where $\rho_p := \sum_{s=1}^N p(s)\rho_s$ for each $p \in \mathfrak{P}([N])$.

Proof of Example 20. Before we prove the claims made in the example, we briefly sketch the argument. Since the A marginals are supported on pairwise orthogonal subspaces, A can perfectly detect, given a block of l outputs of the AVQS, which of the $s^l \in \mathbf{S}^l$ is actually realized. In this way, A obtains state knowledge which helps to achieve the desired rates.
We introduce unitary channels $\mathcal{V}_{A,1}, ..., \mathcal{V}_{A,N}$ and $\mathcal{V}_{B',1}, ..., \mathcal{V}_{B',N}$ where we define $\mathcal{V}_{A,s}(\cdot) := U_s(\cdot)U_s^*$ with the unitaries from (3.105) and consider $\mathcal{V}_{B',s}$ to be the corresponding unitary channel on the space $\mathcal{H}_{B'}$ for each $s \in [N]$. For given blocklength l, we define unitary channels

$$\mathcal{V}_{A,s^l}(\cdot) := \mathcal{V}_{A,s_1} \otimes ... \otimes \mathcal{V}_{A,s_l} \quad \text{and} \quad \mathcal{V}_{B',s^l}(\cdot) := \mathcal{V}_{B',s_1} \otimes ... \otimes \mathcal{V}_{B',s_l} \tag{3.110}$$

for each $s^l = (s_1, ..., s_l) \in \mathbf{S}^l$ accordingly. Thus, the definitions in (3.105) imply

$$\rho_{s^l} = \mathcal{V}_{A,s^l} \otimes \mathrm{id}_{\mathcal{H}_B^{\otimes l}}(\rho_1^{\otimes l}). \qquad (s^l \in [N]^l)$$

Using the projection P_s onto the support of $\rho_{A,s}$ for each $s \in [N]$, we define a quantum instrument

$$\hat{\mathcal{A}} := \{\hat{\mathcal{A}}_s\}_{s=1}^N$$

with $\hat{\mathcal{A}}_s(\cdot) := \mathcal{U}_{A,s} \circ P_s(\cdot)P_s^*$ for each $s \in [N]$, which implies

$$\hat{\mathcal{A}}_{s'} \otimes \mathrm{id}_{\mathcal{H}_B}(\rho_s) = \delta_{ss'}\rho_1 \qquad (s \in [N]). \tag{3.111}$$

It is known from Ref. [HOW07], that for each $\delta > 0$ and sufficiently large blocklength l, there exists an $(l, k_l, \tilde{D}_l)$ $A \to B$ merging $\tilde{\mathcal{M}}_l$ such that

$$F(\tilde{\mathcal{M}}_l \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes n}}(\psi_1^{\otimes l}), \phi_l \otimes \psi_1'^{\otimes l}) \geq 1 - 2^{-lc} \tag{3.112}$$

holds with a constant $c > 0$, where ψ_1 is a purification of ρ_1 and ϕ_l a maximally entangled state shared by A and B with

$$-\frac{1}{l}\log \mathrm{sr}(\phi_l) \leq S(A|B,\rho_1) + \delta \tag{3.113}$$

and where for the classical communication rate

$$\frac{1}{l}\log \tilde{D}_l \leq I(A;E,\rho_1) + \delta \tag{3.114}$$

holds. We combine the instrument $\hat{\mathcal{A}}$ and the unitary channels from (3.110) with $\tilde{\mathcal{M}}_l$ to build a merging LOCC $\mathcal{M}_l$ suitable for merging the AVQS generated by $\hat{\mathfrak{X}}$ and define

$$\mathcal{M}_l := \sum_{s^l \in [N]^l} (\mathcal{V}_{B',s^l} \otimes \mathrm{id}_{\mathcal{H}_B^{\otimes l}}) \circ \tilde{\mathcal{M}} \circ (\hat{\mathcal{A}}_{s^l} \otimes \mathrm{id}_{\mathcal{H}_B^{\otimes l}}).$$

Clearly, $\mathcal{M}_l$ is an $A \to B$ LOCC channel. Explicitly, inspection of the above definition shows, that $\mathcal{M}_l$ is an (l, k_l, D_l) $A \to B$ merging where one of

$$D_l = \tilde{D}_l \cdot N^l \tag{3.115}$$

different classical messages has to be communicated within action of $\mathcal{M}_l$. Moreover, for each $s^l \in [N]^l$, it holds

$$\begin{aligned}
&F(\mathcal{M}_l \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_{s^l}), \phi_l \otimes \psi'_{s^l}) \\
&\overset{(a)}{=} \sum_{m^l \in [N]^l} F((\mathcal{V}_{B',m^l} \otimes \mathrm{id}_{\mathcal{H}_B^{\otimes l}}) \circ \tilde{\mathcal{M}}_l \circ (\hat{\mathcal{A}}_{m^l} \otimes \mathrm{id}_{\mathcal{H}_B^{\otimes l}}) \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_{s^l}), \phi_l \otimes \psi'_{s^l}) \\
&\overset{(b)}{=} F(\tilde{\mathcal{M}}_l \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_1^{\otimes l}), \phi_l \otimes (\mathcal{V}^*_{B',s^l} \otimes \mathrm{id}_{\mathcal{H}_{BE}^{\otimes l}})(\psi'_{s^l})) \\
&\overset{(c)}{=} F(\tilde{\mathcal{M}}_l \otimes \mathrm{id}_{\mathcal{H}_E^{\otimes l}}(\psi_1^{\otimes l}), \phi_l \otimes (\psi'_1)^{\otimes l}) \\
&\overset{(d)}{\geq} 1 - 2^{-lc},
\end{aligned} \tag{3.116}$$

where (a) is the definition of $\mathcal{M}_l$ plus linearity of the fidelity in the first argument in the present situation, (b) is because

$$\hat{\mathcal{A}}_{m^l} \otimes \mathrm{id}_{\mathcal{H}_{BE}^{\otimes l}}(\psi_{s^l}) = \delta_{m^l s^l} \psi_1^{\otimes l} \tag{3.117}$$

holds implied by (3.111) together with the fact, that the fidelity is invariant under action of unitary channels applied simultaneously on both arguments. Equality (c) follows from (3.105), and (d) is by (3.112). It remains to evaluate the rates. It is well known, that for each ensemble $\{q(x), \rho_x\}_{x \in \mathbf{X}}$ of quantum states having pairwise orthogonal supports, it holds

$$S\left(\sum_{x \in \mathbf{X}} q(x)\rho_x\right) = \sum_{s \in \mathbf{X}} q(x) S(\rho_x) + H(q).$$

Thus, for each $p \in \mathfrak{P}([N])$, $\rho_p := \sum_{s\in[N]} p(s)\rho_s$ we yield

$$S(A|B,\rho_p) = S(A|B,\rho_1) + H(p)$$

and

$$I(A;E,\rho_p) = I(A;E,\rho_1) + 2H(p).$$

Taking maxima over all $p \in \mathfrak{P}([N])$ and rearranging equations, we arrive at

$$S(A|B,\rho_1) = \max_{p\in\mathfrak{P}([N])} S(A|B,\rho_p) - \log N \tag{3.118}$$

and

$$I(A;E,\rho_1) = \max_{p\in\mathfrak{P}([N])} I(A;E,\rho_p) - 2\log N. \tag{3.119}$$

Note, that

$$C_{m,\to}(\mathrm{conv}(\hat{\mathfrak{X}})) = \max_{p\in\mathfrak{P}([N])} S(A|B,\rho_p) \tag{3.120}$$

by Proposition 17. Combining (3.118) with (3.113) and (3.120) together with (3.116) shows, that

$$C_m^{AV}(\mathfrak{X}) \le C_m(\mathrm{conv}(\hat{\mathfrak{X}})) - \log N + \delta \tag{3.121}$$

holds. The converse is valid by the merging cost converse for single states [HOW07]. Moreover, by (3.119), our protocols have classical $A \to B$ classical communication rates with

$$\limsup_{l\to\infty} \frac{1}{l}\log D_l = \limsup_{l\to\infty} \frac{1}{l}\log(\tilde{D}_l \cdot N^l) \tag{3.122}$$

$$\le I(A;E,\rho_1) + \delta + \log N \tag{3.123}$$

$$= \max_{p\in\mathfrak{P}([N])} I(A;E,\rho_p) - \log N + \delta \tag{3.124}$$

where (3.122) follows from (3.115), (3.123) is by (3.114), and (3.124) is by (3.119). Since $\delta > 0$ was an arbitrary positive number, we are done. □

We will see in the next section, that applying a version of the robustification technique leads to optimal protocols, if entanglement distillation is considered.

4 One-way entanglement distillation under source uncertainties

4.1 Definitions and results

In this section, we set up the definitions used in this chapter. Since we are mainly interested in the optimal entanglement rates of one-way LOCC distillation protocols, we allow the users free but rate-bounded choice of the classical forward communication. For the definitions in this chapter, we consider $\mathfrak{X}$ to be an arbitrary set $\mathfrak{X} := \{\rho_s\}_{s\in\mathbf{S}} \subset \mathcal{S}(\mathcal{H}_{AB})$ of bipartite density matrices.

4.1.1 Compound quantum sources

Definition 21. *A non-negative number R is an* achievable $A \to B$ entanglement distillation rate for the compound quantum source $\mathfrak{X}$ with classical rate R_c, *if there exists a sequence $\{\mathcal{D}_l\}_{l\in\mathbb{N}}$ of $A \to B$ LOCC channels,*

$$\mathcal{D}_l = \sum_{m=1}^{M_l} \mathcal{A}_{m,l} \otimes \mathcal{B}_{m,l} \qquad (l \in \mathbb{N}) \tag{4.1}$$

such that the conditions

1. $\lim_{l\to\infty} \inf_{s\in\mathbf{S}} F(\mathcal{D}_l(\rho_s^{\otimes l}), \phi_l) = 1$
2. $\liminf_{l\to\infty} \frac{1}{l} \log \mathrm{sr}(\phi_l) \geq R$
3. $\limsup_{l\to\infty} \frac{1}{l} \log M_l \leq R_c$

are fulfilled, where ϕ_l is a maximally entangled state shared by A and B and $\mathrm{sr}(\phi_l)$ is its Schmidt rank for each $l \in \mathbb{N}$.

Definition 22. *The $A \to B$* entanglement distillation capacity for the compound source $\mathfrak{X}$ *is defined*

$$D_{\to}(\mathfrak{X}) := \sup\left\{ R : \begin{array}{l} R \text{ achievable } A \to B \text{ entanglement distillation rate forthe compound} \\ \text{source } \mathfrak{X} \text{ with some classical communication rate } R_c \end{array} \right\}.$$

To introduce some notation we use in this chapter, we state a theorem from Ref. [DW05], where the $A \to B$ entanglement distillation capacity $D_{\to}(\rho)$ of a memoryless bipartite quantum source with perfectly known density matrix ρ was considered.

Theorem 23 ([DW05], Theorem 3.4)**.** *Let ρ be a state on $\mathcal{H}_{AB}$. It holds*

$$D_{\to}(\rho) = \lim_{k\to\infty} \frac{1}{k} \sup_{\mathcal{T}\in\Theta_k} D^{(1)}(\rho^{\otimes k}, \mathcal{T}) \tag{4.2}$$

with

$$D^{(1)}_{\to}(\sigma, \mathcal{T}) := \sum_{\substack{j\in[J]:\\ \lambda_j(\sigma)\neq 0}} \lambda_j(\sigma)\, I_c(A\rangle B, \sigma_j), \tag{4.3}$$

where Θ_k is the set of finite-valued quantum instruments on A's site, i.e.

$$\Theta_k := \left\{ \{\mathcal{T}_j\}_{j=1}^J \subset \mathcal{C}^{\downarrow}(\mathcal{H}_A^{\otimes k}, \mathcal{K}_A) : \sum_{j=1}^J \mathcal{T}_j \in \mathcal{C}(\mathcal{H}_A^{\otimes k}, \mathcal{K}_A),\ J < \infty,\ \dim \mathcal{K}_A < \infty \right\}. \tag{4.4}$$

For each state σ and quantum instrument $\mathcal{T} := \{\mathcal{T}_j\}_{j=1}^J$ on A's site and definitions

$$\lambda_j(\sigma) := \mathrm{tr}(\mathcal{T}_j(\sigma_A)),\ \textit{and}\quad \sigma_j := \frac{1}{\lambda_j}(\sigma)(\mathcal{T}_j \otimes \mathrm{id}_{\mathcal{H}_B})(\sigma)$$

for each j with $\lambda_j(\sigma) \neq 0$.

Remark 24. *It is known from [DW05], that the limit in (4.2) exists for each state, and maximization over instruments in this formula is always realized by an instrument $\mathcal{T} = \{\mathcal{T}_j\}_{j=1}^J$ with $J \leq \dim \mathcal{H}_A^{2k}$ and the operation $\mathcal{T}_j$ described by a single Kraus operator for $1 \leq j \leq J$.*

In order to obtain a compact notation for the capacity functions arising in the entanglement distillation scenarios we consider in this paper, we introduce a one-way LOCC $\hat{\mathcal{T}} := \sum_{j=1}^J \mathcal{T}_j \otimes |e_j\rangle\langle e_j|$ for each instrument $\{\mathcal{T}_j\}_{j=1}^J$ with domain $\mathcal{H}_A$ and an orthonormal system $\{e_j\}_{j=1}^J$ in a suitable space $\mathcal{H}'_B \simeq \mathbb{C}^J$ assigned to B, it holds

$$D^{(1)}(\sigma, \mathcal{T}) = I_c(A\rangle BB', \hat{\mathcal{T}}(\sigma)) \tag{4.5}$$

in (4.3) for each given state σ.

Theorem 25. *Let $\mathfrak{Y} \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$.*

1. *It holds*

$$D_{\to}(\mathfrak{Y}) = \lim_{k\to\infty} \frac{1}{k} \sup_{\mathcal{T}\in\Theta_k} \inf_{\rho\in\mathfrak{Y}} D^{(1)}_{\to}(\rho^{\otimes k}, \mathcal{T}), \tag{4.6}$$

   where the set Θ_k is defined as in (4.4) for each $k \in \mathbb{N}$.

2. *The function in (4.6) behaves regular for compound sources in the following sense. If $\mathfrak{Y}, \mathfrak{Y}' \subset \mathcal{S}(\mathcal{H}_{AB})$ are two nonempty sets of bipartite states with $d_H(\mathfrak{Y}, \mathfrak{Y}') < \delta \leq \frac{1}{2}$, it holds*

$$|D_{\to}(\mathfrak{Y}) - D_{\to}(\mathfrak{Y}')| \leq \nu(\delta) \tag{4.7}$$

4.1.2 Arbitrarily varying quantum sources

Definition 26. *A non-negative number R is an* achievable $A \to B$ entanglement distillation rate for the AVQS generated by a set $\mathfrak{X}$ with classical rate R_c, *if there exists a sequence $\{\mathcal{D}_l\}_{l\in\mathbb{N}}$ of $A \to B$ LOCC channels,*

$$\mathcal{D}_l = \sum_{m=1}^{M_l} \mathcal{A}_{m,l} \otimes \mathcal{B}_{m,l} \qquad (l \in \mathbb{N}) \tag{4.8}$$

such that the conditions

1. $\lim\limits_{l\to\infty} \inf\limits_{s^l\in\mathbf{S}^l} F(\mathcal{D}_l(\rho_{s^l}), \phi_l) = 1$
2. $\liminf\limits_{l\to\infty} \frac{1}{l} \log \mathrm{sr}(\phi_l) \geq R$
3. $\limsup\limits_{l\to\infty} \frac{1}{l} \log M_l \leq R_c$

are fulfilled, where ϕ_l is a maximally entangled state shared by A and B for each $l \in \mathbb{N}$.

Definition 27. *The $A \to B$ entanglement distillation capacity for the AVQS generated by $\mathfrak{X}$ is defined*

$$D_{\to}^{AV}(\mathfrak{X}) := \sup\left\{ R : \begin{array}{l} R \text{ is an achievable } A \to B \text{ entanglement distillation rate for} \\ \text{the AVQS } \mathfrak{X} \text{ with some classical communication rate } R_c \end{array} \right\}.$$

Theorem 28. *Let $\mathfrak{X}$ be a set of states on $\mathcal{H}_A \otimes \mathcal{H}_B$. For the AVQS generated by $\mathfrak{X}$, it holds*

$$D_{\to}^{AV}(\mathfrak{X}) = D_{\to}(\mathrm{conv}(\mathfrak{X})) = \lim_{l\to\infty} \frac{1}{k} \sup_{\mathcal{T}\in\Theta_k} \inf_{\tau\in\mathrm{conv}(\mathfrak{X})} D_{\to}^{(1)}(\tau^{\otimes k}, \mathcal{T}) \tag{4.9}$$

with $D_{\to}^{(1)}$ being the function defined in (4.3), and maximization over instruments on A's systems.

4.2 Entanglement distillation from compound quantum sources - Proofs

The main purpose of this section is, to give a full proof to Theorem 28, which is done in Subsection 4.2.2. As a prerequisite, we demonstrate in Subsection 4.2.1, that the capacity function appearing in Theorem 28 is continuous. We close the section by applying the results obtained to give a new proof for achievability of the entanglement generating capacity for compound quantum channels first proven in [BBN09].

4.2.1 Continuity of entanglement distillation capacities

Continuity was shown for the capacity functions appearing in coding theorems of several quantum channel coding scenarios[LS08], here we state and prove uniform continuity for the entanglement distillation capacity functions.

Lemma 29. *Let $\mathfrak{Y}, \mathfrak{Y}' \subset \mathcal{S}(\mathcal{H}_X \otimes \mathcal{H}_Y)$ be two nonempty sets of bipartite states with Hausdorff distance $0 \leq d_H(\mathfrak{Y}, \mathfrak{Y}') < \epsilon \leq \frac{1}{2}$. It holds for each $k \in \mathbb{N}$ and c.p.t.p map $\mathcal{N}$ with domain $\mathcal{L}(\mathcal{H}_{XY}^{\otimes k})$*

$$\left| \inf_{\tau \in \mathfrak{Y}} I_c(X\rangle Y, \mathcal{N}(\tau^{\otimes k})) - \inf_{\sigma \in \mathfrak{Y}'} I_c(X\rangle Y, \mathcal{N}(\sigma^{\otimes k})) \right| \leq k\nu(\epsilon), \tag{4.10}$$

where the function ν is defined by $\nu(x) := 4x \log \dim \mathcal{H}_X + 2h(x)$ for $x \in (0\frac{1}{2})$ and h being the binary entropy $h(x) := -x \log x - (1-x)\log(1-x)$.

Proof. We show this assertion with sets containing only one state defined $\mathfrak{Y} := \{\tau\}, \mathfrak{Y}' := \{\sigma\}$. The general assertion in (4.10) follows directly by definition of the Hausdorff distance. The argument parallels the one given in Ref. [LS08], Theorem 6 for continuity of the the entropy exchange for channels. Introduce a state $\gamma_{k,n} := \tau^{\otimes n} \otimes \sigma^{\otimes(k-n)}$ for each $0 \leq n \leq k$. By assumption, it holds

$$\|\gamma_{k,n-1} - \gamma_{k,n}\|_1 \leq \epsilon \tag{4.11}$$

for each $0 < n \leq k$, which implies, via the Alicki-Fannes inequality [AF04] for the conditional von Neumann entropy

$$\left| I_c(X\rangle Y, \mathcal{N}(\gamma_{k,n-1})) - I_c(X\rangle Y, \mathcal{N}(\gamma_{k,n})) \right| \leq \nu(\epsilon) \tag{4.12}$$

for each $0 < n \leq k$ by (4.11) and monotonicity of the trace distance under action of $\mathcal{N}$. Further, it holds

$$\begin{aligned}
&\left| I_c(X\rangle Y, \mathcal{N}(\tau^{\otimes k})) - I_c(X\rangle Y, \mathcal{N}(\sigma^{\otimes k})) \right| && (4.13)\\
&= \left| I_c(X\rangle Y, \mathcal{N}(\gamma_{k,k})) - I_c(X\rangle Y, \mathcal{N}(\gamma_{k,0})) \right| && (4.14)\\
&= \left| \sum_{n=1}^{k} \left(I_c(X\rangle Y, \mathcal{N}(\gamma_{k,n-1})) - I_c(X\rangle Y, \mathcal{N}(\gamma_{k,n})) \right) \right| && (4.15)\\
&\leq \sum_{n=1}^{k} \left| I_c(X\rangle Y, \mathcal{N}(\gamma_{k,n-1})) - I_c(X\rangle Y, \mathcal{N}(\gamma_{k,n})) \right|, && (4.16)
\end{aligned}$$

where the first equality above is by definition, and the second by adding some zeros. Estimating each summand in (4.16) by (4.12) concludes the proof. □

Corollary 30. *The one-way entanglement distillation capacity $D_{\rightarrow}$ for memoryless sources with perfectly known source state in (4.2) is a uniformly continuous function (considering the trace distance). Explicitly, it holds for $\rho, \sigma \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with $\|\rho - \sigma\|_1 < \epsilon \leq \frac{1}{2}$, it holds*

$$|D_{\rightarrow}(\rho) - D_{\rightarrow}(\sigma)| \leq \nu(\epsilon). \tag{4.17}$$

4.2.2 Proof of the coding theorem for compound quantum sources

Lemma 31. *Let $\mathfrak{X} := \{\rho_s\}_{s\in\mathbf{S}} \subset \mathcal{S}(\mathcal{H}_{AB})$ be a set of bipartite states on $\mathcal{H}_{AB}$. Then*

$$D_{\rightarrow}(\mathfrak{X}) \geq -\sup_{s\in\mathbf{S}} S(A|B,\rho_s) \tag{4.18}$$

Proof. It suffices to consider the case of a set with $\sup_{s\in\mathbf{S}} S(A|B,\rho_s) < 0$, since rate 0 can always be achieved by using a trivial protocol which distills no entanglement at all. Let $\mathcal{M} := \sum_{k=1}^{D} \mathcal{A}_k \otimes \mathcal{U}_k$ be an L-merging for $\mathfrak{X}$ satisfying

$$\min_{1\leq i\leq N} F(\mathcal{M}\otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_{ABE,i}^{\otimes l}), \phi_l \otimes \psi_{B'BE,i}^{\otimes l}) \geq 1-\epsilon. \tag{4.19}$$

Then $\mathcal{T}(\cdot) := \sum_{k=1}^{D} \mathcal{A}_k \otimes \mathcal{R}_k(\cdot)$ with $\mathcal{R}_k := \mathrm{tr}_{\mathcal{H}_{B'BE}^{\otimes l}} \circ (\mathcal{U}_k \otimes id_{\mathcal{H}_E^{\otimes l}})$ for every k is a one-way entanglement distillation protocol for $\mathfrak{X}$ satisfying

$$F(\mathcal{T}(\rho_i^{\otimes l}), \phi_l) \;\geq\; F_m(\rho_s, \mathcal{M}) \;\geq\; 1-\epsilon. \tag{4.20}$$

for every $s \in \mathbf{S}$ where the first of the above inequalities is justified by the fact that taking partial traces cannot increase fidelity. From Proposition 14 we know, that we find for $\epsilon > 0$ and $l \in \mathbb{N}$ large enough an L_l-merging $\mathcal{M}_l$ for $\mathfrak{X}$ such that

$$L_l \geq \left\lfloor \exp\left(-l(\sup_{s\in\mathbf{S}} S(A|B,\rho_s) + \epsilon) \right) \right\rfloor \tag{4.21}$$

and

$$\inf_{s\in\mathbf{S}} F(\mathcal{M}_l \otimes id_{\mathcal{H}_E^{\otimes l}}(\psi_{ABE,s}^{\otimes l}), \phi_l \otimes \psi_{B'BE,s}^{\otimes l}) \geq 1 - 2^{-nc_4}. \tag{4.22}$$

holds with a constant $c_4 > 0$. Eqns. (4.20) and (4.23) give

$$\inf_{s\in\mathbf{S}} F(\mathcal{T}_l(\rho_s^{\otimes l}), \phi_l) \geq 1 - 2^{-nc_4}. \tag{4.23}$$

The achievability of $-\sup_{s\in\mathbf{S}} S(A|B,\rho_s)$ follows from (4.21) and (4.23). □

The above lemma provides the main building block for determining the one-way entanglement capacity for sets of states, which is done in the following theorem.

Proposition 32. *Let $\mathfrak{Y} \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a set of bipartite states. For each $k \in \mathbb{N}, \delta > 0$, there exists a number $l_0 = l_0(k,\delta)$ and a constant $c_5 = c_5(k,\delta,\mathfrak{X}) > 0$, such for each $l > l_0$, there exists an $A \rightarrow B$ LOCC $\mathcal{D}_l$ fulfilling*

$$\inf_{\rho\in\mathfrak{X}} F(\mathcal{D}_l(\rho^{\otimes l}), \phi_l) \geq 1 - 2^{-lc_5}, \tag{4.24}$$

where ϕ_l is a maximally entangled state shared by A and B with

$$\frac{1}{l}\log \mathrm{sr}(\phi_l) \geq \lim_{k\to\infty} \frac{1}{k} \sup_{\mathcal{T}\in\Theta_k} \inf_{\rho\in\mathfrak{X}} D_{\rightarrow}^{(1)}(\rho^{\otimes k}, \mathcal{T}) - \delta. \tag{4.25}$$

The function $D_{\rightarrow}^{(1)}$ is defined in (4.3), and Θ_k is defined as in (4.4) for each $k \in \mathbb{N}$.

Proof. Our proof parallels the one given in Ref. [DW05] for the single state case. However, for the direct part, we use Lemma 31 instead of the single state hashing bound. To prove achievability, let $\mathcal{T} := \{\mathcal{T}_j\}_{j=1}^J$ be any instrument on $\mathcal{H}_A$, $\mathcal{P} := \{\mathcal{P}_j\}_{j=1}^J$ a set of channels of the form

$$\mathcal{P}_j(\chi) := \chi \otimes |e_j\rangle\langle e_j| \tag{4.26}$$

for every $\chi \in \mathcal{S}(\mathcal{H}_B)$ and $1 \le j \le J$, where $e_1, ..., e_J$ are members of an orthonormal basis of a Hilbert space $\mathcal{H}_{B'}$ located at B's site. Define states

$$\tilde{\rho}_s := \sum_{j=1}^{J} \mathcal{T}_j \otimes \mathcal{P}_j(\rho_s) = \sum_{j,\lambda_j^{(s)}\neq 0} \lambda_j^{(s)} \rho_j^{(s)} \otimes |e_j\rangle\langle e_j|$$

for each $s \in \mathbf{S}$. These preprocessed states have conditional von Neumann entropy

$$S(A|BB', \tilde{\rho}_s) = \sum_{j:\lambda_j^{(s)}\neq 0} \lambda_j^{(s)} S(A|B, \rho_j^{(s)}).$$

Direct application of Lemma 31 proves the claim of the proposition. □

Proof of Theorem 25. Achievability follows directly from Proposition 32. The converse statement can be proven just by the same arguments as given in Ref. [DW05], we give the proof for convenience. We consider an arbitrary (l, k_l) one-way distillation protocol with rate R, given by a LOCC channel with $A \to B$ classical communication

$$\mathcal{T}(\cdot) := \sum_{j=1}^{J} \mathcal{T}_j \otimes \mathcal{R}_j(\cdot)$$

with $\mathcal{T}_j \in \mathcal{C}^{\downarrow}(\mathcal{H}_A^{\otimes l}, \mathcal{K})$ and $\mathcal{R}_j \in \mathcal{C}(\mathcal{H}_B^{\otimes l}, \mathcal{K})$, $1 \le j \le J$, such that for a given $\tau \in (0, \frac{1}{2})$

$$F(\mathcal{T}(\rho_s^{\otimes l}), \phi) \ge 1 - \tau \tag{4.27}$$

holds for all $s \in \mathbf{S}$, where ϕ is a maximally entangled state on $\mathcal{K} \otimes \mathcal{K}$ and $\dim \mathcal{K} = \lfloor 2^{lR} \rfloor$. We fix notations

$$\lambda_j^{(s)} := \mathrm{tr}(\mathcal{T}_j \otimes \mathcal{R}_j(\rho_s^{\otimes l})), \quad \text{and} \quad \omega_j^{(s)} := \frac{1}{\lambda_j^{(s)}} \mathcal{T}_j \otimes R_j(\rho_s^{\otimes l}),$$

$$\rho_j^{(s)} := \frac{1}{\lambda_j^{(s)}} \mathcal{T}_j \otimes id_{\mathcal{H}_B^{\otimes l}}(\rho_s^{\otimes l})$$

for each $s \in \mathbf{S}$, $j \in [J]$ with $\lambda_j^{(s)} \neq 0$. Application of $\mathcal{T}$ on ρ_s results in the state

$$\Omega^{(s)} := \sum_{j:\lambda_j^{(s)}\neq 0} \lambda_j^{(s)} \omega_j^{(s)}.$$

Using the relation from (1.1), (4.27) implies, that

$$\|\Omega^{(i)} - \phi\|_1 \leq 2\sqrt{\tau}$$

holds for all $i \in [N]$, which leads us to

$$|S(A|B,\Omega^{(i)}) - S(A|B,\phi)| \leq \epsilon \tag{4.28}$$

with $\epsilon := 2(2\sqrt{\tau}\log(\dim \mathcal{K}^2) + \eta(2\sqrt{\tau}))$ via twofold application of Fannes' inequality. Eq. (4.28) along with $S(A|B,\phi) = -l \cdot R$ implies

$$lR \leq -S(A|B,\Omega^{(s)}) + 4\sqrt{\tau}\cdot lR + 2\eta(2\sqrt{\tau}). \tag{4.29}$$

Moreover, we have

$$\begin{aligned} S(A|B,\Omega^{(s)}) &\geq \sum_{j:\lambda_j^{(s)}\neq 0} \lambda_j^{(s)} S(A|B,\omega_j^{(s)}) \\ &\geq \sum_{j:\lambda_j^{(s)}\neq 0} \lambda_j^{(s)} S(A|B,\rho_j^{(s)}), \end{aligned} \tag{4.30}$$

where the first inequality is by concavity of the map $\rho \mapsto S(A|B,\rho)$ for quantum states, the second is by application of the quantum data processing inequality. Combining (4.29) and (4.30), we obtain

$$\begin{aligned} lR &\leq -\sup_{s\in\mathbf{S}} \sum_{j:\lambda_j^{(s)}\neq 0} \lambda_j^{(s)} S(A|B,\rho_j^{(s)}) + 4\sqrt{\tau} l\cdot R + 2\eta(2\sqrt{\tau}) \\ &\leq -\min_{\mathcal{T}} \sup_{s\in\mathbf{S}} \sum_{j:\lambda_j^{(s)}\neq 0} \lambda_j^{(s)} S(A|B,\rho_j^{(s)}) + 4\sqrt{\tau} l\cdot R + 2\eta(2\sqrt{\tau}) \\ &\leq D^{(1)}(\mathfrak{X}^{\otimes l}) + 4\sqrt{\tau} l\cdot R + 2\eta(2\sqrt{\tau}) \end{aligned}$$

It remains to show validity of the inequality in (4.7). Assume $d_H(\mathfrak{Y},\mathfrak{Y}') < \delta \leq \frac{1}{2}$. Let $\tau > 0$ be an arbitrary but fixed number, and $\mathcal{Q}$ be an instrument with domain $\mathcal{L}(\mathcal{H}_A^{\otimes l})$, such that

$$\inf_{\rho\in\mathfrak{Y}} I_c(A\rangle BB', \hat{\mathcal{Q}}(\rho^{\otimes l})) \geq \sup_{\mathcal{T}\in\Theta_k} \inf_{\rho\in\mathfrak{Y}} I_c(A\rangle BB', \hat{\mathcal{T}}(\rho^{\otimes l})) - \tau \tag{4.31}$$

holds, where we used our notation from (4.5). Lemma 29 implies

$$\inf_{\rho\in\mathfrak{Y}} I_c(A\rangle BB', \hat{\mathcal{Q}}(\rho^{\otimes l})) \geq \inf_{\rho\in\mathfrak{Y}'} I_c(A\rangle BB', \hat{\mathcal{Q}}(\rho^{\otimes l})) - k\nu(\delta), \tag{4.32}$$

which, together with (4.31) implies

$$\sup_{\mathcal{T}\in\Theta_k} \inf_{\rho\in\mathfrak{Y}'} I_c(A\rangle BB', \hat{\mathcal{T}}(\rho^{\otimes l})) \geq \sup_{\mathcal{T}} \inf_{\rho\in\mathfrak{Y}} I_c(A\rangle BB', \hat{\mathcal{T}}(\rho^{\otimes l})) - \tau - k\nu(\delta). \tag{4.33}$$

Since the above line of reasoning also holds with $\mathfrak{Y},\mathfrak{Y}'$ interchanged and τ can be chosen arbitrarily small, we obtain

$$\left| \sup_{\mathcal{T}\in\Theta_k} \inf_{\rho\in\mathfrak{Y}} D^{(1)}(\rho^{\otimes k},\mathcal{T}) - \sup_{\mathcal{T}\in\theta_k} \inf_{\rho\in\mathfrak{Y}'} D^{(1)}(\rho^{\otimes k},\mathcal{T}) \right| \leq k\nu(\delta). \tag{4.34}$$

The above inequality together with the first assertion of the corollary proves the second one. □

4.2.3 An application: Universal entanglement generation codes for compound quantum channels

In this section, we apply the results obtained so far to give another proof for the direct part of the coding theorem for entanglement generation over compound channels, which was originally given in Ref. [BBN09], Theorem 13. We first recall some definitions from Ref. [BBN09]. Let $\mathfrak{I}$ be a compound quantum channel generated by a set $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ of channels. We consider the uninformed user scenario, where precise knowledge about the identity of the channel is available neither to encoder nor decoder. An *entanglement generating* (l, k_l)*-code* for $\mathfrak{I}$ is a pair $(\mathcal{R}^l, \varphi^l)$ where $\mathcal{R}^l \in \mathcal{C}(\mathcal{H}_B^{\otimes l}, \mathcal{K}^l)$ is a channel with $k_l = \dim \mathcal{K}^l$ and φ_l a pure state on $\mathcal{K}^l \otimes \mathcal{H}_A^{\otimes l}$. A positive number R is an *achievable rate for entanglement generation* over $\mathfrak{I}$ if there is a sequence of (l, k_l)-entanglement generating codes satisfying

1. $\liminf_{l\to\infty} \frac{1}{l} \log k_l \geq R$, and
2. $\liminf_{l\to\infty} \inf_{\mathcal{N}\in\mathfrak{I}} F(\phi_l, (id_{\mathcal{K}^l} \otimes \mathcal{R}^l \circ \mathcal{N}^{\otimes l})(\varphi_l)) = 1$, where ϕ_l denotes a maximally entangled state on $\mathcal{K}^l \otimes \mathcal{K}^l$.

The number

$$E(\mathfrak{I}) := \sup\{R : R \text{ is an achievable rate for entanglement generation over } \mathfrak{I}\}.$$

is called the *entanglement generating capacity* of $\mathfrak{I}$. In the following, we show in a case of compound quantum channel generated by a finite set $\mathfrak{I}$ of c.p.t.p. maps, that our results from the last section imply a coding theorem for entanglement generation. The transition to the case, where the set $\mathfrak{I}$ is arbitrary follows just by applying standard approximation techniques in the same way as it was done in [BBN09]

Theorem 33 (cf. Ref. [BBN09], Th. 13)**.** *Let* $\mathfrak{I} := \{\mathcal{N}_i\}_{i=1}^N$ *be a finite compound quantum channel,* $\mathfrak{I} \subset \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$*. Then*

$$E(\mathfrak{I}) \geq \lim_{l\to\infty} \frac{1}{l} \max_{\rho\in\mathcal{S}(\mathcal{H}_A^{\otimes l})} \min_{1\leq i\leq N} I_c(\rho, \mathcal{N}_i^{\otimes l}) \tag{4.35}$$

holds

Proof. First note that the limit on the r.h.s of (4.35) exists by standard arguments (see Ref. [BBN09], Remark 2). We just have to prove that the number

$$\min_{1\leq i\leq N} I_c(\rho, \mathcal{N}_i) - \epsilon$$

is an achievable rate for every state ρ on $\mathcal{H}_A$ and every $\epsilon > 0$, the rest is by standard blocking arguments. There is nothing to prove for states with $\min_{1\leq i\leq N} I_c(\rho, \mathcal{N}_i) \leq 0$. Therefore let ρ be a

state on $\mathcal{H}_A$ with $\min_{1\leq i\leq N} I_c(\rho,\mathcal{N}_i) > 0$. Consider the set $\mathfrak{X} := \{\rho_i\}_{i=1}^N$ of bipartite states in $\mathcal{H}_{AB}$, where ρ_i is defined

$$\rho_i := (id_{\mathcal{H}_A} \otimes \mathcal{N}_i)(\chi) \tag{4.36}$$

for $1 \leq i \leq N$. Here χ is the pure state on $\mathcal{H}_A \otimes \mathcal{H}_A$ such that the partial trace over any of the two subsystems results in the state ρ. We show that a good entanglement distillation protocol for the set $\mathfrak{X}$ of bipartite states generated by $\mathfrak{I}$ implies the existence of a good entanglement generation code for $\mathfrak{I}$. Following the proof of Lemma 31, there exists an (l, k_l)-distillation protocol $\mathcal{T} = \sum_{k=0}^{D} \mathcal{A}_k \otimes \mathcal{R}_k$ for $\mathfrak{X}$ with $\mathcal{A}_k \in \mathcal{C}^{\downarrow}(\mathcal{H}_A^{\otimes l}, \mathcal{K}^l)$ and $\mathcal{R}_k \in \mathcal{C}(\mathcal{H}_B^{\otimes l}, \mathcal{K}^l)$ for $k \in \{1, ..., D\}$ with D determined by $\dim \mathcal{H}_A$ and $\dim \mathcal{K}^l$ such that

$$\dim \mathcal{K}^l \geq \left\lfloor \exp\left(l \left(\min_{1\leq i\leq N} I_c(\rho,\mathcal{N}_i) - \epsilon \right) \right) \right\rfloor \tag{4.37}$$

and

$$\min_{1\leq i\leq N} F(\mathcal{T}(\rho_i), \phi_l) \geq 1 - 2^{-n\hat{c}} \tag{4.38}$$

with ϕ_l being the maximally entangled state on $\mathcal{K}^l$ and $\hat{c} > 0$ a constant . Notice, that in eq. (4.37), we used the identity

$$I_c(\rho,\mathcal{N}_i) = -S(A|B,\rho_i)$$

for every $i \in \{1, ..., N\}$. The definitions given in eq. (4.36) imply

$$\mathcal{A}_k \otimes \mathcal{R}_k(\rho) = (id_{\mathcal{K}^l} \otimes \mathcal{R}_k \circ \mathcal{N}_i)(\mathcal{A}_k \otimes id_{\mathcal{H}_A^{\otimes l}}(\chi))$$

for every $0 \leq k \leq D$ and $1 \leq i \leq N$. Therefore,

$$\begin{aligned} F(\mathcal{T}(\rho_i), \phi_l) &= \sum_{k=0}^{D} F(id_{\mathcal{K}^l} \otimes \mathcal{R}_k \circ \mathcal{N}_i^{\otimes l}(\mathcal{A}_k \otimes id_{\mathcal{H}_A^{\otimes l}}(\chi)), \phi_l) \\ &= \sum_{k:p_k\neq 0} p_k F(id_{\mathcal{K}^l} \otimes \mathcal{R}_k \circ \mathcal{N}_i^{\otimes l}(\varphi_k), \phi_l) \end{aligned} \tag{4.39}$$

holds for every i, where we used the definitions

$$p_k := \mathrm{tr}(\mathcal{A}_k(\rho)), \text{ and } \qquad \varphi_k := \frac{1}{p_k}(\mathcal{A}_k \otimes id_{\mathcal{H}_A^{\otimes l}})(\chi)$$

for $p_k \neq 0$, $0 \leq k \leq D$. Notice, that $\varphi_0, ..., \varphi_D$ are pure states, because the operations $\mathcal{A}_k$ are pure since they arise from an L-merging (see the proof of Lemma 31). Again because the fidelities are affine functions of the first input, (4.38) and (4.39) imply

$$\sum_{k:p_k\neq 0} p_k F\left(id_{\mathcal{K}^l} \otimes \mathcal{R}_k \circ \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}_i^{\otimes l}(\varphi_k), \phi_l \right) \geq 1 - 2^{-n\hat{c}}. \tag{4.40}$$

The r.h.s. of equation (4.39) is, in fact, an average of fidelities of entanglement generating codes $(\mathcal{R}_1, \varphi_1), ..., (\mathcal{R}_D, \varphi_D)$ with probabilities $p_1, ..., p_D$. This implies the existence of a number $k' \in \{1, ..., D\}$ such that with $\varphi := \varphi_{k'}$ and $\mathcal{R} := \mathcal{R}_{k'}$

$$\min_{1\leq i\leq N} F\left(id_{\mathcal{K}^l} \otimes \mathcal{R} \circ \mathcal{N}_i^{\otimes l}(\varphi), \phi_l\right) \geq 1 - 2^{-n\hat{c}} \tag{4.41}$$

holds. Eqns. (4.41) and (4.37) show that

$$\min_{1\leq i\leq N} I_c(\rho, \mathcal{N}_i) - \epsilon$$

is an achievable rate. □

To conclude this section we compare the proof of Theorem 33 given above with the one given in Ref. [BBN09]. The original achievability proof relies on the fact that good entanglement generation codes can be deduced from entanglement transmission codes working good on maximally mixed states on certain subspaces of the input space of the channels. The passage to arbitrary states is done by a compound version of the so-called BSST-Lemma [Ben+02]. Indeed, one of the results from Ref. [BBN09] is that the entanglement transmission capacity $\mathcal{Q}(\mathfrak{I})$ equals $E(\mathfrak{I})$ for every compound channel $\mathfrak{I}$.
The proof given above follows a more direct route by taking advantage of a direct correspondence between entanglement distillation from bipartite states and entanglement generation over quantum channels, which is very close even in the compound setting. In this way, we have demonstrated, that quantum state merging provides a genuine approach to problems of entanglement generation over quantum channels even in the compound setting.

4.3 Entanglement distillation from arbitrarily varying quantum sources

In this section, we prove a regularized formula for the one-way entanglement distillation capacity where the source is an AVQS generated by a set $\mathfrak{X} \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$.
We first prove the achievability part in case that $\mathfrak{X}$ is finite, where we derive suitable one-way entanglement distillation protocols for the AVQS $\mathfrak{X}$ from entanglement distillation protocols which are universal for the compound source $\mathrm{conv}(\mathfrak{X})$ with fidelity approaching one exponentially fast. In a second step, we generalize this result allowing $\mathfrak{X}$ to be any (not necessarily finite or countable) set on $\mathcal{H}_A \otimes \mathcal{H}_B$. To this end, we approximate $\mathfrak{X}$ by a polytope (which is known to be the convex hull of a finite set of states), where we utilize methods we borrow from Ref. [Ahl+12]. First we state some facts concerning the continuity of the one-way entanglement distillation capacity functions.

4.3.1 AVQS generated by finite sets

In this section, we assume $\mathfrak{X}$ to be a finite set of bipartite states. We show, that sequences of one-way entanglement distillation protocols for the compound source generated by the convex hull of $\mathfrak{X}$ with fidelity going to one exponentially fast can be modified to faithful entanglement distillation schemes for the AVQS $\mathfrak{X}$. We apply Ahlswede's robustification [Ahl86] and elimination [Ahl78] techniques. This method of proof is well-known in classical information theory, and found application also in the quantum setting where it was shown to be a useful approach to determine the entanglement transmission capacity of arbitrarily varying quantum channels (AVQC)[Ahl+12]. The following theorem is the core of the robustification technique. It was first proven in Ref. [Ahl80]. The version below (with a better constant) is from Ref. [Ahl86].

Theorem 34 (Robustification technique, cf. Theorem 6 in Ref. [Ahl86])**.**
Let $\mathbf{S}$ *be a set with* $|\mathbf{S}| < \infty$ *and* $l \in \mathbb{N}$. *If a function* $f : \mathbf{S}^l \to [0,1]$ *satisfies*

$$\sum_{s^l \in \mathbf{S}^l} f(s^l) q(s_1) \cdot \ldots \cdot q(s_l) \geq 1 - \gamma \tag{4.42}$$

for each type q *of sequences in* $\mathbf{S}^l$ *for some* $\gamma \in [0,1]$, *then*

$$\frac{1}{l!} \sum_{\sigma \in \mathfrak{S}_l} f(\sigma(s^l)) \geq 1 - (l+1)^{|\mathbf{S}|} \cdot \gamma \qquad \forall s^l \in \mathbf{S}^l. \tag{4.43}$$

The following theorem is the main result of this section.

Theorem 35. *Let* $\mathfrak{X} := \{\rho_s\}_{s \in \mathbf{S}} \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B), |\mathbf{S}| \leq \infty$. *For the AVQS generated by* $\mathfrak{X}$, *it holds*

$$D_{\rightarrow}^{AV}(\mathfrak{X}) \geq D_{\rightarrow}(\mathrm{conv}(\mathfrak{X})) = \lim_{k \to \infty} \sup_{\mathcal{T} \in \Theta_k} \inf_{p \in \mathfrak{P}(\mathbf{S})} D^{(1)}(\mathcal{T}, \rho_p^{\otimes k}), \tag{4.44}$$

where we use the definition

$$\rho_p := \sum_{s \in \mathbf{S}} p^l(s^l)\, \rho_{s^l} \tag{4.45}$$

for each $p \in \mathfrak{P}(\mathbf{S})$.

Remark 36. *The above statement actually holds with equality in (4.44) which we show in the proof of Corollary 28 below.*

Proof. We show, that each rate R, which is achievable for $A \to B$ entanglement distillation for the compound source generated by $\mathrm{conv}(\mathfrak{X})$, is also an achievable rate for $A \to B$ entanglement distillation for the AVQS generated by $\mathfrak{X}$. We indicate the elements of $\mathrm{conv}(\mathfrak{X})$ by probability distributions on $\mathbf{S}$, since

$$\mathrm{conv}(\mathfrak{X}) = \left\{ \rho_p : \ \rho_p = \sum_{s \in \mathbf{S}} p(s) \rho_s, \ p \in \mathfrak{P}(\mathbf{S}) \right\} \tag{4.46}$$

holds. We know from Proposition 32, that for an achievable $A \to B$ entanglement distillation rate R for the compound source generated by $\mathrm{conv}(\mathfrak{X})$, $\delta > 0$ and each sufficiently large block-length l, there exists a one-way LOCC channel $\tilde{\mathcal{D}}_l$, such that the condition

$$\min_{p \in \mathfrak{P}(\mathbf{S})} F(\tilde{\mathcal{D}}_l(\rho_p^{\otimes l}), \phi_l) \geq 1 - 2^{-lc_5} \tag{4.47}$$

is fulfilled with a maximally entangled state ϕ_l shared by A and B, such that

$$\frac{1}{l} \log \mathrm{sr}(\phi_l) \geq R - \delta \tag{4.48}$$

holds. Note that the minimization in (4.47) is because of (4.46). We define a function $f : \mathbf{S}^l \to [0,1]$ by $f(s^l) := F(\tilde{\mathcal{D}}_l(\rho_{s^l}), \phi_l)$ for each $s^l \in \mathbf{S}^l$, and infer from (4.47), that

$$\sum_{s^l \in \mathbf{S}^l} p(s_1) \cdot \ldots \cdot p(s_l)\ f(s^l)\ \geq 1 - 2^{-lc_5} \tag{4.49}$$

holds for each $p \in \mathfrak{P}(\mathbf{S})$ with a constant $c_5 > 0$. Let

$$\mathcal{U}_\sigma(\cdot) := U_{A,\sigma} \otimes U_{B,\sigma}(\cdot) U_{A,\sigma}^* \otimes U_{B,\sigma}^*, \tag{4.50}$$

for each permutation $\sigma \in \mathfrak{S}_l$, be the unitary channel, which permutes the tensor factors in $\mathcal{H}_{AB}^{\otimes l}$ according to σ, (with unitary matrices $U_{A,\sigma}, U_{B,\sigma}$ permuting the tensor bases on $\mathcal{H}_A^{\otimes l}$ resp. $\mathcal{H}_B^{\otimes l}$). It holds

$$\rho_{\sigma(s^l)} = \mathcal{U}_\sigma(\rho_{s^l}), \tag{4.51}$$

and consequently

$$f(\sigma(s^l)) = F(\tilde{\mathcal{D}}_l \circ \mathcal{U}_\sigma(\rho_{s^l}), \phi_l) \tag{4.52}$$

for each $s^l \in \mathbf{S}^l, \sigma \in \mathfrak{S}_l$. The functions in (4.52) fulfill the conditions of Theorem 34, which in turn implies, that

$$(1 - (l+1)^{|\mathbf{S}|}) \cdot 2^{-lc_5} \leq \frac{1}{l!} \sum_{\sigma \in \mathfrak{S}_l} F(\tilde{\mathcal{D}}_l \circ \mathcal{U}_\sigma(\rho_{s^l}), \phi_l) \tag{4.53}$$

$$= F(\hat{\mathcal{D}}_l(\rho_{s^l}), \phi_l) \tag{4.54}$$

is valid with the definition $\hat{\mathcal{D}}_l := \frac{1}{l!} \sum_{\sigma \in \mathfrak{S}_l} \tilde{\mathcal{D}}_l \circ \mathcal{U}_\sigma$. Notice, that $\hat{\mathcal{D}}_l$ is an $A \to B$ LOCC channel either. However, $\hat{\mathcal{D}}_l$ is not a reasonable protocol for entanglement distillation regarding the classical communication cost. Implementation of $\hat{\mathcal{D}}_l$ demands $A \to B$ communication of a number of classical messages increased by a factor $l!$ compared to the requirements of $\tilde{\mathcal{D}}_l$, which leads to super-exponential growth of required classical messages and consequently unbounded classical communication rates. We remark here, that for a coordination of the permutations in $\hat{\mathcal{D}}_l$, common randomness accessible to A and B, which is known to be a weaker resource than $A \to B$ communication, would suffice. Nonetheless, the asymptotic common randomness

consumption of the protocol would be above any rate either. We will a apply the well-known derandomization technique which first appeared in Ref. [Ahl78] to construct $A \to B$ LOCC channel with reasonable classical communication requirements (actually, we will show, that we can approximate the classical cost of $A \to B$ distillation of the compound source $\mathrm{conv}(\mathfrak{X})$.
Let $X_1, ..., X_{K_l}$ be a sequence of i.i.d. random variables, each distributed uniformly on $\mathfrak{S}_l$. We define a function $g : \mathfrak{S}_l \times \mathbf{S}^l \to [0,1]$ by

$$g(\sigma, s^l) = 1 - F(\tilde{\mathcal{D}}_l \circ \mathcal{U}_\sigma(\rho_{s^l}), \phi_l) \qquad (\sigma \in \mathfrak{S}_l, s^l \in \mathbf{S}^l). \tag{4.55}$$

One readily verifies, that

$$\mathbb{E}\left[g(X_1, s^l)\right] = 1 - F(\hat{\mathcal{D}}_l(\rho_{s^l}), \phi_l) \leq (l+1)^{|\mathbf{S}|}\, 2^{-lc_5} := \epsilon_l \tag{4.56}$$

holds for each $s^l \in \mathbf{S}^l$. Thus, for each $s^l \in \mathbf{S}^l$, and $\nu_l \in (0,1)$, we yield

$$\begin{aligned}
\Pr\left(\sum_{k=1}^{K_l} g(X_k, s^l) > K_l\nu_l\right) &= \Pr\left(\prod_{k=1}^{K_l} \exp(g(X_k, s^l)) > 2^{K_l\nu_l}\right) && (4.57)\\
&\leq 2^{-K_l\nu_l} \cdot \mathbb{E}\left[\exp(g(X_k, s^l)\right]^{K_l} && (4.58)\\
&\leq 2^{-K_l\nu_l} \cdot (1 + \mathbb{E}\left[\exp(g(X_k, s^l)\right])^{K_l} && (4.59)\\
&\leq 2^{-K_l\nu_l} \cdot 2^{K_l \log(1+\epsilon_l)} && (4.60)\\
&\leq 2^{-K_l(\nu_l - 2\epsilon_l)}. && (4.61)
\end{aligned}$$

Eq. (4.58) above is by Markov's inequality, (4.59) follows from the fact, that $\exp(x) \leq 1 + x$ holds for $x \in [0,1]$, (4.60) is by (4.56), and (4.61) follows from the inequality $\log(1+x) \leq 2x$ being valid for $x \in (0,1)$. From (4.57)-(4.61) and application of de Morgan's laws, it follows

$$\begin{aligned}
\Pr\left(\forall s^l \in \mathbf{S}^l : \frac{1}{K_l}\sum_{k=1}^{K_l} g(X_k, s^l) \leq \nu_l\right) &\geq 1 - |\mathbf{S}|^l \cdot 2^{-K_l(\nu_l - 2\epsilon_l)} && (4.62)\\
&\geq 1 - 2^{-l\frac{(\theta-\kappa)}{2}}, && (4.63)
\end{aligned}$$

for large enough l,where the last line results from the choosing $\nu_l = 2^{-l\kappa}$ and $K_l = 2^{l\theta}$ with $\theta, \kappa > 0$. If we choose κ and θ in a way, that they fulfill $0 < \kappa < \theta < c$, the r.h.s. of (4.63) is strictly positive and we find a realization $\sigma_1, ..., \sigma_{K_l}$ of $X_1, ..., X_{K_l}$, such that for each $s^l \in \mathbf{S}^l$

$$\begin{aligned}
2^{-l\kappa} &\geq \frac{1}{K_l}\sum_{k=1}^{K_l} g(\sigma_k, s^l) && (4.64)\\
&= 1 - \frac{1}{K_l}\sum_{k=1}^{K_l} F(\tilde{\mathcal{D}}_l \circ \mathcal{U}_{\sigma_k}(\rho_{s^l}), \phi_l) && (4.65)\\
&= 1 - F(\mathcal{D}_l(\rho_{s^l}), \phi_l), && (4.66)
\end{aligned}$$

where we defined $\mathcal{D}_l := \frac{1}{K_l}\sum_{k=1}^{K_l} \tilde{\mathcal{D}} \circ \mathcal{U}_{\sigma_k}$. With (4.48) and (4.66), it is shown, that for each sufficiently large blocklength l, we find a one-way entanglement distillation protocol with

$$\min_{s^l \in \mathbf{S}^l} F(\mathcal{D}_l(\rho_{s^l}), \phi_l) \geq 1 - 2^{-l\kappa}, \quad \text{and} \qquad \frac{1}{l}\log \mathrm{sr}(\phi_l) \geq R - \delta. \tag{4.67}$$

Notice, that the number of different classical messages to be communicated by A within application of $\mathcal{D}_l$ is increased by a factor $2^{l\theta}$ compared to the message transmission demanded by $\tilde{\mathcal{D}}_l$, i.e. the communication rate is increased by θ (which we can choose to be an arbitrarily small fixed number). □

4.3.2 General AVQS

In this section, we generalize the results of the preceding section, admitting the AVQS to be generated by any not necessarily finite or countable set $\mathfrak{X}$ of states on $\mathcal{H}_A \otimes \mathcal{H}_B$. We approximate the closed convex hull of $\mathfrak{X}$ by a polytope, which is known as the convex hull of a finite set of points and apply Theorem 35, together with continuity properties of the capacity function. The proof strategy has some similarities with the argument given in Ref. [Ahl+12] for entanglement transmission over general arbitrarily varying quantum channels. To prepare ourselves for the approximation, we need some notation and results from convex geometry which we state first. For a subset A of a normed space $(V, \|\cdot\|)$, $\overline{A}$ is the closure and $\mathrm{aff}A$ is the affine hull of A. If A is a convex set, the relative interior $\mathrm{ri}A$ is the interior and the relative boundary $\mathrm{rebd}A$ of A are the interior and boundary of A regarding the topology on $\mathrm{aff}A$ induced by $\|\cdot\|$.

Lemma 37 (Ref. [Ahl+12], Lemma 34)**.** *Let A, B be compact sets in $\mathbb{C}^n$ with $A \subset B$ and*

$$d_H(\mathrm{rebd}B, A) = t > 0, \tag{4.68}$$

where $\|\cdot\|$ denotes any norm on $\mathbb{C}^n$. Let P a polytope with $A \subset P$ and $d_H(A, P) \leq \delta$, where $\delta \in (0, t]$ and d_H is the Hausdorff distance induced by $\|\cdot\|$. Then $P' := P \cap \mathrm{aff}A$ is also a polytope and $P \subset B$.

With the above statement and the assertions of the preceding section, we are prepared to prove the following theorem which is the main result of this section.

Theorem 38. *Let $\mathfrak{X} := \{\rho_s\}_{s\in\mathbf{S}}$ be a set of states on $\mathcal{H}_A \otimes \mathcal{H}_B$. For each $\delta > 0$ and $k \in \mathbb{N}$, there exists a number $l_0 \in \mathbb{N}$, such that for each $l > l_0$, there is an $A \to B$ LOCC channel $\mathcal{D}_l$ fulfilling*

$$\inf_{s\in\mathbf{S}^l} F(\mathcal{D}_l(\rho_{s^l}), \phi_l) \geq 1 - 2^{-lc_6} \tag{4.69}$$

with a maximally entangled state ϕ_l shared by A and B and a constant $c_6 > 0$, such that

$$\frac{1}{l}\log \mathrm{sr}(\phi_l) \geq \frac{1}{k}\sup_{\mathcal{T}\in\Theta_k}\inf_{\tau\in\mathrm{conv}(\mathfrak{X})} D^{(1)}_{\rightarrow}(\tau^{\otimes k}, \mathcal{T}) - \delta \tag{4.70}$$

holds, where the function $D^{(1)}_{\rightarrow}$ is defined in (4.3).

Proof. Let $\mathcal{T} := \{\mathcal{T}_j\}_{j=1}^J$ be any instrument with domain $\mathcal{L}(\mathcal{H}_A^{\otimes k})$, $\delta > 0$. Dealing only with the nontrivial case, we show, that

$$\inf_{\rho\in\mathrm{conv}(\mathfrak{X})} \frac{1}{k} I_c(A\rangle BB', \hat{\mathcal{T}}(\rho^{\otimes k})) - \delta > 0 \tag{4.71}$$

is an achievable rate (remember our notation from (4.5)). Since the Hausdorff distance between $\mathrm{conv}(\mathfrak{X})$ and $\overline{\mathrm{conv}(\mathfrak{X})}$ is zero, it makes no difference if we consider the set $\overline{\mathrm{conv}(\mathfrak{X})}$ instead. We briefly describe the strategy of our proof. We approximate the set $\overline{\mathrm{conv}(\mathfrak{X})}$ from the outside by a polytope P_η. Since P_η, as a polytope, is the convex hull of a finite set of points, Theorem 35 can be applied. A technical issue (cf. Ref. [Ahl+12]) is, to ensure, that the approximating polytope completely consists of density matrices, i.e. $P_\eta \subset \mathcal{S}(\mathcal{H}_{AB})$. We achieve this by a slight depolarization of the states in $\mathrm{conv}(\mathfrak{X})$, such that the resulting set does not touch the boundary of $\mathcal{S}(\mathcal{H}_{AB})$. Define, for $\gamma \in [0,1]$ the channel $\mathcal{N}_\gamma \in \mathcal{C}(\mathcal{H}_A \otimes \mathcal{H}_B)$ by $\mathcal{N}_\gamma := \mathcal{N}_{A,\gamma} \otimes \mathcal{N}_{B,\gamma}$, where $\mathcal{N}_{X,\gamma}$ is the γ-depolarizing channel on the subsystem X, $X = A, B$, i.e

$$\mathcal{N}_\gamma(\tau) = (1-\gamma)^2\tau + \gamma(1-\gamma)(\tau_A \otimes \pi_B + \pi_A \otimes \tau_B) + \gamma^2(\pi_A \otimes \pi_B) \tag{4.72}$$

for each $\tau \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, were π_A, π_B are maximally mixed states and τ_A, τ_B are the marginals of τ on $\mathcal{H}_A$, $\mathcal{H}_B$. Notice, that $\mathcal{N}_\gamma$ is defined in terms of local depolarizing channels on the subsystems. This is required, since we are restricted to one-way LOCC channels. It holds

$$\begin{aligned} \|\mathcal{N}_\eta(\tau) - \tau\|_1 &\leq \|(1-\eta)^2\tau - \tau\|_1 + \eta(1-\eta)\|\tau_A \otimes \pi_B + \pi_A \otimes \tau_B\|_1 && (4.73)\\ &\quad + \eta\|\pi_A \otimes \pi_B\|_1 && (4.74)\\ &\leq 6\eta && (4.75) \end{aligned}$$

for each state τ on $\mathcal{H}_A \otimes \mathcal{H}_B$. Moreover, it holds $\overline{\mathcal{N}_\eta(\mathrm{conv}(\mathfrak{X}))} = \mathcal{N}_\eta(\overline{\mathrm{conv}(\mathfrak{X})}) \subset \mathrm{ri}\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, which implies

$$\inf\left\{\|\rho - \rho'\|_1 :\ \rho \in \overline{\mathcal{N}_\eta(\mathrm{conv}(\mathfrak{X}))}, \rho' \in \mathrm{rebd}(\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B))\right\} > 0. \tag{4.76}$$

Therefore, due to of Lemma 37 and Theorem 3.1.6 in Ref. [Web94], there exists, for each small enough number $\eta > 0$, a polytope $P_\eta := \mathrm{conv}(\{\tau_e\}_{e\in E_\eta}) \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\mathcal{N}_\eta(\mathrm{conv}(\mathfrak{X})) \subset P_\eta$ and

$$d_H(\mathcal{N}_\eta(\mathrm{conv}(\mathfrak{X})), P_\eta) \leq \eta. \tag{4.77}$$

Applying Theorem 35 to the finite AVQS generated by the extremal set $\{\tau_e\}_{e\in E}$ of the polytope P_η, we know, that for each sufficiently large blocklength l, there exists an $A \to B$ LOCC channel $\hat{\mathcal{D}}_l$ such that

$$F(\hat{\mathcal{D}}_l(\tau_{e^l}), \phi_l) \geq 1 - 2^{-lc_6} \tag{4.78}$$

holds with a maximally entangled state ϕ_l shared by A and B for each $e^l \in E^l$ with Schmidt rank fulfilling

$$\frac{1}{l}\log \mathrm{sr}(\phi_l) \geq \frac{1}{k}\inf_{\tau\in P_\eta} I_c(A\rangle BB', \hat{\mathcal{T}}(\tau^{\otimes k})) - \frac{\delta}{2}. \tag{4.79}$$

Since $\mathcal{N}_\eta(\mathrm{conv}(\mathfrak{X})) \subset P_\eta$ holds, the depolarized version $\mathcal{N}_\eta(\rho_s)$ of each state ρ_s, $s \in \mathbf{S}$ can be written as a convex combination of elements from $\{\tau_e\}_{e\in E_\eta}$, i.e.

$$\mathcal{N}_\eta(\rho_s) = \sum_{e\in E_\eta} q(e|s)\ \tau_e \tag{4.80}$$

with a probability distribution $q(\cdot|s)$ on E_η for each $s \in \mathbf{S}$. We define a one-way LOCC channel $\mathcal{D}_l$ by $\mathcal{D}_l := \hat{\mathcal{D}}_l \circ \mathcal{N}_\eta^{\otimes l}$ and deduce

$$\begin{aligned}
F(\mathcal{D}_l(\rho_{s^l}), \phi_l) &= F(\hat{\mathcal{D}}_l(\mathcal{N}_\eta^{\otimes l}(\rho_{s^l})), \phi_l) && (4.81)\\
&= F\left(\hat{\mathcal{D}}_l\left(\bigotimes_{i=1}^{l}\mathcal{N}_\eta(\rho_{s_i})\right), \phi_l\right) && (4.82)\\
&= F\left(\hat{\mathcal{D}}_l\left(\bigotimes_{i=1}^{l}\sum_{e_i\in E} q(e_i|s_i)\tau_{e_i}\right), \phi_l\right) && (4.83)\\
&= \sum_{e_1\in E}\cdots\sum_{e_l\in E}\prod_{i=1}^{l} p(e_i|s_i) F\left(\hat{\mathcal{D}}_l(\tau_{e_i}), \phi_l\right) && (4.84)\\
&= \sum_{e^l\in E_\eta^l} q^l(e^l|s^l)\, F(\hat{\mathcal{D}}_l(\tau_{e^l}), \phi_l) && (4.85)\\
&\geq 1 - 2^{-lc_6} && (4.86)
\end{aligned}$$

for each $s^l = (s_1, ..., s_l) \in \mathbf{S}^l$ where we used (4.80) in (4.83) and (4.86) is by (4.78). To complete the proof, we show, that for small enough η,

$$\inf_{\rho\in\mathrm{conv}(\mathfrak{X})} I_c(A\rangle BB', \hat{\mathcal{T}}(\rho^{\otimes k})) \geq \inf_{\tau\in P_\eta} I_c(A\rangle BB', \hat{\mathcal{T}}(\tau^{\otimes k})) - \frac{k\delta}{2} \tag{4.87}$$

holds. For each $\rho \in \mathrm{conv}(\mathfrak{X})$, $\tau \in P_\eta$, we have

$$\begin{aligned}
\|\rho - \tau\|_1 &\leq \|\rho - \mathcal{N}_\eta(\rho)\|_1 + \|\mathcal{N}_\eta(\rho) - \tau\|_1 && (4.88)\\
&\leq 6\eta + \|\mathcal{N}_\eta(\rho) - \tau\|_1 && (4.89)
\end{aligned}$$

where the last estimation is by (4.75). From (4.89), we can conclude, that

$$d_H(\mathrm{conv}(\mathfrak{X}), P_\eta) \leq d_H(\mathcal{N}_\eta(\mathrm{conv}(\mathfrak{X}), P_\eta) + 6\eta \leq 7\eta \tag{4.90}$$

holds, which implies, via Lemma 29,

$$\left|\inf_{\rho\in\mathrm{conv}(\mathfrak{X})} I_c(A\rangle BB', \hat{\mathcal{T}}(\rho^{\otimes k})) - \inf_{\tau\in P_\eta} I_c(A\rangle BB', \hat{\mathcal{T}}(\tau^{\otimes k}))\right| \leq k\nu(7\eta). \tag{4.91}$$

If now η is chosen small enough to ensure $\nu(7\eta) < \frac{\delta}{2}$, (4.87), and we conclude, collecting inequalities, that the entanglement rate of $\mathcal{D}_l$ is

$$\begin{aligned}
\frac{1}{l}\log \mathrm{sr}(\phi_l) &\geq \frac{1}{k}\inf_{\tau\in P_\eta} I_c(A\rangle BB', \hat{\mathcal{T}}(\tau^{\otimes k})) - \frac{\delta}{2} && (4.92)\\
&\geq \frac{1}{k}\inf_{\rho\in\mathrm{conv}(\mathfrak{X})} I_c(A\rangle BB', \hat{\mathcal{T}}(\rho^{\otimes k})) - \delta && (4.93)
\end{aligned}$$

where (4.92) is (4.79), (4.93) is by (4.87). □

Proof of Theorem 28. The rightmost equality in (4.9) is Corollary 25.1. We prove the first equality. Achievability directly follows from Theorem 38. For the converse statement, let $\mathfrak{X} := \{\rho_s\}_{s\in\mathbf{S}}$ and $\sigma \in \mathrm{conv}(\mathfrak{X})$. By Carathéordory's Theorem (see e.g. Ref. [Web94], Theorem 2.2.4.), σ can be written as a finite convex combination of elements of $\mathfrak{X}$, say

$$\sigma = \sum_{s\in\mathbf{S}'} p(s)\rho_s. \tag{4.94}$$

with $|\mathbf{S}'| < \infty$. Thus, for an $A \to B$ LOCC channel $\mathcal{D}_l$ for blocklength l with suitable maximally entangled state ϕ_l, it holds

$$\begin{aligned}\inf_{s^l\in\mathbf{S}^l} F(\mathcal{D}_l(\rho_{s^l}),\phi_l) &\leq \min_{s^l\in\mathbf{S}'^l} F(\mathcal{D}_l(\rho_{s^l}),\phi_l) && (4.95)\\ &\leq \sum_{s^l\in\mathbf{S}'^l} p^l(s^l)F(\mathcal{D}_l(\rho_{s^l}),\phi_l) && (4.96)\\ &= F(\mathcal{D}_l(\sigma^{\otimes l}),\phi_l). && (4.97)\end{aligned}$$

Since (4.97) holds for each element of $\mathrm{conv}(\mathfrak{X})$, each rate R which is an $A \to B$ achievable entanglement distillation rate for the AVQS generated by $\mathfrak{X}$ is also achievable for the compound quantum source generated by $\mathrm{conv}(\mathfrak{X})$, thus the converse statement in Corollary 25.1 applies. □

Having determined the one-way entanglement distillation capacity $D_{\to}^{AV}$, the continuity properties of the capacity function on the r.h.s. of (4.91) imply the following corollary.

Corollary 39. *Identifying each set of states with its closure, $D_{\to}^{AV}$ is uniformly continuous in the metric defined by the Hausdorff distance on compact sets of density matrices. If $\mathfrak{X}, \mathfrak{X}' \subset \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ are two compact sets with $d_H(\mathfrak{X}', \mathfrak{X}) < \epsilon \leq \frac{1}{2}$ it holds*

$$|D_{\to}^{AV}(\mathfrak{X}') - D_{\to}^{AV}(\mathfrak{X})| \leq \nu(\epsilon). \tag{4.98}$$

Remark 40. *Corollary 39 classifies the AVQS one-way entanglement distillation task as well-behaved in the following sense. Two different AVQS with generating sets being near in the Hausdorff sense will have approximately equal capacities.*
An example for a situation where "capacity" is a more fragile quantity is transmission of classical messages over an arbitrarily varying quantum channel. The capacity C_{random} for classical message transmission using correlated random codes is continuous, while it can be shown, that in some cases, the capacity using deterministic codes, C_{det}, is discontinuous on certain points [BJ14a].

5 Secret-key distillation for compound classical-quantum-quantum sources

5.1 Basic definitions and main Result

In this section we give precise definitions of the secret-key distillations task and the corresponding capacities of compound memoryless cqq sources with and without assumption of SMI. We also state the two main results Theorem 48 and Theorem 49.

5.1.1 Source model

A *compound memoryless quantum source* generated by a set of density matrices $\mathfrak{I} \subset \mathcal{S}(\mathcal{K})$ on a Hilbert space $\mathcal{K}$ is the source described by the set of possible output density matrices

$$\mathfrak{I}^{\otimes n} := \{\rho^{\otimes n} : \ \rho \in \mathfrak{I}\}$$

for each blocklength $n \in \mathbb{N}$. This source definition models a situation, where the source statistics is memoryless, but the generating density matrix is not known. The communication parties processing the source, only can be sure, that the output statistics is governed by memoryless extensions of a density matrix from $\mathfrak{I}$. In this paper, the compound sources considered are generated by tripartite classical-quantum density matrix of the form

$$\rho := \sum_{x \in \mathcal{X}} p(x) \left|x\right\rangle \left\langle x\right| \otimes \rho_{BE,x} \ \in \ \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$$

which is the coherified way to represent a statistics where A receives outputs of a classical source with distribution $p \in \mathfrak{P}(\mathcal{X})$, while B, and E receive quantum systems with joint state $\rho_{BE,x} \in \mathcal{S}(\mathcal{H}_{BE})$, dependent on the letter x. If a system is classical, we regard the basis used for coherifying the systems as fixed once and for all (we fix it to be the canonical basis $\{x\}_{x \in \mathcal{X}}$). Note, that ρ can be alternatively described by the pair (p, V) with $p \in \mathfrak{P}(\mathcal{X})$ being a probability distribution on $\mathcal{X}$ and $V \in \mathcal{CQ}(\mathcal{X}, \mathcal{H}_{BE})$ with

$$V(x) := \rho_{BE,x},$$

notations, we will use interchangeably. Note, that

$$I(X; BE, \rho) = \chi(p, V).$$

We define the class of density matrices in $\mathcal{S}(\mathcal{H}_{ABE})$ with classical A-system, $\mathcal{H}_A := \mathbb{C}^{|\mathcal{X}|}$ by

$$\mathcal{S}_{cqq}(\mathcal{H}_{ABE}) := \left\{ \rho = \sum_{x\in\mathcal{X}} p(x)\,|x\rangle\langle x| \otimes \rho_x : \; p \in \mathfrak{P}(\mathcal{X}) \text{ and } \rho_x \in \mathcal{S}(\mathcal{H}_{BE}) \; (x \in \mathcal{X}) \right\}.$$

Since bipartite sources with a classical and a quantum subsystem also occur, we also define

$$\mathcal{S}_{cq}(\mathcal{H}_{AB}) := \left\{ \rho = \sum_{x\in\mathcal{X}} p(x)\,|x\rangle\langle x| \otimes \rho_x : \; p \in \mathfrak{P}(\mathcal{X}) \text{ and } \rho_x \in \mathcal{S}(\mathcal{H}_{B}) \; (x \in \mathcal{X}) \right\}.$$

To increase notational flexibility within our considerations, we define for each given set $\mathfrak{I} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$

$$\mathcal{P}_{\mathfrak{I}} := \left\{ p \in \mathfrak{P}(\mathcal{X}) : \; \exists \rho \in \mathfrak{I} \text{ with } \rho_A = \sum_{x\in\mathcal{X}} p(x)\,|x\rangle\langle x| \right\}, \text{ and}$$

$$\mathfrak{I}_p := \left\{ \rho \in \mathfrak{I} : \; \rho_A := \sum_{x\in\mathcal{X}} p(x)\,|x\rangle\langle x| \right\}$$

for each $p \in \mathcal{P}$. With these notations, $\mathcal{P}_{\mathfrak{I}}$ is the set of marginal probability distributions which can occur at the sender's site, while $\mathfrak{I}_p$ collects for each p the set of possible occurring cqq density matrices under the constraint that p generates the marginal distributions on the sender's systems. For a more efficient notation of the capacity formulas appearing below, we also define the following sets of marginal distributions deriving from states in $\mathfrak{I}_p$ by

$$\mathfrak{I}_p^{AB} := \{\rho_{AB} : \; \rho \in \mathfrak{I}_p\}, \quad \mathfrak{I}_p^{AE} = \{\rho_{AE} : \; \rho \in \mathfrak{I}_p\}$$

for each $p \in \mathcal{P}_{\mathfrak{I}}$.
In this paper, the systems labeled A, and B belong to the legitimate communication parties, while E labels the systems of the eavesdropper. The definitions in the next section model a situation, where the legitimate parties do not know, which density matrix from $\mathfrak{I}$ governs the source statistics (except the SMI case, where A knows his/her marginal statistics). The eavesdropper instead, may know the source statistics and the protocols applied by the legitimate users.

5.1.2 Secret key generation from compound cqq sources: Definitions and results

For a cqq source with fixed density matrix $\rho \in \mathcal{S}_{cqq}(\mathcal{X}, \mathcal{H}_{ABE})$, a secret key generation protocol for given blocklength n is performed, informally speaking, as follows. The A-party generates from his/her source output messages l and m where m is the the key value for A and l is broadcasted to the remaining parties via a noiseless channel. The legitimate receiver subsequently determines a key value by applying a quantum measurement, which can be chosen according to the received l. This results in a tuple (K, K', Λ, X^n), where K (K') is the key random value of A (B), Λ the random variable representing the public transmission, and X^n the classical random variable initially received by A. The formal definition for the described type of protocol is as follows.

Definition 41. *An* (n, M, L) (forward) secret-key distillation protocol *for states on* $\mathcal{S}_{cqq}(\mathcal{H}_{ABE})$ *is a pair* (T, D)*, with* $T : \mathcal{X}^n \to \mathfrak{P}([L] \times [M])$ *being a stochastic matrix, and* $D = \{D_{lm}\}_{l\in[L],m\in[M]}$ *being a set of matrices,* $0 \leq D_{lm} \leq \mathbb{1}_{\mathcal{H}_B}^{\otimes n}$ *such that*

$$\sum_{m=1}^{M} D_{lm} = \mathbb{1}_{\mathcal{H}_B}^{\otimes n}$$

holds for each $l \in [L]$.

We will also consider the situation, where the sender has full knowledge of the statistics of his/her part of the source. If this is assumed, the sender can choose the stochastic matrix of a protocol according to this knowledge.

Definition 42. *An* (n, M, L) (forward) secret-key distillation protocol with sender marginal information (SMI) *for a set* $\mathfrak{I} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$ *is a family* $(T_p, D)_{p\in\mathcal{P}_\mathfrak{I}}$*, with* (T_p, D) *being an* (n, M, L) *forward secret key distillation protocol for states on* $\mathcal{S}_{cqq}(\mathcal{H}_{ABE})$ *for each* $p \in \mathfrak{P}_\mathfrak{I}$.

Next, we define the performance of (n, L, M) forward secret-key distillation protocols with and without SMI performed in a compound source generated by a set $\mathfrak{I} := \{\rho_s\}_{s\in S} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$. For a protocol with SMI, (T_p, D) is performed, if the cqq density matrix is from $\mathfrak{I}_p$. It is convenient, to express the aftermath in coherified manner by the state

$$\rho_{\Lambda KK'E^n,s} := \sum_{l=1}^{L}\sum_{m,m'=1}^{M}\sum_{x^n\in\mathcal{X}^n} p^n(x^n)\, T_p(l,m|x^n)\, |l\rangle\langle l| \otimes |m\rangle\langle m| \otimes |m'\rangle\langle m'| \otimes \mathrm{tr}_{\mathcal{H}_B^{\otimes n}}((D_{lm'} \otimes \mathbb{1}_{\mathcal{H}_E}^{\otimes n})V^{\otimes n}(x^n)). \tag{5.1}$$

We are especially interested in the marginal state

$$\rho_{\Lambda KE^n,s} := \sum_{l=1}^{L}\sum_{m=1}^{M}\sum_{x^n\in\mathcal{X}^n} p^n(x^n)\, T_p(l,m|x^n)\mathrm{tr}(D_{lm'}V_B^{\otimes n}(x^n))\ |l\rangle\langle l| \otimes |m\rangle\langle m| \otimes \rho_{E,x^n},$$

and the probability distribution (K_s, K_s') belonging to the key given by

$$P_{KK',s}(m,m') = \langle m\otimes m', \rho_{KK'}\, m\otimes m'\rangle = \sum_{l=1}^{L}\sum_{x^n\in\mathcal{X}^n} p^n(x^n)\, T_p(l,m|x^n)\mathrm{tr}(D_{lm'}\rho_{B,x^n})$$

for all $m, m' \in [M]$ when the state governing the statistics of the source is

$$\rho_s = \sum_{x\in\mathcal{X}} p(x)\ |x\rangle\langle x| \otimes \rho_{BE,x}.$$

The case of application of a protocol without SMI can be regarded as the special case, where T_p does not depend on p. The following definition quantifies the performance of each (n, M, L) forward secret-key distillation protocol with SMI when performed on a set of cqq density matrices. The corresponding definition for the case without SMI is easily obtained by dropping all text in brackets from the following definitions.

Definition 43. *An (n, M, L) forward secret-key distillation protocol (with SMI) for a set $\mathfrak{I} := \{\rho_s\}_{s\in S} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$ is an (n, M, L, λ) forward secret-key distribution protocol (with SMI) for $\mathfrak{I}$ if the following two assertions are satisfied simultaneously for all $s \in S$.*

1. $\Pr(K_s \neq K'_s) \leq \lambda$.
2. $\log M - H(K_s) + I(K; E^n\Lambda, \rho_{\Lambda K E^n, s}) \leq \lambda$.

The first condition above is a bound on the probability, that key values mismatch. The second one guarantees for small λ, that the key is approximately equidistributed and secure. The left hand side of the inequality in 2. of the above definition can be regarded as a quantum version of the so-called *security index* introduced in classical information theory [CN04]. For a pair (K, Z) of classical random variables, the *security index* of K against Z is given by the expression

$$S_{SID}(K|Z) := \log \mathrm{supp}(P_K) - H(K) + I(K; Z).$$

The security index is well-known as a useful criterion for quantifying equidistribution and the degree of decoupling from the eavesdropper (see e.g. [CK11] for more information). From the classical security index, also the above introduced quantum version stems its operational significance. If $(K, \Lambda, Z)_s$ is a tuple of random variables with K_s being the key random variable, Λ_s belonging to the public message of the protocol and Z_s the classical random variable obtained by measurement on the eavesdropper's system, the second condition in Definition 43 implies $S_{SID}(K_s\Lambda_s|Z_s) \leq \lambda$, because

$$\begin{aligned} S_{SID}(K_s|\Lambda_s, Z_s) &= \log M - H(K_s) + I(K_s; Z_s, \Lambda_s) \\ &\leq \log M - H(K_s) + I(K; E^n\Lambda, \rho_{\Lambda K E^n, s}) \\ &\leq \lambda \end{aligned} \tag{5.2}$$

holds by the Holevo bound [Hol73].

Remark 44. *In their work [DW05], Devetak and Winter proposed a slightly stronger security criterion to be satisfied instead of the one used in Definition 43. The authors of the present paper feel, that in general the security criterion therein will hardly be satisfied by universal secret-key distillation protocols in general. However, the results in the subsequent sections applied to the case of a compound source $\mathfrak{I}$ with $|\mathfrak{I}| = 1$ show, that imposing the weaker criterion from Definition 43 does not lead to higher capacities than in [DW05] in case of perfectly known source statistics.*

Definition 45. *A nonnegative number R is called an* achievable secret-key distillation rate for $\mathfrak{I}$ (with SMI)*, if for each $\epsilon > 0, \delta > 0$ exist numbers n_0 and $0 < R_c < \infty$ such for each possible marginal state ρ_A there is an (n, M, L, ϵ) secret-key distillation protocol for $\mathfrak{I}$ (with SMI), such that*

$$M \geq \exp(n(R - \delta)), \quad \textit{and} \quad L \leq \exp(nR_c) \tag{5.3}$$

for each $n > n_0$. We define the forward secret-key capacity of $\mathfrak{I}$ with SMI *by*

$$K_{\rightarrow,SMI}(\mathfrak{I}) := \sup\{R \geq 0 : \ R \text{ is an achievable secret key distillation rate for } \mathfrak{I} \text{ with SMI}\}. \tag{5.4}$$

and the forward secret-key capacity of $\mathfrak{I}$ without SMI *by*

$$K_{\rightarrow}(\mathfrak{I}) := \sup\{R \geq 0 : \ R \text{ is an achievable secret key distillation rate for } \mathfrak{I} \text{ without SMI}\}.$$

What we define next, is a regularity condition on generating sets for compound cqq sources.

Definition 46 (Regularity Condition)**.** *We call a set $\mathfrak{I} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$*

- ϵ-regular*, if there is a $\delta > 0$ such that the implication*

$$\|p - q\|_1 < \delta \quad \Rightarrow \quad d_H(\mathfrak{I}_p^{AB}, \mathfrak{I}_q^{AB}) + d_H(\mathfrak{I}_p^{AE}, \mathfrak{I}_q^{AE}) < \epsilon$$

  holds for each pair $p, q \in \mathcal{P}_{\mathfrak{I}}$, where d_H denotes the Hausdorff distance generated by the trace norm distance on the underlying matrix spaces.

- regular*, if $\mathfrak{I}$ is ϵ-regular for each $\epsilon > 0$.*

Remark 47. *The regularity condition given above aims to cover an as large as possible class of reasonable sets of cqq density matrix under the condition that general protocol constructions are successful. The reader interested in detailed discussion of this condition is referred to section 5.4. In Section 5.4.2 we show using results from the theory of set-valued functions, that the above condition of regularity can be weakened somewhat to include even a larger class of cqq sources.*

The following two theorems state the main results proven in this paper.

Theorem 48. *Let $\mathfrak{I}$ be a regular set of cqq density matrices on $\mathcal{H}_{ABE}$. It holds*

$$K_{\rightarrow}(\mathfrak{I}) = \lim_{k\to\infty} \frac{1}{k} K_{\rightarrow}^{(1)}(\mathfrak{I}^{\otimes k}), \tag{5.5}$$

where for a set $\mathfrak{A} := \{\sum_{y\in\mathcal{Y}} p(y) \ |y\rangle\langle y| \otimes \sigma_y\}$ on some space,

$$K_{\rightarrow}^{(1)}(\mathfrak{A}) := \inf_{p\in\mathcal{P}_{\mathfrak{A}}} \sup_{\Gamma:=T\leftarrow U\leftarrow Y_p} \left(\inf_{\sigma\in\mathfrak{A}_p} I(U;B|T,\sigma_\Gamma) - \sup_{\sigma\in\mathfrak{A}_p} I(U;E|T,\sigma_\Gamma) \right).$$

The supremum above is over all Markov chains $T \leftarrow U \leftarrow Y_p$ resulting from application of Markov transition matrices $P_{T|U}$, $P_{U|Y}$ on p for each $p \in p$, and

$$\sigma_{TU} := \sum_{y\in\mathcal{Y}} \sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}} P_{T|U}(t|u) P_{U|Y}(u|y) p(y) \ |t\rangle\langle t| \otimes |u\rangle\langle u| \otimes \sigma_y$$

for given transition matrices $P_{T|U}$, $P_{U|Y}$ when

$$\sigma = \sum_{y\in\mathcal{Y}} p(y) \ |y\rangle\langle y| \otimes \sigma_y.$$

Theorem 49. *Let $\mathfrak{I}$ be a set of cqq density matrices on $\mathcal{H}_{ABE}$. It holds*

$$K_{\rightarrow,SMI}(\mathfrak{I}) = \lim_{k\to\infty} \frac{1}{k} K_{\rightarrow}^{(1)}(\mathfrak{I}^{\otimes k}), \tag{5.6}$$

where the function $K^{(1)}$ is defined in the preceding theorem.

Notice, that the inequality

$$K_{\rightarrow}(\mathfrak{I}) \leq K_{\rightarrow,SMI}(\mathfrak{I}) \tag{5.7}$$

holds for each $\mathfrak{I}$. This can be directly observed from the definitions of achievable rates given above. The next section is devoted to giving a full argument which justifies the claim of Theorem 48. We now give short outline of the proof. In a sequence of propositions with increasing level of approximation we prepare ourselves for proving the assertion

$$K_{\rightarrow}(\mathfrak{I}) \geq K_{\rightarrow}^{(1)}(\mathfrak{I}) \tag{5.8}$$

in Proposition 55. For this reason, we first design suitable protocols of suboptimal rate for the special case of a source parameterized by a full Cartesian product of probability distributions and cq channels. We improve the bounds in Proposition 54, where we derive protocols suitable for the same type of source, but including sender preprocessing of the source by a fixed Markov chain for optimization. Finally, this result is combined with a fine-grained approximation of an arbitrary regular source by a number of sources of type subject to the mentioned propositions. The proof of achievability (i.e. the lower bound in (5.6)) follows almost immediately from (5.8), since we show, that regularity of $\mathfrak{I}$ implies, for each $k \in \mathbb{N}$, regularity of the set $\mathfrak{I}^{\otimes k}$ of all k-fold tensor extensions for states from $\mathfrak{I}$.
In Section 5.3 we give a full proof of Theorem 49. The achievability part therein is derived also from the results gathered in Section 5.2. The protocol construction used for proving achievability in Theorem 48 can be employed in case of SMI. To do so we use a certain type of finite covering in Hausdorff space to decompose a general set $\mathfrak{I}$ into a finite family of regular sets. Moreover, we provide a proof to the converse assertion.
The reader may ask, whether Theorem 48 may hold also without assumption of regularity. We give a negative answer to this question in Section 5.4, where an example of a cqq set of density matrices with

$$K_{\rightarrow}(\mathfrak{I}) < K_{\rightarrow,SMI}(\mathfrak{I})$$

is established.

5.2 Secret-key distillation without state knowledge

In this chapter, we give a detailed argument to prove Theorem 48. The first assertion, we prove is on a restricted type of cqq density matrices. Assume $\mathcal{Q} \subset \mathfrak{P}(\mathcal{Y})$ be a set of probability

distributions and $\mathcal{V} \subset \mathcal{CQ}(\mathcal{Y}, \mathcal{K}_{BE})$ be a set of cq-channels. We define

$$\rho_{(p,V)} := \sum_{y\in\mathcal{Y}} p(y)\ |y\rangle\langle y| \otimes V(y) \qquad\qquad (p \in \mathcal{Q},\ V \in \mathcal{V}),$$

and the set

$$\mathfrak{J} := \left\{\rho_{(p,V)}\right\}_{(p,V)\in\mathcal{Q}\times\mathcal{V}}. \tag{5.9}$$

We set for each $V \in \mathcal{V}$, $V_B = \mathrm{tr}_{\mathcal{K}_B} \circ V$, and $V_E = \mathrm{tr}_{\mathcal{K}_E} \circ V$.

Proposition 50. *Let $\mathfrak{J}$ be the source defined in (5.9), and $\delta > 0$. There is a constant $c_1 > 0$ and a number n_0 such that for each $n > n_0$ there is an (n, M, L, μ) forward secret-key distillation protocol with*

$$\begin{aligned}
\frac{1}{n}\log M &\geq \inf_{q\in\mathcal{Q}}\left(\inf_{V\in\mathcal{V}} \chi(q, V_B) - \sup_{V\in\mathcal{V}} \chi(q, V_E)\right) - \delta \\
\frac{1}{n}\log L &\leq \sup_{p\in\mathcal{P}_{\mathfrak{J}}}\left(H(p) - \inf_{V\in\mathcal{V}} \chi(p, V_B)\right) + \delta \\
\mu &\leq 2^{-\sqrt[16]{n}c_1}.
\end{aligned}$$

Within the proof of Proposition 50 we will use some auxiliary results, we introduce first. The following lemma states, for given compound memoryless classical-quantum channel (DMcqC) existence of random codes being of constant composition (i.e. all codewords having the very same type) and equidistributed over the typical sets. The assertion is a direct consequence of coding results stated in Appendix A, where also the basic definitions regarding codes for message transmission over compound DMcq channels can be found.

Lemma 51. *Let $\mathcal{V} \subset CQ(\mathcal{X}, \mathcal{K})$ be a set of cq channels. For each $\gamma > 0$, there is a number $n_1(\gamma, \mathcal{V})$ such that for each $n > n_1(\gamma)$ and each type $\lambda \in \mathfrak{T}(n, \mathcal{X})$ the following assertion is true.*
It exists a random (n, M_λ)-code

$$\mathcal{C}(U) := (U_m, D_m(U))_{m=1}^{M_\lambda}$$

fulfilling the following three properties

1. *$U = (U_1, \ldots, U_{M_\lambda})$ is an i.i.d. sequence of random variables, such that U_m is equidistributed on T^n_λ for each $m \in [M_\lambda]$,*
2. *$M_\lambda \geq \exp\left(n\left(\inf_{V\in\mathcal{V}} \chi(\lambda, V) - \gamma\right)\right)$*
3. *$\mathbb{E}\left[\sup_{V\in\mathcal{V}} \overline{e}(\mathcal{C}(U), V^{\otimes n})\right] \leq 2^{-\sqrt[16]{n}\hat{c}}$.*

with a constant $\hat{c}(\gamma, \mathcal{V}) > 0$ (independent of λ).

Proof. We need only consider types with

$$\inf_{V\in\mathcal{V}} \chi(\lambda, V) - \gamma > 0, \tag{5.10}$$

since for all other types, the bounds in the assertion of the lemma can be satisfied by trivial coding. Setting $\delta := \frac{\gamma}{2}$ in Proposition 70 in Appendix A, ensures us, that for each large enough blocklength n and each type $\lambda \in \mathfrak{T}(n, \mathcal{X})$ the hypothesis of Proposition 71 is fulfilled with an M'_λ which fulfills

$$M'_\lambda \;\geq\; \exp\left(n\left(\inf_{V\in\mathcal{V}} \chi(\lambda, V) - \frac{\gamma}{2}\right)\right) \;>\; 2^{n\frac{\gamma}{2}} \tag{5.11}$$

and $\mu \leq 2^{-\sqrt[16]{n}c(\frac{\gamma}{2})}$. Note, that the rightmost inequality in (5.11) is satisfied because we only consider types, which fulfill the condition in (5.10). Setting $\vartheta := \frac{1}{2}$, we conclude with Proposition 71, that we find, for large enough $n \in \mathbb{N}$ and random (n, M_λ) message transmission code fulfilling the properties demanded. □

The following matrix-concentration inequality results from the powerful matrix Chernov bound [AW02] and was proven in [DW05].

Proposition 52 ([DW05], Prop. 2.4)**.** *Let $n \in \mathbb{N}$, $W \in CQ(\mathcal{X}, \mathcal{K})$, $\lambda \in \mathfrak{T}(k, \mathcal{X})$, $U := (U_1, \ldots, U_M)$ an i.i.d. sequence of random variables generically equidistributed on T^n_λ, and*

$$\sigma_{n,\lambda}(W) := \frac{1}{|T^n_\lambda|} \sum_{x^n \in T^n_\lambda} W^{\otimes n}(x^n).$$

For each $\epsilon, \delta > 0$, there is a number $k := k(\epsilon, \delta)$, such that if $n > k$, then

$$\Pr\left(\left\| \frac{1}{M} \sum_{m=1}^{M} W^{\otimes n}(U_m) - \sigma_{n,\lambda}(W) \right\| \geq \epsilon\right) \leq 2 \cdot \dim \mathcal{K}^n \cdot \exp(-M\Delta_n \cdot \epsilon)$$

holds with

$$\Delta_n := -\frac{1}{288 \ln 2} \cdot \exp(-n(\chi(\lambda, W) - \delta))$$

The next assertion will help us to approximate the set $\mathcal{V}$ assumed in Proposition 50 by a finite subset in a suitable way.

Lemma 53. *Let $\mathcal{V} \subset \mathcal{CQ}(\mathcal{X}, \mathcal{K})$ be a set of classical quantum channels. For each $\alpha \in (0, \frac{1}{e})$ exists a subset $\mathcal{V}_\alpha \subset \mathcal{V}$, which fulfills the following three conditions.*

1. $|\mathcal{V}_\alpha| \leq \left(\frac{6}{\alpha}\right)^{2|\mathcal{X}| \dim \mathcal{K}^2}$
2. *Given any $n \in \mathbb{N}$, to each $V \in \mathcal{V}$ exists a $W \in \mathcal{V}_\alpha$, such that*

   $$\|V^{\otimes n}(x^n) - W^{\otimes n}(x^n)\|_1 \leq 2n\alpha$$

   holds for each $x^n \in \mathcal{X}^n$.

3. *For each $p \in \mathfrak{P}(\mathcal{X})$, it holds*

$$\left| \min_{W \in \mathcal{V}_\alpha} \chi(p, W) - \inf_{V \in \mathcal{V}} \chi(p, V) \right| \leq 2\alpha \log \frac{\dim \mathcal{K}}{2\alpha}.$$

Proof of Proposition 50. Set

$$R := \inf_{q \in \mathcal{Q}} \left(\inf_{V \in \mathcal{V}} \chi(q, V_B) - \sup_{V \in \mathcal{V}} \chi(q, V_E) \right),$$

and let $\delta > 0$ be a number small enough for fulfilling $R - \delta > 0$, otherwise, there is nothing left to prove. Let $\frac{1}{2} > \eta > 0$, be fixed and small enough such that the inequality

$$12\eta + \log \dim \mathcal{K}_{BE} + 4h(2\eta) \leq \frac{\delta}{16}. \tag{5.12}$$

is valid. Let $n \in \mathbb{N}$ be large enough to simultaneously satisfy

$$\frac{1}{n} \leq \frac{1}{16}\delta \quad \text{and} \quad \frac{1}{n} \leq 2\eta. \tag{5.13}$$

Define

$$\mathfrak{T}_n := \mathfrak{T}(n, \mathcal{Y}) \cap \mathcal{Q}_\eta,$$

where $\mathcal{Q}_\eta := \{q \in \mathfrak{P}(\mathcal{Y}) : \ \exists p \in \mathcal{Q} : \|p - q\|_1 \leq \eta\}$ is the η-blowup of $\mathcal{Q}$ regarding the variational distance. We set for each probability distribution $q \in \mathfrak{P}(\mathcal{Y})$

$$\begin{aligned} \chi_{B,q} &:= \inf_{V \in \mathcal{V}} \chi(q, V_B), \\ \chi_{E,q} &:= \sup_{V \in \mathcal{V}} \chi(q, V_E), \\ \chi_q &:= \chi_{B,q} - \chi_{E,q}, \end{aligned}$$

and

$$\chi_n := \min_{\lambda \in \mathfrak{T}_n} \chi_\lambda.$$

Our choice of η and n implies

$$d_H(\mathfrak{T}_n, \mathcal{Q}) \leq d_H(\mathfrak{T}_n, \mathcal{Q}_\eta) + d_H(\mathcal{Q}_\eta, \mathcal{Q}) \leq \frac{1}{2n} + \eta \leq 2\eta, \tag{5.14}$$

where the first inequality above is the triangle inequality for the Hausdorff distance, and the second is by (5.13). From (5.14), and (5.12) together with twofold application of Lemma 75, we infer

$$|\chi_n - R| \leq \frac{\delta}{16}. \tag{5.15}$$

Set, for each $\lambda \in \mathfrak{T}_n$

$$\begin{aligned} L_\lambda &:= \left\lfloor \exp\left(n(H(\lambda) - \chi_{B,\lambda} + \tfrac{3}{4}\delta)\right)\right\rfloor, \\ S_\lambda &:= \left\lceil \exp\left(n(\chi_{E,\lambda} + \chi_\lambda - \chi_n + \tfrac{3}{4}\delta)\right)\right\rceil, \text{ and} \\ M &:= \left\lfloor \exp\left(n(R - \delta)\right)\right\rfloor. \end{aligned}$$

The above definitions, together with (5.15) and the second inequality of (5.13) the bounds

$$M \cdot S_\lambda \leq \exp\left(n(\chi_{B,\lambda} - \tfrac{7}{8}\delta)\right), \tag{5.16}$$

$$M \cdot L_\lambda \leq \exp\left(n(H(\lambda) - \chi_{E,\lambda} - \tfrac{7}{8}\delta)\right), \text{ and} \tag{5.17}$$

$$\Gamma_\lambda := \frac{L_\lambda \cdot S_\lambda \cdot M}{|T_\lambda^n|} \geq 2^{n\frac{11}{8}\delta}. \tag{5.18}$$

are valid. The strategy for the rest of the proof is the following. We will in a first step, generate a class of one-way-secret key distribution protocols for $\mathfrak{J}$, and then show, that with high probability, the protocols meet the properties demanded.
Define for each $\lambda \in \mathfrak{T}_n$ a random matrix

$$U^{(\lambda)} := (U^\lambda_{lms})_{(l,m,s)\in[L_\lambda]\times[M]\times[S_\lambda]}$$

with all entries being independent and generically equidistributed on T^n_λ. We collect the matrices defined above in an independent family

$$\mathbf{U} := \{U^{(\lambda)}\}_{\lambda\in\mathfrak{T}_n}.$$

Define, for each $y^n \in \mathcal{Y}^n$, $\lambda \in \mathfrak{T}_n$ a random set

$$A_\lambda(y^n, \mathbf{U}) := \left\{(\lambda, l, m, s) : \ U^\lambda_{lms} = y^n\right\}.$$

Obviously, the sets defined above fulfill for each outcome $\mathbf{u}$ of $\mathbf{U}$, $\lambda \in \mathfrak{T}_n$ the following relations

$$\begin{aligned} A_\lambda(y^n, \mathbf{u}) &= \varnothing && (y^n \notin T^n_\lambda), \\ A_\lambda(y^n, \mathbf{u}) \cap A_\lambda(z^n, \mathbf{u}) &= \varnothing && (y^n \neq z^n), \\ \text{and} \bigcup_{y^n\in T^n_\lambda} A_\lambda(y^n, \mathbf{u}) &= \{\lambda\} \times [L_\lambda] \times [M] \times [S_\lambda]. \end{aligned} \tag{5.19}$$

We regard, for each $\lambda \in \mathfrak{T}_n$ and $l \in [L_\lambda]$,

$$U_{\lambda,l} := (U^\lambda_{lms})_{(m,s)\in[M]\times[S_\lambda]}$$

as a random i.i.d. constant composition codebook of size $M \cdot S_\lambda$ with codewords equidistributed over T^n_λ. Since we have the bound in (5.16), we know from Lemma 51 , that there is a random $(n, M \cdot S_\lambda)$ constant composition code

$$\mathcal{C}_{\lambda,l}(U_{\lambda,l}) := (U^\lambda_{lms}, D^\lambda_{lms})_{(m,s)\in[M]\times[S_\lambda]}$$

for the compound DMcqC generated by $\mathcal{V}_B := \{V_B : V \in \mathcal{V}\}$ which has expected average error bounded

$$\mathbb{E}\left[\sup_{V\in\mathcal{V}_\lambda} \overline{e}(\mathcal{C}_{\lambda,l}(U_{\lambda,l}), V_B^{\otimes n})\right] \leq 2^{-\sqrt[16]{n}\hat{c}} =: \beta_0 \tag{5.20}$$

with a strictly positive constant $\hat{c}$ independent of λ. Define, for $\beta_3 > 0$, $\lambda \in \mathfrak{T}_n$ a random set

$$B_\lambda(\mathbf{U}, \beta_3) := \left\{l \in [L_\lambda] : \max_{V\in\mathcal{V}} \overline{e}\left(\mathcal{C}_{\lambda,l}(U_{\lambda,l}), V_B^{\otimes n}\right) < \beta_3\right\},$$

which collects all indices $l \in [L_\lambda]$, such that $\mathcal{C}_{\lambda,l}$ is β_3-good regarding the average error criterion. Define a random stochastic matrix

$$T_{\mathbf{U}} : \mathcal{X} \to \mathfrak{P}\left(\mathfrak{T}(n,\mathcal{Y}) \times [L_\lambda] \times [M] \times [S_\lambda]\right)$$

with entries

$$T_{\mathbf{U}}(\lambda, l, m, s|y^n) := \begin{cases} |A_\lambda(y^n, \mathbf{U})|^{-1} & \text{if } (\lambda, l, m, s) \in A_\lambda(y^n, \mathbf{U}) \\ 0 & \text{otherwise} \end{cases}$$

for each $\lambda \in \mathfrak{T}_n$. The values of $T_{\mathbf{U}}(\lambda, l, m, s|y^n)$ with λ being not in $\mathfrak{T}_n$ will be of no special interest for us, so they may be defined in any consistent way. Let, for each $\lambda \in \mathfrak{T}_n$, $V \in \mathcal{V}$

$$\sigma_\lambda(V) := \frac{1}{|T_\lambda^n|} \sum_{y^n \in T_\lambda^n} V_E^{\otimes n}(y^n).$$

Note, that

$$\sigma_\lambda(V) = \mathbb{E}\left[V_E^{\otimes n}(U_{lms}^\lambda)\right] \tag{5.21}$$

holds for all $(l, m, s) \in [L_\lambda] \times [M] \times [S_\lambda]$. Define, for $\lambda \in \mathfrak{T}_n$, $\beta_1, \beta_2, \beta_3 > 0$ and each outcome $\mathbf{u}$ of $\mathbf{U}$, the following sets.

$$C_\lambda^{(1)}(\beta_1) := \{\mathbf{u} : \forall y^n \in T_\lambda^n : (1-\beta_1)\Gamma_\lambda \leq |A_\lambda(y^n, \mathbf{u})| \leq (1+\beta_1\Gamma_\lambda)\}$$

$$C_\lambda^{(2)}(\beta_2) := \left\{\mathbf{u} : \forall (l,m) \in [L_\lambda] \times [M], V \in \mathcal{V} : \|\frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} V_E^{\otimes n}(u_{lms}^\lambda) - \sigma_\lambda(V)\|_1 \leq \beta_2\right\}$$

$$C_\lambda^{(3)}(\beta_3) := \{\mathbf{u} : |B_\lambda(\mathbf{u}, \beta_3)| \geq (1-2\beta_3)L_\lambda\}, \quad \text{and}$$

$$A := \bigcap_{\lambda\in\mathfrak{T}_n} \bigcap_{i=1}^{3} C_\lambda^{(i)}(\beta_i).$$

Eventually, we will show, that if an outcome $\mathbf{u}$ of $\mathbf{U}$ is an element of A, it generates a suitable protocol for our needs. First we make sure, that for the right choice of parameters and each large enough blocklength, A is actually nonempty, which we do by actually bounding the r.h.s. of

$$\Pr(A^c) \leq \sum_{i=1}^{3} \sum_{\lambda\in\mathfrak{T}_n} \Pr\left(C_\lambda^{(i)}(\beta_i)^c\right) \tag{5.22}$$

away from one. In the following we separately deriva bound on each of the summands on the right hand side of (5.22). Let λ be a type from $\mathcal{T}_n$. Note, that

$$|A_\lambda(y^n, \mathbf{u})| = \sum_{l=1}^{L_\lambda} \sum_{m=1}^{M} \sum_{s=1}^{S_\lambda} \mathbb{1}_{y^n}(u^\lambda_{lms})$$

holds, where $\mathbb{1}$ is the indicator function, therefore,

$$\mathbb{E}[|A_\lambda(y^n, \mathbf{u})|] = \frac{L_\lambda M S_\lambda}{|T^n_\lambda|} = \Gamma_\lambda \geq 2^{-n\frac{11}{8}\delta}$$

where the rightmost inequality above results from (5.18). We infer

$$\begin{aligned}
\Pr\left(C^{(1)}_\lambda(\beta_1)^c\right) &= \sum_{y^n \in \mathcal{Y}^n} \Pr\left(\{\mathbf{u} : \ |A_\lambda(y^n, \mathbf{u})| \notin ((1-\beta_1)\Gamma_\lambda, (1+\beta_1)\Gamma_\lambda)\}\right) \\
&\leq 2|\mathcal{Y}|^n \cdot \exp\left(-\Gamma_\lambda \beta_1^2/(2\ln 2)\right) \\
&\leq 2 \cdot \exp(-2^{n\frac{9}{8}\delta})
\end{aligned}$$

for large enough blocklength n, where the first inequality above is by Chernov-bounding with Proposition 72, and the second is by (5.18) together with the choice

$$\beta_1 = 2^{-n\frac{\delta}{8}}. \tag{5.23}$$

To bound the summands with $i = 2$, we choose an approximating set $\hat{\mathcal{V}}_n$ for $\mathcal{V}$ according to Lemma 53 with parameter

$$\alpha := 2^{-\sqrt[16]{n}\hat{c}/(16|\mathcal{Y}|\dim \mathcal{K}^2_{BE})}.$$

which is possible with cardinality

$$|\hat{\mathcal{V}}_n| \leq 2^{-\sqrt[16]{n}\frac{\hat{c}}{4}}$$

as long as n is large enough. Let for given $V \in \mathcal{V}$, $W \in \hat{\mathcal{V}}_n$ be a channel, such that $\|V(x) - W(x)\| \leq 2\alpha$. It holds for each $\lambda \in \mathfrak{T}_n$, $l \in [L_\lambda]$, $m \in [M]$

$$\begin{aligned}
\|\frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} V_E^{\otimes n}(y^n) - \sigma_\lambda(V)\|_1 &\leq \|\frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} W_E^{\otimes n}(y^n) - \sigma_\lambda(W)\|_1 \qquad (5.24) \\
&\quad + \|\frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} V_E^{\otimes n}(y^n) - \frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} W_E^{\otimes n}(y^n)\|_1 \\
&\quad + \|\sigma_\lambda(V) - \sigma_\lambda(W)\|_1 \\
&\leq \|\frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} W_E^{\otimes n}(y^n) - \sigma_\lambda(W)\|_1 + 2n\alpha. \qquad (5.25)
\end{aligned}$$

If we now choose

$$\beta_2 = 4n\alpha \tag{5.26}$$

We can bound

$$\begin{aligned}
&\Pr\left(C_\lambda^{(2)}(\beta_2)^c\right)\\
&= \Pr\left(\exists (l,m), V \in \mathcal{V} : \left\| \frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} V_E^{\otimes n}(u_{lms}^\lambda) - \sigma_\lambda(V)\right\|_1 > \beta_2\right)\\
&\leq \Pr\left(\exists (l,m), V \in \hat{\mathcal{V}}_n : \left\| \frac{1}{S_\lambda}\sum_{s=1}^{S_\lambda} V_E^{\otimes n}(u_{lms}^\lambda) - \sigma_\lambda(V)\right\|_1 > \frac{\beta_2}{2}\right)\\
&\leq 4\cdot L_\lambda M|\hat{\mathcal{V}}_n|\cdot(\dim \mathcal{K}_E)^n \exp\left(-S_\lambda\cdot 2^{-n(\chi_{E,\lambda}-\frac{\delta}{4})}\frac{\beta_2}{576\ln 2}\right)\\
&\leq \exp\left(2^{-n\frac{\delta}{4}}\right)
\end{aligned}$$

where the first inequality is by (5.25), the second by application of Proposition 52, and the last inequality holds for each large enough blocklength. At last,

$$\begin{aligned}
\Pr\left(C_\lambda^{(3)}(\beta_3)^c\right) &= \Pr\left(\frac{1}{L_\lambda}\sum_{l\in L_\lambda} \mathbb{1}_{B_\lambda(\beta_3,\mathbf{U})^c}(l) \geq 2\beta_3\right)\\
&< \frac{\mathbb{E}\left[\mathbb{1}_{B_\lambda(\beta_3,\mathbf{U})^c}(l)\right]}{2\beta_3}\\
&< \frac{\beta_0}{2\beta_3^2}.
\end{aligned} \tag{5.27}$$

The first inequality above is Markov's inequality applied, the second can be justified as follows. It holds

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_{B_\lambda(\beta_3,\mathbf{U})^c}(l)\right] &= \Pr\left(B_\lambda(\mathbf{U},\beta_3)(l)^c\right))\\
&= \Pr\left(\max_{V\in\mathcal{V}} \overline{e}(\mathcal{C}_{\lambda,l}, V_B^{\otimes n}) \geq \beta_3\right)\\
&\leq \frac{\mathbb{E}\left[\max_{V\in\mathcal{V}} \overline{e}(\mathcal{C}_{\lambda,l}, V_B^{\otimes n})\right]}{\beta_3}\\
&\leq \frac{\beta_0}{\beta_3}.
\end{aligned} \tag{5.28}$$

The first inequality above is again by Markov-bounding. The second is by (5.20). By combination of the estimate in (5.27), and the choice

$$\beta_3 = 2^{-\sqrt[16]{n}\hat{c}/4} \tag{5.29}$$

we yield, for large enough blocklength

$$\Pr\left(C_\lambda^{(3)}(\beta_3)^c\right) \leq 2^{-\sqrt[16]{n}\hat{c}}. \tag{5.30}$$

Combining all the bounds derived above with (5.22), choosing the blocklength large enough, we arrive with our choice of the parameters β_i, $1 \leq i \leq 3$ at

$$\Pr\left(A^c\right) \leq |\mathfrak{T}_n| 2^{-\sqrt[16]{n}\tilde{c}}$$

with a strictly positive constant $\tilde{c}$ if n is large enough. Since $|\mathfrak{T}_n| \leq |\mathfrak{T}(n,\mathcal{Y})| \leq (n+1)^{|\mathcal{Y}|}$, we have for each large enough blocklength

$$\Pr\left(A^c\right) \leq \frac{1}{2},$$

which implies, that A is nonempty. Define

$$L := \max_{\lambda \in \mathfrak{T}_n} L_\lambda, \quad \text{and} \quad S := \max_{\lambda \in \mathfrak{T}_n} S_\lambda.$$

We choose any $\mathbf{u} \in A$ and define a stochastic map

$$\begin{aligned} T:\ \mathcal{Y}^n &\to \mathfrak{T}(n,\mathcal{Y}) \times [L] \times [M] \\ y^n &\mapsto T(\lambda,l,m|y^n) := \sum_{s=1}^{S} T_{\mathbf{u}}(\lambda,l,m,s|y^n). \end{aligned}$$

and

$$D := (D_{lm}^\lambda)_{\lambda \in \mathfrak{T}(n,\mathcal{Y}), l \in [L], m \in [M]},$$

where

$$D_{lm}^\lambda := \sum_{s=1}^{S} D_{lms}^\lambda \qquad\qquad (\lambda,l,m) \in \mathfrak{T}(n,\mathcal{Y}) \times [L] \times [M]$$

with D_{lms}^λ being the decoding set with index (m,s) from the code $\mathcal{C}_{\lambda,l}(\mathbf{u})$. Note that some entries of $T_{\mathbf{u}}$ as well as some of the decoding matrices are have been not defined yet (e.g. $L > L_\lambda$ may occur for some λ)), we populate the undefined entries of $T_{\mathbf{u}}$ with zeros, and add zero matrices, and add arbitrary but consistently, where decoding matrices are undefined.
To fit the above objects to the definition of a one-way secret key distillation protocol, we consider each public message as a tuple $\mathbf{l} = (\lambda, l)$. With the above definitions, $\mathcal{D} := (T, D)$ is an (n, M, L) secret key distillation protocol with

$$\frac{1}{n} \log M \geq R - \delta \tag{5.31}$$

by definition of M. It remains, to show, that actually the bound on the performance μ stated is fulfilled. We fix an arbitrary member

$$\rho_t = \sum_{y \in \mathcal{Y}} p(y) \left|y\right\rangle \left\langle y\right| \otimes V(y)$$

from $\mathfrak{J}$. We first show, that

$$\Pr(K_t \neq K'_t) \leq \mu. \tag{5.32}$$

holds. By construction, we have $\mathfrak{T}(n,\mathcal{Y})_\setminus \subset Q_\eta^c$, which together with well known type-bounds implies

$$p^n\left(\bigcup_{\lambda\in\mathfrak{T}(n,\mathcal{Y})_\setminus\mathfrak{T}_n} T_\lambda^n\right) \leq 2^{-nc\eta^2}$$

with a universal constant $c > 0$. It holds

$$\begin{aligned}\Pr(K_t \neq K'_t) &= \sum_{y^n\in\mathcal{Y}^n} p^n(y^n)\cdot\Pr\left(K_t \neq K'_t | Y_t^n = y^n\right)\\ &\leq \sum_{\lambda\in\mathfrak{T}_n}\sum_{y^n\in T_\lambda^n} p^n(y^n)\cdot\Pr\left(K_t \neq K'_t | Y_t^n = y^n\right) + 2^{-nc\eta^2}.\end{aligned} \tag{5.33}$$

We upper-bound for each $\lambda \in \mathfrak{T}_n$ the corresponding summand on the r.h.s. of (5.33). For each $y^n \in T_\lambda^n$, we have

$$\begin{aligned}&p^n(y^n)\cdot\Pr\left(K_t \neq K'_t | Y_t^n = y^n\right)\\ &= \sum_{m=1}^M \sum_{m'\neq m} p^n(y^n)\cdot\Pr\left(K_t = m, K'_t = m' | Y_t^n = y^n\right)\\ &= \sum_{m=1}^M \sum_{m'\neq m}\sum_{l=1}^L p^n(y^n)\cdot\Pr\left(K_t = m, K'_t = m', \Lambda_t = (\lambda,l) | Y_s^n = y^n\right)\\ &= \sum_{m=1}^M \sum_{m'\neq m}\sum_{l=1}^{L_\lambda} \operatorname{tr}\left(D_{lm'}^\lambda V_B^{\otimes n}(y^n)\right)\cdot p^n(y^n)\cdot T(\lambda,l,m|y^n)\\ &= \sum_{m=1}^M \sum_{m'\neq m}\sum_{l=1}^{L_\lambda}\sum_{s,s'=1}^{S} \operatorname{tr}\left(D_{lm's'}^\lambda V_B^{\otimes n}(y^n)\right)\cdot p^n(y^n)\cdot T_{\mathbf{u}}(\lambda,l,m,s|y^n).\end{aligned} \tag{5.34}$$

On the r.h.s. of (5.34), only the summands survive, where (λ,l,m,s) is in $A_\lambda(y^n,\mathbf{u})$ by definition of $T_{\mathbf{u}}$. For the nonzero summands, we can estimate

$$T_{\mathbf{u}}(\lambda,l,m,s|y^n) = \frac{1}{|A_\lambda(y^n,\mathbf{u})|} \leq ((1-\beta_3)\Gamma_\lambda)^{-1} \leq 2\left(\frac{|T_\lambda^n|}{L_\lambda\cdot M\cdot S_\lambda}\right). \tag{5.35}$$

The left of the above inequalities is because $\mathbf{u} \in A$, the right holds, if n is large enough. We define the abbreviation $A_\lambda(y^n) := A_\lambda(y^n,\mathbf{u})$. Counting only the nonzero summands, and using the estimate in (5.35), we yield

$$\begin{aligned}&p^n(y^n)\cdot\Pr\left(K_t \neq K'_t | Y_t^n = y^n\right)\\ &\leq \sum_{\substack{(l,m,s):\\(\lambda,l,m,s)\in A_\lambda(y^n)}} \sum_{m'\neq m}\sum_{s'=1}^{S} \operatorname{tr}\left(D_{lm's'}^{(\lambda)} V_B^{\otimes n}(u_{lms}^\lambda)\right)\frac{2|T_\lambda^n|\cdot p^n(u_{lms}^\lambda)}{L_\lambda S_\lambda M}\\ &\leq \frac{2}{L_\lambda M_\lambda S_\lambda}\sum_{\substack{(l,m,s):\\(\lambda,l,m,s)\in A_\lambda(y^n)}} \sum_{m'\neq m}\sum_{s'=1}^{S} \operatorname{tr}\left(D_{lm's'}^{(\lambda)} V_B^{\otimes n}(u_{lms}^\lambda)\right),\end{aligned}$$

where in the last inequality, we noted, since u^{λ}_{lms} is of type λ, $|T^n_\lambda| \cdot p^n(u^{\lambda}_{lms}) = p^n(T^n_\lambda) \leq 1$ holds. Since, for each type $\lambda \in \mathfrak{T}_n$

$$\bigcup_{y^n \in T^n_\lambda} A_\lambda(y^n) = \{\lambda\} \times [L_\lambda] \times [M] \times [S_\lambda]$$

holds by construction (see (5.19)), we have (with some rearrangements of terms)

$$\begin{aligned}\sum_{y^n \in T^n_\lambda} p^n(y^n) \cdot \Pr\left(K_t \neq K'_t | Y^n_t = y^n\right) &= \frac{2}{L_\lambda} \sum_{l=1}^{L_\lambda} \frac{1}{S_\lambda M_\lambda} \sum_{m=1}^{M} \sum_{m' \neq m} \sum_{s,s'=1}^{S} \operatorname{tr}\left(D^{(\lambda)}_{lm's'} V_B^{\otimes n}(u^{\lambda}_{lms})\right) \\ &= \frac{2}{L_\lambda} \sum_{l=1}^{L_\lambda} \overline{e}(\mathcal{C}_{\lambda,l}(u_{\lambda,l}), V_B^{\otimes n}) \\ &\leq \frac{2}{L_\lambda}(L_\lambda \cdot \beta_3 + L_\lambda \cdot 2\beta_3) = 6\beta_3. \end{aligned} \tag{5.36}$$

The last inequality holds, because we have chosen our protocol in a way, that for at least a fraction of $1 - 2\beta_3$, the code $C^{\lambda,l}(u_{\lambda,l})$ is β_3-good regarding the average error criterion (i.e. $\mathbf{u}$ is in $A \subset C^{(3)}_\lambda(\beta_3)$).
Collecting inequalities, we arrive at

$$\Pr(K_t \neq K'_t) \leq (n+1)^{|\mathcal{Y}|} \cdot 6\beta_3 + 2^{-nc\eta^2} \leq 2^{-\sqrt[16]{n}\hat{c}/8} \tag{5.37}$$

for large enough n by (5.29). Next, we show, that the key is almost equidistributed. For each $\lambda \in \mathfrak{T}(n, \mathcal{Y})$, we consider the probability distribution $P_{K_t,\lambda}$ on $[M]$, given by

$$\begin{aligned} P_{K_t,\lambda}(m) &:= \sum_{y^n \in T^n_\lambda} \frac{\Pr(K_t = m | Y^n_t = y^n)}{|T^n_\lambda|} \\ &= \sum_{y^n \in T^n_\lambda} \sum_{l=1}^{L_\lambda} \sum_{s=1}^{S_\lambda} \frac{T_u(\lambda, l, m, s | y^n)}{|T^n_\lambda|} \\ &= \sum_{l=1}^{L_\lambda} \sum_{s=1}^{S_\lambda} \frac{T_u(\lambda, l, m, s | u^{\lambda}_{lms})}{|T^n_\lambda|} \end{aligned} \tag{5.38}$$

Using the properties of the protocol constructed together with (5.38), we arrive at

$$\frac{1}{1+\beta_1}\frac{1}{M} \leq P_{K_t,\lambda}(m) \leq \frac{1}{1-\beta_1}\frac{1}{M},$$

for each $\lambda \in \mathfrak{T}_n$, from which we infer, that

$$\|P_{K,\lambda} - \pi_{[M]}\|_1 \leq 2\beta_1.$$

is true for all $\lambda \in \mathfrak{T}_n$. We conclude

$$\|P_{K_t} - \pi_{[M]}\|_1 \leq \sum_{\lambda \in \mathfrak{T}_n} p^n(T^n_\lambda) \, \|P_{K_t,\lambda} - \pi_{[M]}\|_1 \leq 2\beta_1 + 2 \cdot 2^{-nc\eta^2} \leq 3\beta_1,$$

where the last inequality holds if n is large enough. This implies

$$H(K_t) \geq \log M - 3\beta_1 \log \frac{M}{3\beta_1} \geq \log M - \frac{\mu}{2}$$

if n is large enough. It remains to bound $I(K; E^n, \Lambda, \rho_{\Lambda K E^n, t})$. We will actually show, that $\rho_{\Lambda K E^n, t}$ is close to a state γ_t whith $I(K; E^n, \Lambda, \gamma_t) = 0$. Define

$$\gamma_t := \sum_{\lambda \in \mathfrak{T}(n,\mathcal{Y})} p^n(T_\lambda^n) \left|\lambda\right\rangle \left\langle\lambda\right| \otimes \gamma_{t,\lambda}$$

where set for each $\lambda \in \mathfrak{T}(n, \mathcal{Y})$

$$\gamma_{t,\lambda} := \frac{1}{L_\lambda \cdot M} \sum_{l=1}^{L_\lambda} \sum_{m=1}^{M} \left|l \otimes m\right\rangle \left\langle l \otimes m\right| \otimes \sigma_\lambda(V).$$

We write $\rho_{\Lambda K E^n, t}$ in the form

$$\rho_{\Lambda K E^n, t} = \sum_{\lambda \in \mathfrak{T}(n,\mathcal{Y})} p^n(T_\lambda^n) \left|\lambda\right\rangle \left\langle\lambda\right| \otimes \tilde{\rho}_{t,\lambda},$$

where we defined

$$\tilde{\rho}_{t,\lambda} := \sum_{y^n \in T_\lambda^n} \sum_{l=1}^{L_\lambda} \sum_{m=1}^{M} \sum_{s=1}^{S_\lambda} \frac{T_u(\lambda, l, m, s|y^n)}{|T_\lambda^n|} \left|l \otimes m\right\rangle \left\langle l \otimes m\right| \otimes V_E^{\otimes n}(y^n).$$

We first consider $\lambda \in \mathfrak{T}_n$. Note, that if (λ, l, m, s) is a member of $A_\lambda(\mathbf{u}, y^n)$,

$$\frac{1}{1+\beta_1} \Gamma_\lambda^{-1} \leq T_{\mathbf{u}}(\lambda, l, m, s|y^n) \leq \frac{1}{1-\beta_1} \Gamma_\lambda^{-1},$$

while being zero otherwise, which implies

$$\sum_{y^n \in t_\lambda^n} \sum_{l=1}^{L_\lambda} \sum_{m=1}^{M} \sum_{s=1}^{S_\lambda} \left| \frac{T_{\mathbf{u}}(\lambda, l, m, s|y^n)}{|T_\lambda^n|} - \frac{1}{L_\lambda M |T_\lambda^n| S_\lambda} \right| \leq 2\beta_1.$$

Also, we know, that for each $l \in L_\lambda$, $s \in S_\lambda$

$$\left\| \frac{1}{S_\lambda} \sum_{s=1}^{S_\lambda} V_E^{\otimes n}(y^n) \right\|_1 \leq 2\beta_2.$$

We define

$$\tau := \sum_{y^n \in T_\lambda^n} \sum_{l=1}^{L_\lambda} \sum_{m=1}^{M} \sum_{s=1}^{S_\lambda} \frac{1}{L_\lambda \cdot M \cdot S_\lambda \cdot |T_\lambda^n|} \left|l \otimes m\right\rangle \left\langle l \otimes m\right| \otimes V_E^{\otimes n}(y^n),$$

and obtain

$$\begin{aligned}\|\tilde{\rho}_{t,\lambda} - \gamma_{t,\lambda}\|_1 &\leq \|\tilde{\rho}_{t,\lambda} - \tau\|_1 + \|\tau - \gamma_{t,\lambda}\|_1 \\ &\leq \sum_{y^n \in T_\lambda^n} \sum_{l=1}^{L_\lambda} \sum_{m=1}^{M} \sum_{s=1}^{S_\lambda} \left| \frac{T_{\mathbf{u}}(\lambda, l, m, s|y^n)}{|T_\lambda^n|} - \frac{1}{L_\lambda M |T_\lambda^n| S_\lambda} \right| + \left\| \sum_{y^n \in T_\lambda^n} V_E^{\otimes n}(y^n) - \sigma_\lambda(V) \right\|_1 \\ &\leq 2(\beta_1 + \beta_2).\end{aligned}$$

Therefore,

$$\begin{aligned}\|\rho_{K\Lambda E^n,t} - \gamma_t\|_1 &\leq \sum_{\lambda \in \mathfrak{T}(n,\mathcal{Y})} p^n(T_\lambda^n)\ \|\tilde{\rho}_{t,\lambda} - \gamma_{t,\lambda}\|_1 \\ &\leq 2(\beta_1 + \beta_2) + 2^{-nc\eta^2} \\ &\leq \frac{\mu}{12}\end{aligned}$$

By using the well-known Alicki-Fannes type bound for the quantum mutual information, we infer

$$I(K; \Lambda, E^n, \rho_{K\Lambda E^n,t}) \leq I(K; \Lambda, E^n, \gamma_t) + \frac{1}{2}\mu \log(L \cdot \dim \mathcal{K}_E^{\otimes n}) + h\left(\frac{\mu}{12}\right) \leq \mu,$$

where the last inequality is by the fact, that γ_t is an uncorrelated state, together with a large enough choice of n. □

Next, we prove an achievability result for the same simple kind of compound source as in the previous proposition, but the lower bound on the key rate derived including possible preprocessing of the source outputs for the sender by Markov chains. For each set A of probability distributions on $\mathcal{Y}$, we denote its diameter (regarding the variational distance) by

$$\operatorname{diam}(A) := \sup\{\|q - q'\|_1 : \ q, q' \in A\}$$

Proposition 54. *Let $\mathcal{P} \subset \mathfrak{P}(\mathcal{Y})$ be a set of probability distributions, $\operatorname{diam}(\mathcal{P}) \leq \Delta$, $\mathcal{V} \subset \mathcal{CQ}(\mathcal{Y}, \mathcal{K}_{BE})$, and $\mathcal{U}, \mathcal{T}$ finite alphabets. Define*

$$\mathfrak{I} := \left\{ \rho_{(p,V)} := \sum_{y \in \mathcal{Y}} p(y)\ |y\rangle\langle y| \otimes V(y) \right\}_{(p,V) \in \mathcal{P} \times \mathcal{V}}.$$

For each $P_{T|U} : \mathcal{U} \to \mathfrak{P}(\mathcal{T})$, $P_{U|Y} : \mathcal{Y} \to \mathfrak{P}(\mathcal{U})$ stochastic matrices, and $\delta > 0$, there is a number n_0, such that for each $n > n_0$ we find an (n, M, L, μ)-secret-key distillation protocol for $\mathfrak{I}$ which fulfills

$$\begin{aligned}\mu &\leq 2^{-\sqrt[16]{n} c_2} \\ \frac{1}{n} \log L &\leq \sup_{p \in \mathcal{P}} \inf_{\rho \in \mathfrak{I}_p} S(U|BT, \tilde{\rho}) + \log|\mathcal{T}| \\ \frac{1}{n} \log M &\geq \inf_{p \in \mathcal{P}} \left(\inf_{\rho \in \mathfrak{I}_p} I(U; B|T, \tilde{\rho}) - \sup_{\rho \in \mathfrak{I}_p} I(U; E|T, \tilde{\rho}) \right) - \delta - 12\Delta \log|\mathcal{U}| - 4h(\Delta), \qquad (5.39)\end{aligned}$$

where $h(x) = -x\log x - (1-x)\log(1-x)$, $(x \in (0,1))$ *is the binary entropy, and* c_2 *is a strictly positive constant. We used the definition*

$$\tilde{\rho} := \sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}} \sum_{y\in\mathcal{Y}} P_{T|U}(t|u) P_{U|Y}(u|y) p(y) \ |u\rangle\langle u| \otimes |t\rangle\langle t| \otimes V(y)$$

for each state

$$\rho = \sum_{y\in\mathcal{Y}} p(y) \ |y\rangle\langle y| \otimes V(y) \qquad (p \in \mathcal{P}, V \in \mathcal{V}).$$

Proof. For the proof, we define a set of effective cqq density matrices $\hat{\mathfrak{I}}$ on which we apply Proposition 50 from which we derive existence of a certain forward secret-key distillation protocol for $\hat{\mathfrak{I}}$. Afterwards, we show, that this protocol can be modified to a forward secret-key distillation protocol for $\mathfrak{I}$ which has the stated properties. Define, for each $p \in \mathcal{P}$ a probability distribution $q_p \in \mathfrak{P}(\mathcal{U})$ by

$$q_p(u) := \sum_{y\in\mathcal{Y}} P_{U|Y}(u|y) p(y) \qquad (u \in \mathcal{U}),$$

a stochastic matrix $W_p : \mathcal{U} \to \mathfrak{P}(\mathcal{Y})$ by

$$W_p(y|u) := \begin{cases} \frac{p(y) P_{U|Y}(u|y)}{q_p(u)} & \text{if } q_p(u) > 0 \\ \frac{1}{|\mathcal{Y}|} & \text{otherwise,} \end{cases}$$

and a classical-quantum channel $\hat{V}_p : \mathcal{U} \to \mathcal{S}(\mathcal{K}_{BE})$ by

$$\hat{V}_p(u) := \sum_{y\in\mathcal{Y}} W_p(u|y) V(y) \qquad (V \in \mathcal{V}).$$

Moreover, we define, introducing spaces $\mathcal{K}_{B'} = \mathcal{K}_{E'} = \mathbb{C}^{|\mathcal{T}|}$ a classical-quantum channel $\tilde{V} : \mathcal{U} \to \mathcal{S}(\mathcal{K}_{B'} \otimes \mathcal{K}_{E'})$ by

$$\tilde{V}(u) := \sum_{t\in\mathcal{T}} P_{T|U}(t|u) \ |t\rangle\langle t| \otimes |t\rangle\langle t| \qquad (u \in \mathcal{U}). \tag{5.40}$$

Define for each $(p, p', V) \in \mathcal{P} \times \mathcal{P} \times \mathcal{V}$ a state

$$\hat{\rho}_{(p,p',V)} := \sum_{u\in\mathcal{U}} q_p(u) \ |u\rangle\langle u| \otimes \hat{V}_{p'}(u) \otimes \tilde{V}(u).$$

We define a set of classical-quantum channels $\hat{\mathcal{V}} := \{\hat{V}_{p'} \otimes \tilde{V} : \ p' \in \mathcal{P}, V \in \mathcal{V}\}$ and a set

$$\hat{\mathfrak{I}} := \left\{\rho_{(p,p',V)} : \ p, p' \in \mathcal{P}, V \in \mathcal{V}\right\} \subset \mathcal{S}_{cqq}(\mathbb{C}^{|\mathcal{U}|} \otimes \mathcal{K}_{BE} \otimes \mathcal{K}_{B'E'})$$

of cqq density matrices. Note, that $\hat{\mathfrak{I}}$ meets the specifications of Proposition 50 (the states in $\hat{\mathfrak{I}}$ are parameterized by $\mathcal{P} \times \hat{\mathcal{V}}$). We apply Proposition 50 on $\hat{\mathfrak{I}}$ and infer in case of sufficiently large

blocklength existence of an $(n, M, \hat{L}, \mu)$ secret key distillation protocol $\hat{\mathcal{D}} = (\hat{T}, \hat{D})$ for $\hat{\mathfrak{I}}$ with a stochastic matrix

$$\hat{T} : \mathcal{U} \to [M] \times [\hat{L}]$$

and

$$\hat{D} := \left\{\hat{D}_{lm}\right\}_{(l,m)\in[\hat{L}]\times[M]} \subset \mathcal{L}(\mathcal{K}_{BB'}^{\otimes n})$$

being a POVM such that the key rate is lower-bounded by

$$\begin{aligned}\frac{1}{n}\log M &\geq \inf_{p\in\mathcal{P}}\left(\inf_{(p,p',V)\in\{p\}\times\mathcal{P}\times\mathcal{V}}\chi(q_p, \hat{V}_{B,p'}\otimes\tilde{V}_{B'}) - \sup_{(p,p',V)\in\{p\}\times\mathcal{P}\times\mathcal{V}}\chi(q_p, \hat{V}_{E,p'}\otimes\tilde{V}_{E'})\right) - \delta \\ &\geq \inf_{(p,p',V)\in\mathcal{P}^2\times\mathcal{V}}\chi(q_p, \hat{V}_{B,p'}\otimes\tilde{V}_{B'}) - \sup_{(p,p',V)\in\mathcal{P}^2\times\mathcal{V}}\chi(q_p, \hat{V}_{E,p'}\otimes\tilde{V}_{E'}) - \delta \qquad (5.41)\end{aligned}$$

holds, and for each $s := (p, p', V)$ the inequalities

$$\begin{aligned}\Pr(\hat{K}_s \neq \hat{K}'_s) \leq \mu, \text{ and} \\ \log M - H(\hat{K}_s) + I(\hat{K}; E^n E'^n \hat{\Lambda}, \hat{\rho}_{\hat{K}\hat{\Lambda}E^n E'^n, s}) \leq \mu \qquad (5.42)\end{aligned}$$

being satisfied with $\mu = 2^{-\sqrt[16]{n}c_2}$ with a constant $c_2 > 0$. Notice, that the cq-channel $\tilde{V}_B$ defined in (5.40) has classical structure in the sense, that all its output quantum states are diagonal in the orthogonal basis $\{t\}$. Consequently, we can assume, that for each $l \in [\hat{L}]$, $m \in [M]$ the corresponding effect $\hat{D}_{lm}$ has the form

$$\hat{D}_{lm} = \sum_{t^n\in\mathcal{T}^n} D_{lt^n m} \otimes |t^n\rangle\langle t^n| .$$

We define the POVM

$$D := \{D_{lt^n m}\}_{(l,t^n,m)\in[\hat{L}]\times\mathcal{T}^n\times[M]}$$

and the stochastic matrix $T : \mathcal{Y} \to [\hat{L}] \times \mathcal{T}^n \times [M]$ by

$$T(l, t^n, m|y^n) := \sum_{u\in\mathcal{U}^n} P_{T|U}^n(t^n|u^n)\hat{T}(l, m|u^n)P_{U|Y}^n(u^n|y^n) \quad ((l, t^n, m, y^n) \in [L] \times \mathcal{T}^n \times [M] \times \mathcal{Y}^n). \qquad (5.43)$$

With these definitions, $\mathcal{D} := (T, D)$ is an $(n, M, \hat{L} \cdot |\mathcal{T}|^n)$ secret-key distillation protocol for $\mathfrak{I}$.

It holds for each $s := (p, V) \in \mathcal{P} \times \mathcal{V}$, $m, m' \in [M]$

$$
\begin{aligned}
&P_{KK',s}(m,m') \\
&= \sum_{l=1}^{\hat{L}} \sum_{t^n \in \mathcal{T}^n} \sum_{y^n \in \mathcal{Y}^n} p^n(y^n) T(l, t^n, m | y^n) \mathrm{tr}(D_{lt^n m'} V_B^{\otimes n}(y^n)) && (5.44) \\
&= \sum_{u^n \in \mathcal{U}^n} \sum_{l=1}^{\hat{L}} \sum_{t^n \in \mathcal{T}^n} \sum_{y^n \in \mathcal{Y}^n} p^n(y^n) P_{T|U}^n(t^n|u^n) \hat{T}(l, m|u^n) P_{U|Y}^n(u^n|y^n) \mathrm{tr}(D_{lt^n m'} V_B^{\otimes n}(y^n)) && (5.45) \\
&= \sum_{u^n \in \mathcal{U}^n} \sum_{l=1}^{\hat{L}} \sum_{t^n \in \mathcal{T}^n} \sum_{y^n \in \mathcal{Y}^n} q_p^n(u^n) W_p^n(y^n|u^n) \\
&\times \hat{T}(l, m|u^n) \cdot \mathrm{tr}((D_{lt^n m'}) \otimes |t^n\rangle\langle t^n|)(V_B^{\otimes n}(y^n) \otimes \tilde{V}_{B'}^{\otimes n}(u^n))) && (5.46) \\
&= \sum_{u^n \in \mathcal{U}^n} \sum_{l=1}^{\hat{L}} q_p^n(u^n) \hat{T}(l, m|u^n) \mathrm{tr}(\hat{D}_{lm}(\hat{V}_p^{\otimes n}(u^n) \otimes \tilde{V}_{B'}^{\otimes n}(u^n))) && (5.47) \\
&= P_{\hat{K}\hat{K}',(p,p,V)}(m,m'). && (5.48)
\end{aligned}
$$

The equality in (5.44) is holds by definition, (5.45) is valid by (5.43). The equality in (5.46) is justified by definition of q_p, W_p, and the fact, that

$$\mathrm{tr}(|t\rangle\langle t| \tilde{V}_B(u)) = P_{T|U}(t|u) \qquad (t \in \mathcal{T}, u \in \mathcal{U})$$

holds by definition of $\tilde{V}$. From (5.48), we directly infer

$$\Pr(K_s \neq K_s') = \Pr(\hat{K}_{(p,p,V)} \neq \hat{K}'_{(p,p,V)}) \leq \mu, \text{ and} \tag{5.49}$$

$$H(K_s) = H(\hat{K}_{(p,p',V)}) \tag{5.50}$$

Notice, that by definition of $\hat{\rho}_{(p,p,V)}$ is (up to unitaries permuting tensor factors) equal to $\tilde{\rho}_{(p,V)}$, i.e.

$$
\begin{aligned}
\hat{\rho}_{(p,p,V)} &= \sum_{u \in \mathcal{U}} q_p(u) \ |u\rangle\langle u| \otimes \hat{V}_p(u) \otimes \tilde{V}(u) \\
&= \sum_{t \in \mathcal{T}} \sum_{u \in \mathcal{U}} \sum_{y \in \mathcal{Y}} P_{T|U}(t|u) P_{U|Y}(u|y) p(y) \ |u\rangle\langle u| \otimes V(y) \otimes |t\rangle\langle t| \otimes |t\rangle\langle t|
\end{aligned}
$$

holds for each $(p, V) \in \mathcal{P} \times \mathcal{V}$. Consequently, it follows

$$I(K; \hat{\Lambda} T E, \hat{\rho}_{K\hat{\Lambda}TE^n,s}) = I(\hat{K}; E E^n \hat{\Lambda}, \hat{\rho}_{\hat{K}\hat{\Lambda}E^n E'^n,(p,p,V)}). \tag{5.51}$$

The inequalities contained in (5.50) and (5.51) together with the one in (5.42) yield

$$\log M - H(K_s) + I(K; \hat{\Lambda} T E, \hat{\rho}_{K\hat{\Lambda}TE^n,s}) \leq \mu$$

for each $s = (p, V)$, which, together with (5.49) makes (T, D) an $(n, \hat{L} \cdot |\mathcal{T}|^n, M, \mu)$ forward secret-key distillation protocol for $\mathfrak{I}$. At last, we have to show, that M indeed satisfies the

bound stated in (5.39). We will therefore, lower-bound the right-hand side of (5.41). Because Markov-processing does never increase the variational distance, it holds

$$\|q_p - q_{p'}\|_1 \leq \|p - p'\|_1 \leq \operatorname{diam}(\mathcal{P}) \leq \Delta \tag{5.52}$$

for each $p, p' \in \mathcal{P}$, where the rightmost inequality is by assumption. We obtain for each $V \in \mathcal{V}$

$$\begin{aligned}\left|\chi\left(q_p, \hat{V}_{B,p} \otimes \tilde{V}_B\right) - \chi\left(q_{p'}, \hat{V}_{B,p} \otimes \tilde{V}_{B'}\right)\right| &\leq 6\|q_p - q_{p'}\|_1 \ \log|\mathcal{U}| + 2h(\|q_p - q_{p'}\|_1) \\ &\leq 6\Delta \log|\mathcal{U}| + 2h(\Delta). \end{aligned} \tag{5.53}$$

The first equality above is by application of Lemma 74 which can be found in Appendix A.2, the second by (5.52). The bound in (5.53) directly implies

$$\begin{aligned}\inf_{(p',V)} \chi\left(q_p, \hat{V}_{B,p'} \otimes \tilde{V}_{B'}\right) &\geq \inf_{V} \chi\left(q_p, \hat{V}_{B,p} \otimes \tilde{V}_{B'}\right) - \Delta \log|\mathcal{U}| - 2h(\Delta) \\ &= \inf_{\rho \in \mathfrak{J}_p} I(U, BB', \hat{\rho}) - \Delta \log|\mathcal{U}| - 2h(\Delta). \end{aligned} \tag{5.54}$$

for each $p \in \mathcal{P}$. The equality in (5.54) holds by the identity

$$\chi\left(q_p, \hat{V}_{B,p} \otimes \tilde{V}_{B'}\right) = I(U, BB', \hat{\rho}_{(p,p,V)}).$$

By similar reasoning, we also yield the bound

$$\sup_{(p,p',V)} \chi\left(q_p, \hat{V}_{E,p'} \otimes \tilde{V}_{E'}\right) \leq \sup_{\rho \in \mathfrak{J}_p} I(U, EE', \hat{\rho}) + 6\Delta \log|\mathcal{U}| + 2h(\Delta). \tag{5.55}$$

Combination of (5.54) and (5.55) for each $p \in \mathcal{P}$ ensures us, that

$$\begin{aligned}&\inf_{(p,p',V)\in\mathcal{P}^2\times\mathcal{V}} \chi(q_p, \hat{V}_{B,p'} \otimes \tilde{V}_{B'}) - \sup_{(p,p',V)\in\mathcal{P}^2\times\mathcal{V}} \chi(q_p, \hat{V}_{E,p'} \otimes \tilde{V}_{E'}) \\ &\geq \inf_{p\in\mathcal{P}} \left(\inf_{\rho\in\mathfrak{J}_p} I(U, BB', \hat{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U, EE', \hat{\rho}) \right) - 12\Delta \log|\mathcal{U}| - 4h(\Delta) \end{aligned}$$

holds. Note, that the identities

$$S(\tilde{\rho}_{BT}) = S(\hat{\rho}_{BB'}), \text{ and } \quad S(\tilde{\rho}_{ET}) = S(\hat{\rho}_{EE'}),$$

are valid. Moreover, for each $\rho \in \mathfrak{J}$ the equalities

$$I(U; B|T, \tilde{\rho}) = H(P_{UT}) + S(\tilde{\rho}_{BT}) - S(\tilde{\rho}_{UBT}) - H(P_T), \text{ and} \tag{5.56}$$

$$I(U; E|T, \tilde{\rho}) = H(P_{UT}) + S(\tilde{\rho}_{ET}) - S(\tilde{\rho}_{UET}) - H(P_T), \tag{5.57}$$

hold by definition, where P_T, P_{TU} are the distributions for the random variables T and TU. It is important, to notice here, that the distributions P_U and P_{TU} depend only on the Markov chain

on the sender's systems. Therefore, we have for each $p \in \mathcal{P}$

$$\begin{aligned}
&\inf_{\rho\in\mathfrak{J}_p} I(U;B|T,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U;E|T,\tilde{\rho}) \\
&= \inf_{\rho\in\mathfrak{J}_p} \left(S(\tilde{\rho}_{BT}) - S(\tilde{\rho}_{UBT})\right) - \sup_{\rho\in\mathfrak{J}_p} \left(S(\tilde{\rho}_{ET}) - S(\tilde{\rho}_{UET})\right) \\
&= \inf_{\rho\in\mathfrak{J}_p} \left(S(\tilde{\rho}_{BB'}) - S(\tilde{\rho}_{UBB'})\right) - \sup_{\rho\in\mathfrak{J}_p} \left(S(\tilde{\rho}_{EE'}) - S(\tilde{\rho}_{UEE'})\right) \\
&= \inf_{\rho\in\mathfrak{J}_p} \left(H(q_p) + S(\hat{\rho}_{BB'}) - S(\hat{\rho}_{UBB'})\right) - \sup_{\rho\in\mathfrak{J}_p} \left(H(q_p) + S(\hat{\rho}_{EE'}) - S(\hat{\rho}_{UEE'})\right) \\
&= \inf_{\rho\in\mathfrak{J}_p} I(U;BB',\hat{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U;EE',\hat{\rho}). \qquad (5.58)
\end{aligned}$$

Collecting the bounds obtained, we can prove the desired lower bound on the key rate. It holds

$$\begin{aligned}
\frac{1}{n}\log M &\geq \inf_{(p,p',V)\in\mathcal{P}^2\times\mathcal{V}} \chi(q_p, \hat{V}_{Bp'} \otimes \tilde{V}_{B'}) - \sup_{(p,p',V)\in\mathcal{P}^2\times\mathcal{V}} \chi(q_p, \hat{V}_{E,p'} \otimes \tilde{V}_{E'}) - \delta \\
&= \inf_{p\in\mathcal{P}} \left(\inf_{\rho\in\mathfrak{J}_p} I(U;BB',\hat{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U;EE',\hat{\rho}) \right) - \delta - 12\Delta\log|\mathcal{U}| - 4h(\Delta) \\
&= \inf_{p\in\mathcal{P}} \left(\inf_{\rho\in\mathfrak{J}_p} I(U;B|T,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U;E|T,\tilde{\rho}) \right) - \delta - 12\Delta\log|\mathcal{U}| - 4h(\Delta).
\end{aligned}$$

The first inequality above is by (5.41), the first inequality is the one from (5.54), while the last inequality is by (5.58). We are done. □

Proposition 55. *Let $\Delta > 0$, and $\mathfrak{J} \subset \mathcal{S}_{cqq}(\mathcal{Y}, \mathcal{K}_{BE})$ be a Δ-regular set of cqq density matrices on $\mathcal{H}_{ABE}$. For all $z, z' \in \mathbb{N}$, it holds*

$$K_{\rightarrow}(\mathfrak{J}) \geq \tilde{K}^{(1)}_{\rightarrow}(\mathfrak{J}, z, z) - \delta - f_{reg}(z, \Delta),$$

with a function $f_{reg} : \mathbb{N} \times \mathbb{R}^+ \to \mathbb{R}^+$ such that $f(r,\Delta) \to 0$ $(\Delta \to 0)$. For a set $\mathfrak{A} := \{\sum_{y\in\mathcal{Y}} p(y) \ |y\rangle\langle y| \otimes \sigma_y\}$ on some space, and $z, z' \in \mathbb{N}$, the function $\tilde{K}^{(1)}_{\rightarrow}(\mathfrak{A}, z, z')$ is defined

$$\tilde{K}^{(1)}_{\rightarrow}(\mathfrak{A}) := \inf_{p\in\mathcal{P}_{\mathfrak{A}}} \sup_{\Gamma:=T\leftarrow U\leftarrow Y_p} \left(\inf_{\sigma\in\mathfrak{A}_p} I(U;B|T,\sigma_\Gamma) - \sup_{\sigma\in\mathfrak{A}_p} I(U;E|T,\sigma_\Gamma) \right).$$

The supremum above is over all Markov chains $T \leftarrow U \leftarrow Y_p$ resulting from application of Markov transition matrices $P_{T|U} : \mathcal{U} \to \mathcal{T}$, $P_{U|Y} : \mathcal{Y} \to \mathcal{U}$ on p for each $p \in p$ with $|\mathcal{U}| = z$, $|\mathcal{T}| = z'$, and

$$\sigma_{TU} := \sum_{y\in\mathcal{Y}} \sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}} P_{T|U}(t|u) P_{U|Y}(u|y) p(y) \ |t\rangle\langle t| \otimes |u\rangle\langle u| \otimes \sigma_y$$

for given transition matrices $P_{T|U}$, $P_{U|Y}$ and

$$\sigma = \sum_{y\in\mathcal{Y}} p(y) \ |y\rangle\langle y| \otimes \sigma_y. \qquad (5.59)$$

Proof. Assume the set $\mathfrak{J}$ is parameterized such that $\mathcal{P} \subset \mathfrak{P}(\mathcal{Y})$ is the set of possible marginal distributions on the sender's system, while to each $p \in \mathcal{P}$ a set $\mathcal{V}_p \subset \mathcal{CQ}(\mathcal{Y}, \mathcal{K}_{BE})$ is associated, i.e.

$$\mathfrak{J} = \left\{ \rho := \sum_{y \in \mathcal{Y}} q(y) \left|y\right\rangle \left\langle y\right| \otimes V(y) : \ q \in \mathcal{P}, V \in \mathcal{V}_p \right\}$$

Let $\delta > 0$, $z, z' \in \mathbb{N}$ be arbitrary but fixed numbers. We show, that

$$\tilde{K}^{(1)}(\mathfrak{J}, z, z') - \delta - f_{reg}(z, \Delta)$$

with $f_{reg}(z, \Delta)$ being defined

$$f_{reg}(z, \Delta) := 32\Delta \log(z \cdot \dim \mathcal{K}_{BE}) + 24h(\Delta)$$

is an achievable forward secret-key distillation rate for $\mathfrak{J}$. Note that the function defined above indeed has the properties claimed above. The strategy of proof will be as follows. We will equip $\mathcal{P}$ with a regular non-intersecting covering, where we utilize the set of types for large enough blocklength to define such. With the right choice of parameters, we obtain a finite family of sources which approximate $\mathfrak{J}$ and have the addition property for fulfilling the hypotheses of Proposition 54. Combining the protocols obtained for each member of the family with an estimation on the first $\sqrt{n}$ letters for blocklength n leads us to a universal protocol for $\mathfrak{I}$.
We begin setting up the covering of $\mathcal{P}$. Define for each $k, l \in \mathbb{N}$, $\lambda \in \mathfrak{T}(k, \mathcal{Y})$ a set

$$\mathfrak{T}_{\lambda,l} := \left\{ q \in \mathfrak{P}(\mathcal{Y}) : \ \forall y \in \mathcal{Y} : \lambda(y) - \tfrac{l}{2k} < q(y) \leq \lambda(y) + \tfrac{l}{2k} \right\}.$$

Notice, that the diameter of $\mathfrak{T}_{\lambda,l}$ is bounded by

$$\operatorname{diam}(\mathfrak{T}_{\lambda,l}) \leq \frac{l \cdot |\mathcal{Y}|}{k},$$

and the sets in the family being pairwise non-intersecting for $l = 1$. We fix k to be specified later, define $\mathcal{P}_\lambda := \mathfrak{T}_{\lambda,3} \cap \mathcal{P}$ for each $\lambda \in \mathfrak{T}(k, \mathcal{Y})$, and denote by $\hat{\mathfrak{T}}$ the collection of all λ with $\mathcal{P}_\lambda$ being nonempty. We construct sets of cqq density matrices, which fit the specifications demanded in Proposition 54, we define

$$\hat{\mathcal{V}}_\lambda := \bigcup_{q \in \mathcal{P}_\lambda} \mathcal{V}_q, \quad \text{and} \quad \hat{\mathfrak{J}}_{p,\lambda} := \left\{ \sum_{y \in \mathcal{Y}} p(y) \left|y\right\rangle \left\langle y\right| \otimes V(y) \right\}_{V \in \tilde{\mathcal{V}}_\lambda}$$

for each $\lambda \in \hat{\mathfrak{T}}$. Fix the number k large enough, to ensure us, that for each $\lambda \in \tilde{\mathfrak{T}}$, p, p' are in $\mathcal{P}_\lambda$ implies

$$d_H(\mathfrak{J}_p^{AB}, \mathfrak{J}_{p'}^{AB}) + d_H(\mathfrak{J}_p^{AE}, \mathfrak{J}_{p'}^{AE}) \leq \Delta,$$

and, in addition $\operatorname{diam}(\mathcal{P}_\lambda) \leq \Delta$. Notice, that this choice of k is indeed possible because we assumed $\mathfrak{J}$ to be Δ-regular. We consider the family $\{\hat{\mathfrak{J}}_\lambda\}_{\lambda \in \hat{\mathfrak{T}}}$, where for each λ, $\hat{\mathfrak{J}}_\lambda$ is the set of density matrices defined by

$$\hat{\mathfrak{J}}_\lambda := \bigcup_{p \in \mathcal{P}_\lambda} \mathfrak{J}_{p,\lambda} = \left\{ \sum_{y \in \mathcal{Y}} p(y) \left|y\right\rangle \left\langle y\right| \otimes V(y) : \ p \in \mathcal{P}_\lambda, V \in \hat{\mathcal{V}}_\lambda \right\}.$$

The rightmost of the above equalities holds by construction. Notice, that since $\mathcal{P}_\lambda$ has diameter bounded, and $\mathfrak{J}_\lambda$ is parameterized by the full Cartesian product $\mathcal{P}_\lambda \times \hat{\mathcal{V}}_\lambda$, Proposition 54 can be applied in each case. Choose for each $\lambda \in \hat{\mathfrak{T}}$, stochastic matrices $P_{T|U,\lambda} : \mathcal{U} \to \mathfrak{P}(\mathcal{T})$ and $P_{U|Y,\lambda} : \mathcal{Y} \to \mathfrak{P}(\mathcal{U})$ such that

$$\begin{aligned}&\inf_{\rho\in\mathfrak{J}_p} I(U_\lambda;B|T_\lambda,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U_\lambda;E|T_\lambda,\tilde{\rho})\\ &\geq \sup_{T\leftarrow U\leftarrow Y}\left(\inf_{\rho\in\mathfrak{J}_p} I(U;B|T,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U;E|T,\tilde{\rho})\right) - \frac{\delta}{2}\end{aligned} \tag{5.60}$$

is fulfilled. Resulting from the choices made, it also holds

$$\begin{aligned}&\inf_{\rho\in\hat{\mathfrak{J}}_{p,\lambda}} I(U_\lambda;B|T_\lambda,\tilde{\rho}) - \sup_{\rho\in\hat{\mathfrak{J}}_{p,\lambda}} I(U_\lambda;E|T_\lambda,\tilde{\rho})\\ &\geq \inf_{\rho\in\mathfrak{J}_p} I(U_\lambda;B|T_\lambda,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U_\lambda;E|T_\lambda,\tilde{\rho}) - f_{reg}(\Delta,z,z')\end{aligned} \tag{5.61}$$

for each $\lambda \in \tilde{\mathfrak{T}}, p \in \mathcal{P}_\lambda$. The inequality above is by continuity together with properties of our construction and definition of f_{reg}. The full argument for justification can be found in Appendix A.3. Combining the above estimates, we have for each $\lambda \in \hat{\mathfrak{T}}$

$$\begin{aligned}&\inf_{p\in\mathcal{P}_\lambda}\left(\inf_{\rho\in\hat{\mathfrak{J}}_{p,\lambda}} I(U_\lambda;B|T_\lambda,\tilde{\rho}) - \sup_{\rho\in\hat{\mathfrak{J}}_{p,\lambda}} I(U_\lambda;E|T_\lambda,\tilde{\rho})\right)\\ &\geq \inf_{p\in\mathcal{P}_\lambda}\left(\inf_{\rho\in\mathfrak{J}_p} I(U_\lambda;B|T_\lambda,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U_\lambda;E|T_\lambda,\tilde{\rho})\right) - f_{reg}(\Delta,z,z')\\ &\geq \inf_{p\in\mathcal{P}_\lambda}\sup_{T\leftarrow U\leftarrow Y}\left(\inf_{\rho\in\mathfrak{J}_p} I(U;B|T,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U;E|T,\tilde{\rho})\right) - \frac{\delta}{2} - \frac{1}{2}f_{reg}(\Delta,z,z')\\ &\geq \inf_{p\in\mathcal{P}}\sup_{T\leftarrow U\leftarrow Y}\left(\inf_{\rho\in\mathfrak{J}_p} I(U;B|T,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U;E|T,\tilde{\rho})\right) - \frac{\delta}{2} - \frac{1}{2}f_{reg}(\Delta,z,z')\\ &= \tilde{K}^{(1)}_{\rightarrow}(\mathfrak{J},z,z') - \frac{\delta}{2} - \frac{1}{2}f_{reg}(\Delta,z,z').\end{aligned} \tag{5.62}$$

fulfilled. The first inequality above holds by (5.61), the second is by (5.60). Let the blocklength $n \in \mathbb{N}$ be fixed. We set $n = a_n + b_n$ with $a_n := \lceil\sqrt{n}\rceil$, $b_n := n - a_n$, and consider the decomposition

$$\mathcal{Y}^n = \mathcal{Y}^{a_n} \times \mathcal{Y}^{b_n}.$$

Applying Proposition 54, to each of the sets $\hat{\mathfrak{J}}_\lambda$, we infer for each large enough n, $\lambda \in \hat{\mathfrak{T}}$ existence of an $(b_n, M, L, \hat{\vartheta})$ secret-key distillation protocol $(\hat{T}_\lambda, \hat{D}_\lambda)$ for $\hat{\mathfrak{J}}_\lambda$ with

$$\hat{\vartheta} \leq 2^{-\sqrt[16]{b_n}c_\lambda} \leq 2^{\sqrt[16]{b_n}c}$$

with a strictly positive constant c_λ and $c := \min_{\lambda \in \tilde{\mathfrak{T}}} c_\lambda$ and

$$M = \left\lfloor \exp\left(b_n \left(\tilde{K}^{(1)}_{\rightarrow}(\mathfrak{J}) - \frac{3\delta}{4} - f_{reg}(\Delta, z, z') \right) \right) \right\rfloor. \tag{5.63}$$

Note that the combination of M and θ is indeed possible is justified by combining the claim of Proposition 54 and the bound in (5.62). Next, we define a two-phase protocol, where the first a_n letters from the source observed by the sender are used to estimate $\lambda \in \hat{fT}$, while the protocol $(\hat{D}_\lambda, \hat{T}_\lambda)$ for the estimated parameter λ is applied on the remaining b_n outputs of the source. Tor formalize this strategy, we define a stochastic matrix

$$T : \mathcal{Y}^n \to \mathfrak{P}([L] \times \mathfrak{T}(k, \mathcal{Y}) \times [M])$$

with entries

$$T(l, \theta, m | y^n) := \hat{T}_\theta(l, m | y^{b_n})\ \delta_{\theta \xi(y^{a_n})} \qquad (l, m, \theta, y^n) \in [L] \times [M] \times \hat{\mathfrak{T}} \times \mathcal{Y}^n)$$

for each $\mu \in \hat{\mathfrak{T}}$, where we defined a function $\xi : \mathcal{Y}^{a_n} \to \mathfrak{T}(k, \mathcal{Y})$ which maps each y^{a_n} to the unique member $\lambda = \xi(y^{a_n})$ such that $\mathfrak{T}_{\lambda,1}$ contains the type of y^{a_n}. Notice, that some of the entries may are undefined, if $\hat{\mathfrak{T}}$ does not contain all elements of $\mathfrak{T}(k, \mathcal{Y})$. In this case, entries can be defined in any consistent way, because they will be of no further relevance. Moreover, we introduce matrices

$$D_{lm\theta} := \mathbb{1}^{\otimes a_n}_{\mathcal{H}_B} \otimes D^\theta_{lm} \in \mathcal{L}(\mathcal{H}^{\otimes n}_B),$$

where D^θ_{lm} is the corresponding effect from the POVM D_θ associated to θ. With these definitions, it is clear, that (T, D) is an $(n, M, L \cdot |\hat{\mathfrak{T}}|, \vartheta)$ forward secret-key distillation protocol for $\mathfrak{J}$, with a number ϑ we will bound below. Let

$$\rho := \sum_{y \in \mathcal{Y}} p(y)\, |y\rangle \langle y| \otimes V(y)$$

be any fixed member of $\mathfrak{J}$, and λ_0 the unique type in $\hat{\mathfrak{T}}$ such that $p \in T_{\lambda_0,1}$. It is important to notice, that not only for λ_0, but also for each $\theta \in \hat{\mathfrak{T}}$ with $\theta \in T_{\lambda_0,3}$, ρ is also a member of $\hat{\mathfrak{J}}_\theta$. Assuming application of the protocol to ρ, we suppress indicating the chosen member in the following formulas. By definition, it holds

$$\begin{aligned} P_{KK'\Lambda\Theta|Y^n}(m, m', l, \theta | y^n) &= \hat{T}_\theta(m, l | y^{b_n}) \cdot \delta_{\theta, \xi(y^{a_n})} \cdot \mathrm{tr}(D^\theta_{lm'} V^{\otimes b_n}(y^{b_n})) \\ &= P^\theta_{\hat{K}\hat{K}'\hat{\Lambda}|Y^{b_n}}(m, m', l, \theta | y^{b_n}) \cdot \delta_{\theta, \xi(y^{a_n})}, \end{aligned}$$

with $P^\theta_{\hat{K}\hat{K}'\hat{\Lambda}|Y^{b_n}}(m, m', l, \theta | y^{b_n})$ being the conditional distribution generated by $(\hat{T}_\theta, \hat{D}_\theta)$. We define the sets

$$\iota_1 := \{y^{a_n} :\ \xi(y^{a_n}) \in \mathfrak{T}_{\lambda,1}\}, \quad \text{and} \quad \iota_3 := \{y^{a_n} :\ \xi(y^{a_n}) \in \mathfrak{T}_{\lambda,3}\}. \qquad (\lambda \in \mathfrak{T}(k, \mathcal{Y}))$$

It holds

$$\begin{aligned}
P_{KK'}(m,m') &= \sum_{\theta\in\mathfrak{T}(k,\mathcal{Y})}\sum_{y^n\in\mathcal{Y}^n}\sum_{l=1}^{L} P_{KK'\Lambda\Theta|\mathcal{Y}^n}(m,m',l,\theta|y^n)p^n(y^n)\\
&= \sum_{\theta\in\mathfrak{T}(k,\mathcal{Y})}\sum_{y^{a_n}\in\mathfrak{T}_{\theta,1}} p^{a_n}(y^{a_n})\sum_{y^{b_n}\in\mathcal{Y}^{b_n}}\sum_{l=1}^{L} P^{\theta}_{\hat{K}\hat{K}'\hat{\Lambda}|Y^{b_n}}(m,m',l|y^{b_n})p^{b_n}(y^{b_n})\\
&= \sum_{\theta\in\mathfrak{T}(k,\mathcal{Y})} p^{a_n}(\iota_{1,\theta})\; P^{\theta}_{KK'}(m,m')
\end{aligned}$$

for each $m,m'\in[M]$. We denote the key random variables produced by performing $(\hat{T}_\theta,\hat{D}_\theta)$ on $\rho^{\otimes b_n}$ by $\hat{K}_\theta$ and $\hat{K}'_\theta$. We directly obtain

$$\begin{aligned}
\Pr(K\neq K') &= \sum_{\theta\in\mathfrak{T}(k,\mathcal{Y})} p^{a_n}(\iota_{\theta,1})\cdot\Pr(\hat{K}_\theta\neq\hat{K}'_\theta)\\
&\leq p^{a_n}(\iota_{\lambda_0,3})\cdot\hat{\vartheta}+p^{a_n}(\iota^c_{\lambda_0,3})\\
&\leq 2^{\sqrt[16]{b_n}c}+2^{-a_n\frac{c}{k^2}}\\
&\leq 2\hat{\vartheta}
\end{aligned}\tag{5.64}$$

The first inequality above is by the fact, that the protocol associated to each $\theta\in\hat{\mathfrak{T}}\cap\mathfrak{T}_{\lambda_0,3}$ is $\hat{\vartheta}$-good for ρ by construction. The second inequality is by standard type bounds. Explicitly, we have by construction $y^{a_n}\in\iota^c_{\lambda_0,3}$ implying

$$\left|\frac{1}{a_n}N(e|y^{a_n})-p(e)\right|>\frac{1}{k}$$

for all $e\in\mathcal{Y}$, where $N(e|y^{a_n})$ is the number of occurrences of the letter e in y^{a_n}. Consequently

$$p^{a_n}(\iota^c_{\lambda_0,3})\leq p^{a_n}\left(\left(T^{a_n}_{p,\frac{1}{k}}\right)^c\right)\leq 2^{-a_n\frac{c}{k^2}}\leq\hat{\vartheta},\tag{5.65}$$

where c is a universal, strictly positive constant, and the last inequality holds for large enough choice of n. Also, it holds

$$\begin{aligned}
&H(K)-I(K;\Lambda\Theta E^n,\rho_{K\Lambda\Theta E^n})\\
&=H(K)-I(K;\Theta)-I(K;\Lambda E^n|\Theta,\rho_{K\Lambda\Theta E^n})\\
&=H(K|\Theta)-I(K;\Lambda E^n|\Theta,\rho_{K\Lambda\Theta E^n})\\
&=\sum_{\theta\in\mathfrak{T}}p^{a_n}(\iota_{\theta,1})\;\left(H(K|\Theta=\theta)-I(K;\Lambda E^n|\Theta=\theta,\rho_{K\Lambda\Theta E^n})\right)\\
&=\sum_{\theta\in\mathfrak{T}}p^{a_n}(\iota_{\theta,1})\;\left(H(\hat{K}^\theta)-I(\hat{K}^\theta;\Lambda E^{b_n},\rho^\theta_{\hat{K}\hat{\Lambda}E^{b_n}})\right),
\end{aligned}\tag{5.66}$$

where the first equality is the chain rule for the quantum mutual information applied, the second holds by definition of the classical mutual information. The third equality results from the fact, that if θ is the estimate obtained in the first a_n outputs of the source, $(\hat{T}_\theta,\hat{D}_\theta)$ is performed on the

remaining b_n outputs, which determines the conditional quantities as generated from application of the protocol. Therefore, we obtain

$$\begin{aligned}
&\log M - H(K) + I(K;\Lambda\Theta E^n, \rho_{K\Lambda\Theta E^n}) \\
&= \sum_{\theta\in\mathfrak{T}} p^{a_n}(\iota_{\theta,1}) \left(\log M - H(\hat{K}^\theta) + I(\hat{K}^\theta; \hat{\Lambda}^\theta E^{b_n}, \rho^\theta_{\hat{K}\hat{\Lambda}E^{b_n}})\right) \\
&\leq p^{a_n}(\iota_{\lambda_0,3}) \cdot \hat{\vartheta} + p^{a_n}(\iota_{\lambda_0,3}) \cdot (2 \cdot \log M + \log L + b_n \cdot \log\dim \mathcal{K}_{BE}) \\
&\leq 2\hat{\vartheta}
\end{aligned} \tag{5.67}$$

where the equality above follows from (5.66), the first inequality is by the fact, that the protocol $(\hat{T}_\theta, \hat{D}_\theta)$ is $\hat{\vartheta}$-good for ρ whenever θ is a direct grid point neighbour of λ_0, i.e. $\mathfrak{T}_{\theta,1} \subset \mathfrak{T}_{\lambda_0,3}$. Moreover we applied the ultimate bound $I(A;B,\rho) \leq 2\log\dim \mathcal{H}_A \otimes \mathcal{H}_B$ which holds for each state σ on any Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. The last inequality holds with a large enough choice of n by application of the bound in (5.65). The bounds obtained in (5.67) and (5.64 show us, that (T, D) is actually an (n, M, L, ϑ) forward secret-key distillation protocol for $\mathfrak{J}$, with $\vartheta \leq 2\hat{\vartheta}$, and, since $b_n/n \to 1$ for $n \to \infty$, it holds

$$\begin{aligned}
\frac{1}{n}\log M &\geq \frac{b_n}{n}\tilde{K}^{(1)}_{\rightarrow}(\mathfrak{J}) - \frac{3\delta}{4} - f_{reg}(\Delta, z, z') \\
&\geq \tilde{K}^{(1)}_{\rightarrow}(\mathfrak{J}) - \delta - f_{reg}(\Delta, z, z')
\end{aligned}$$

if n is large enough, where the first inequality is from (5.63). □

To prove achievability of the multi-letter formula claimed in Theorem 48, we have to ensure ourselves, that regularity conditions do not break down when considering the set $\mathfrak{I}^{\otimes n} := \{\rho^{\otimes n} : \rho \in \mathfrak{I}\}$ instead of a set $\mathfrak{I}$ of cqq density matrices. The following two basic lemmas will turn out to be sufficient for our needs.

Lemma 56. *Let $\mathfrak{I}, \mathfrak{J} \subset \mathcal{L}(\mathcal{K})$ be any two sets of density matrices. It holds for each $n \in \mathbb{N}$*

$$d_H(\mathfrak{I}^{\otimes n}, \mathfrak{J}^{\otimes n}) \leq n \cdot d_H(\mathfrak{I}, \mathfrak{J}),$$

where d_H is the Hausdorff distance induced by the trace norm on the underlying space.

Proof. The inequality

$$\|a^{\otimes n} - b^{\otimes n}\|_1 \leq n \cdot \|a - b\|_1$$

valid for any two matrices $a, b \in \mathcal{L}(\mathcal{K})$ inherits to the Hausdorff distance. It holds

$$\sup_{a\in\mathfrak{I}} \inf_{b\in\mathfrak{J}} \|a^{\otimes n} - b^{\otimes n}\|_1 \leq n \cdot \sup_{a\in\mathfrak{I}} \inf_{b\in\mathfrak{J}} \|a - b\|_1.$$

□

Lemma 57. *Let $\mathfrak{I}$ be a set of cqq density matrices. It holds*

$$\mathfrak{I}\ \epsilon - \textit{regular} \ \Rightarrow\ \mathfrak{I}^{\otimes k}\ k\cdot\epsilon - \textit{regular}.$$

for each $k \in \mathbb{N}$.

Proof. Is by direct application of Lemma 56 and the definition of regularity. □

We now obtained sufficient preparations to tackle the proof of achievability in Theorem 48. Before we head to the proof, we ensure ourselves, that the limit in (5.5) indeed exists.

Lemma 58. *Let $\mathfrak{I}$ be a set of cqq density matrices on $\mathcal{H}_{ABE}$. It holds*

$$\sup_{k\in\mathbb{N}}\frac{1}{k}K^{(1)}(\mathfrak{I}^{\otimes k}) = \lim_{k\to}\frac{1}{k}K^{(1)}(\mathfrak{I}^{\otimes k}).$$

Proof. The assertion of the lemma follows from application of Fekete's lemma [Fek23] on the sequence $K^{(1)}(\mathfrak{I}^{\otimes k})$. We check that the hypotheses of Fekete's lemma are fulfilled. Clearly, the sequence is bounded. We show, that it is also superadditive, i.e. $K^{(1)}(\mathfrak{I}^{\otimes(k+l)}) \geq K^{(1)}(\mathfrak{I}^{\otimes k}) + K^{(1)}(\mathfrak{I}^{\otimes l})$ being valid for all $k, l \in \mathbb{N}$. We can for each k write $K^{(1)}(\mathfrak{I}^{\otimes k})$ in the form

$$K^{(1)}(\mathfrak{I}^{\otimes k}) \ = \ \inf_{p\in\mathcal{P}_{\mathfrak{I}}}\sup_{z,z'\in\mathbb{N}}\hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z, z')$$

where we defined

$$\hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z, z') := \sup_{T\leftarrow U\leftarrow X_p}\left(\inf_{\sigma\in\mathfrak{I}_p} I(U;B|T,\tilde{\sigma}) - \sup_{\sigma\in\mathfrak{I}_p} I(U;E|T,\tilde{\sigma})\right),$$

with the outer maximization above being over all Markov chains generated by transition matrices $P_{U|X} : \mathcal{X} \to \mathfrak{P}(\mathcal{X})$ and $P_{T|U} : \mathcal{U} \to \mathfrak{P}(\mathcal{T})$ with alphabets of cardinalities $|\mathcal{U}| = z$, $|\mathcal{T}| = z'$, and

$$\tilde{\sigma} := \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} P_{T|U}(t|u)\ P_{U|X}(u|x)\ p(x)\,|u\rangle\langle u|\,|t\rangle\langle t|\otimes V(x)$$

for

$$\sigma = \sum_{x\in\mathcal{X}} p(x)\,|x\rangle\langle x|\otimes V(x).$$

Notice, that for each $p \in \mathcal{P}_{\mathfrak{I}}$, $z, z' \in \mathbb{N}$

$$\hat{K}^{(1)}(\mathfrak{I}^{\otimes(k+l)}, p, z, z') \geq \hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z, z') + \hat{K}^{(1)}(\mathfrak{I}^{\otimes l}, p, z, z'),$$

and moreover, for each $z_1 \leq z_2$, $z_1' \leq z_2'$,

$$\hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z_2, z_2') \geq \hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z_1, z_1')$$

holds for each $k \in \mathbb{N},\ p \in \mathcal{P}_{\mathfrak{I}}$. We obtain

$$\begin{aligned}\hat{K}^{(1)}(\mathfrak{I}^{\otimes(k+l)}, p, z_2, z_2') &\geq \hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z_2, z_2') + \hat{K}^{(1)}(\mathfrak{I}^{\otimes l}, p, z_2, z_2') \\ &\geq \hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z_2, z_2') + \hat{K}^{(1)}(\mathfrak{I}^{\otimes l}, p, z_1, z_1')\end{aligned}$$

Consequently, it holds

$$\sup_{z,z' \in \mathbb{N}} \hat{K}^{(1)}(\mathfrak{I}^{\otimes(k+l)}, p, z, z') \geq \sup_{z,z' \in \mathbb{N}} \hat{K}^{(1)}(\mathfrak{I}^{\otimes k}, p, z, z') + \sup_{z,z' \in \mathbb{N}} \hat{K}^{(1)}(\mathfrak{I}^{\otimes l}, p, z, z')$$

for each $p \in \mathcal{P}\mathfrak{I}$. We conclude

$$K^{(1)}(\mathfrak{I}^{\otimes(k+l)}) \geq K^{(1)}(\mathfrak{I}^{\otimes k}) + K^{(1)}(\mathfrak{I}^{\otimes l})$$

□

Proof of Theorem 48. We first prove achievability, i.e. validity of the inequality

$$K_{\rightarrow}(\mathfrak{I}) \geq \lim_{k \to \infty} \frac{1}{k} K_{\rightarrow}^{(1)}(\mathfrak{I}^{\otimes k})$$

Let $z.z', k \in \mathbb{N}$ and $\delta > 0$ be arbitrary and fixed. We show, that

$$\frac{1}{k} \tilde{K}_{\rightarrow}^{(1)}(\mathfrak{I}^{\otimes} k) - \delta$$

is an achievable forward secret key distillation rate. We apply Proposition 55 with $\mathfrak{J} = \mathfrak{I}^{\otimes k}$, and conclude, that for each large enough blocklength $g \in \mathbb{N}$, we find an (l, M, L, ϑ) forward secret-key distillation protocol for $\mathfrak{I}^{\otimes k}$ with $\vartheta \leq 2^{-\sqrt[16]{g} c_3}$ with a constant $c_3 > 0$, and

$$\frac{1}{g} \log M \geq \tilde{K}^{(1)}(\mathfrak{I}^{\otimes k}, z, z') - \frac{2}{3}\delta \tag{5.68}$$

where we chose Δ small enough to satisfy $f_{reg}(z, z', \Delta) \leq \frac{\delta}{3}$. Since an (g, K, M, ϑ) protocol for $\mathfrak{I}^{\otimes k}$ is obviously an $(g \cdot k, M, L, \vartheta)$ protocol for $\mathfrak{I}$, we obtained sufficient protocols for all large enough blocklengths being integer multiples of k. We can achieve sufficient protocols also for the remaining blocklengths just by wasting resources. To be explicit, let $n = k \cdot g + r$ with $0 < r < k$ and assume $(\hat{T}_{gk}, \hat{D}_{gk})$ being an $(g \cdot k, M, L, \mu)$ protocol for $\mathfrak{I}$. Define a protocol (T_n, D_n) for blocklength n by setting

$$T_n(l, m|x^n) = \hat{T}_{gk}(l, m|(x_1, \ldots, x_{g \cdot k})) \qquad (x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n)$$

and effects

$$D_{n,lm} := \hat{D}_{gk,lm} \otimes \mathbb{1}_{\mathcal{H}_B}^{\otimes k}$$

for each $l \in [L], m \in [M]$. It is clear, that (T_n, D_n) is an (n, M, L, μ) forward secret-key distillation protocol for $\mathfrak{I}$ with rate

$$\frac{1}{n} \log M = \frac{1}{g \cdot k + r} \log M \geq \frac{1}{g \cdot k} \log M - \frac{\delta}{3} \tag{5.69}$$

if n is large enough. It follows from (5.68) and (5.69), that we actually achieve

$$\frac{1}{k}\tilde{K}^{(1)}(\mathfrak{I}^{\otimes k}, z, z') - \delta$$

Since δ, z, z' where arbitrary, we are done. We do not give a detailed argument for the converse inequality here, since the assertion directly follows from (5.7) together with a converse proof for the case of a source with SMI given in the next section. □

5.3 Secret-key distillation with sender marginal information (SMI)

In this section, we assume that the sender has perfect knowledge of his/her marginal distribution deriving from the source statistics. We will prove the achievability part of Theorem 49 by decomposing each compound cqq source into a *finite* collection of regular compound cqq sources. To obtain such an approximation, we need the following basic assertion. For a given set X, we use the notation 2^X for the power set.

Lemma 59. *Let d_H be the Hausdorff distance on $2^{\mathbb{R}^n}$ generated by the 1-norm distance on $\mathbb{R}^n$. Let $A \subset \mathbb{R}^n$ be a subset of $\mathbb{R}^n$ with $\operatorname{diam}(A) \leq a < \infty$. For each $\Delta > 0$, there exists a family $\mathcal{R}_A := \{\tilde{A}_\omega\}_{\omega=1}^{\Omega} \subset 2^{\mathbb{R}^n} \setminus \{\varnothing\}$ with the following properties.*

1. $\Omega \leq \exp\left(\left(\frac{n\cdot a}{\Delta}\right)^n\right)$.
2. *For each $B \subset A$ there exists $\omega \in [\Omega]$, such that*

$$d_H(B, \tilde{A}_\omega) \leq \Delta, \quad \textit{and} \quad B \subset \tilde{A}_k.$$

Proof. Equip $\mathbb{R}^n$ with the regular pairwise-disjoint covering, generated by the n-dimensional half-open cubes

$$\left[\left(k_1\frac{\Delta}{n}, \ldots, k_n\frac{\Delta}{n}\right), \left((k_1+1)\frac{\Delta}{n}, \ldots, (k_n+1)\frac{\Delta}{n}\right)\right) \qquad ((k_1, \ldots, k_n) \in \mathbb{Z}^n).$$

Since $\operatorname{diam}(A) \leq a$ is assumed, we do not need more than $K := \left(\frac{n\cdot a}{\Delta}\right)^n$ of these cubes to cover A. Let $\{G_k\}_{k=1}^{K}$ be any parameterization of the family of cubes intersecting with A by $[K] := \{1, \ldots K\}$. Define, for each $\omega \subset [K]$

$$\tilde{A}_\omega := \bigcup_{k\in\omega} G_k.$$

We show, that

$$\mathcal{R}_A := \{\tilde{A}_\omega\}_{\omega=1}^{\Omega}$$

indeed has the properties stated in the lemma. The first property is fulfilled by the bound on K and the fact, that there are not more, than 2^K different values for ω. The member

$$\omega := \{k \in [K] : \ G_k \cap A \neq \varnothing\} \tag{5.70}$$

fulfills the properties demanded for the second property. □

Proof of Theorem 49. For proving achievability, we the following strategy is applied. We approximate $\mathfrak{I}$ by a finite family $\{\mathfrak{I}_\omega\}_{\omega\in\Omega}$ and apply Theorem 48 for each degree of regularity. Let $\mathfrak{I} := \{\rho_s\}_{s\in S}$ be a parameterization of $\mathfrak{I}$, and $\{\rho_{A,t}\}_{t\in T}$ be a parameterization of the set of marginal states on $\mathcal{H}_A$ which derive from members of $\mathfrak{I}$. Fix an arbitrary $\Delta > 0$ and let $\{\tilde{A}_\omega\}_{\omega\in\Omega}$ be an approximation of $\mathfrak{I}$ with the properties stated in Lemma 59 with parameter λ. Note, that by identifying $\mathbb{C}$ to $\mathbb{R}^2$ in the usual way, the approximation satisfies

$$|\Omega| \leq \exp\left(\left(\frac{4\dim\mathcal{H}_{ABE}^2}{\Delta}\right)^{4\dim\mathcal{H}_{ABE}^2}\right) < \infty,$$

where we only use the fact, the cardinality of Ω is finite. Let, for each $t \in T$, $\omega(t)$ the element of Ω as defined in (5.70) for $\mathfrak{I}_t$. It holds

$$\mathfrak{I}_t \subset \tilde{A}_\omega, \quad \text{and} \quad d_H(\tilde{A}_\omega, \mathfrak{I}_t) \leq \Delta. \tag{5.71}$$

Define

$$\tilde{\mathfrak{I}}_\alpha := \bigcup_{t:\ \omega(t)=\alpha} \mathfrak{I}_t \qquad\qquad (\alpha \in \Omega).$$

The family $\{\tilde{\mathfrak{I}}_\alpha\}_{\alpha\in\Omega}$ is decomposition of $\mathfrak{I}$ into a family of pairwise disjoint sets of cqq density matrices with the additional feature, that for each $\alpha \in \Omega$, $\tilde{\mathfrak{I}}_\alpha$ is 4Δ-regular, which can be justified as follows. For each t, t' with $\omega(t) = \omega(t') := \beta$, it holds

$$d_H(\mathfrak{I}_t, \mathfrak{I}_{t'}) \leq d_H(\mathfrak{I}_t, \tilde{A}_\beta) + d_H(\tilde{A}_\beta, \mathfrak{I}_{t'}) \leq 2\Delta. \tag{5.72}$$

The left of the above inequalities is the triangle inequality for the Hausdorff distance applied, the right hand inequality is by (5.71). Therefore, we infer, using monotonicity of the Hausdorff distance under taking partial traces,

$$d_H(\mathfrak{I}_t^{AB}, \mathfrak{I}_{t'}^{AB}) + d_H(\mathfrak{I}_t^{AE}, \mathfrak{I}_{t'}^{AE}) \leq 2 \cdot d_H(\mathfrak{I}_t, \mathfrak{I}_{t'}) \leq 4\Delta.$$

From applying Proposition 55 on each of the sets $\tilde{\mathfrak{I}}_\beta$, $t \in T$, we know, that for each given $\delta, \mu > 0$, there is a number $k_0(\beta)$, such that we find for each $n > k_0(\beta)$ an $(n, M_\beta, L_\beta, \mu_\beta)$ forward secret-key distillation protocol $(T^{(\beta)}, D^{(\beta)})$ for $\tilde{\mathfrak{I}}_\beta$, with

$$\begin{aligned} \log M_\beta &\geq \tilde{K}^{(1)}(\mathfrak{I}_\beta, z, z') - f_{reg,\beta}(z, \Delta) - \delta \\ &\geq \tilde{K}^{(1)}(\mathfrak{I}, z, z') - f_{reg,\beta}(z, \Delta) - \delta \end{aligned} \tag{5.73}$$

for each $z, z' \in \mathbb{N}$ with a function $f_{reg,\beta}$ as stated in Proposition 55. Moreover, we have bounds

$$\mu_\beta \leq 2^{-\sqrt[16]{n}c_\beta}, \quad \text{and} \quad L_\beta \leq 2^{nR_{c,\beta}}$$

with constants $c_\beta > 0$ and $R_{c,\beta} \in \mathbb{R}^+$ for each $\beta > 0$. We define $c := \min_{\beta\in\Omega}$, and c_β, $R_c := \min_{\beta\in\Omega} R_{c,\beta}$, $L = 2^{nR_c}$, $M := \min_{\beta\in\Omega}$. If we define a stochastic matrix T_t with entries

$$T_t(\beta, l, m|x^n) := T^{(\beta)}(l, m|x^n) \cdot \delta_{\beta\omega(t)} \qquad (\beta \in \Omega, l \in [L], m \in [)$$

and effects

$$D_{\beta lm} := D^{(\beta)}_{lm} \qquad (\beta \in \Omega, l \in [L], m \in [M]),$$

Then $((T_t, D))_{t\in T}$ with $D := \{D_{\beta lm}\}_{(\beta,l,m)\in\Omega\times[L]\times[M]}$ is an $(n, M, |\Omega| \cdot L, \mu)$ secret-key distillation protocol for $\mathfrak{I}$ with SMI, such that

$$\log M \geq \tilde{K}^{(1)}(\mathfrak{I}, z, z') - f_{reg,\beta}(z, \Delta) - \delta$$

holds by (5.73). Since $\Delta > 0$ was arbitrary,

$$\tilde{K}^{(1)}(\mathfrak{I}, z, z') - 2\delta$$

is achievable for each $z, z' \in \mathbb{N}$. Consequently, it holds

$$K_{\to,SMI} \geq \sup_{z,z'\in\mathbb{N}} \tilde{K}^{(1)}(\mathfrak{I}, z, z') - 2\delta = \tilde{K}^{(1)}(\mathfrak{I}) - 2\delta$$

The same reasoning can be applied for $\mathfrak{I}^{\otimes k}$, $k \in \mathbb{N}$, which implies, that

$$\frac{1}{k} K^{(1)}(\mathfrak{I}^{\otimes k})$$

is achievable as well. It remains to prove the converse inequality. Assume $\mathcal{P} \subset \mathfrak{P}(\mathcal{X})$ to be the set of marginal probability distributions on the sender's systems deriving from $\mathfrak{I}$. Define $\mathcal{V}_p \subset \mathcal{CQ}(\mathcal{X}, \mathcal{H}_{BE})$ to be the set of classical-quantum channels associated to each $p \in \mathcal{P}$. I.e. Fix $k \in \mathbb{N}$, and assume $(T, D_p)_{p\in\mathcal{P}}$ to be an (k, M, L, μ) forward secret-key distillation protocol for the set

$$\mathfrak{I}_p := \left\{ \rho_{p,V} := \sum_{x\in\mathcal{X}} p(x) \ |x\rangle\langle x| \otimes V(x) : \ V \in \mathcal{V}_p \right\}$$

of density matrices from $\mathfrak{I}$ having sender marginal distribution p. Fix any $p \in \mathcal{P}$ we suppress the index p for the next lines. Denote by

$$\rho_{\Lambda K K' E^n, V}$$

the state resulting from performing (T_p, D) on ρ_V according to (5.1) for each $V \in \mathcal{V}_p$. Note that the resulting pair (Λ, K) of random variables does not depend on the chosen state ρ_V since all state in $\mathfrak{I}_p$ have same sender marginal distribution. Since $\log M - H(K)$ and

$I(K;\Lambda E^n,\rho_{\Lambda K E^n,V})$ are nonnegative by definition of the protocol and non-negativity of the quantum mutual information, the inequalities

$$\log M - H(K) \leq \mu, \quad \text{and} \quad \sup_{V\in\mathcal{V}} I(K;\Lambda E^n,\rho_{\Lambda K E^n,V}) \leq \mu \tag{5.74}$$

are simultaneously fulfilled. Moreover, we have

$$\begin{aligned} H(K) &= I(K;K'_V) + H(K|K'_V) \\ &\leq I(K;K'_V) + \mu\log M + h(\mu) \\ &\leq I(K;K'_V\Lambda) + \mu\log M + h(\mu) \\ &\leq I(K;B^n\Lambda,\rho_{K\Lambda B^n,V}) + \mu\log M + h(\mu), \end{aligned} \tag{5.75}$$

where the first inequality is by Fano's inequality together with the assumption $\Pr(K\neq K'_V)\leq\mu$, while the last two inequalities follow from the data processing inequalities for the classical and quantum mutual information. We infer

$$\begin{aligned} \log M &\leq H(K) + \mu \\ &\leq \inf_{V\in\mathcal{V}} I(K;B^n\Lambda,\rho_{K\Lambda B^n,V}) + \mu + \mu\log M + h(\mu) \\ &\leq \inf_{V\in\mathcal{V}} I(K;B^n\Lambda,\rho_{K\Lambda B^n,V}) - \sup_{V\in\mathcal{V}} I(K;E^n\Lambda,\rho_{K\Lambda E^n,V}) + 2\mu + \mu\log M + h(\mu) \\ &\leq \inf_{V\in\mathcal{V}} I(K;B^n|\Lambda,\rho_{K\Lambda B^n,V}) - \sup_{V\in\mathcal{V}} I(K;E^n|\Lambda,\rho_{K\Lambda E^n,V}) + 2\mu + \mu\log M + h(\mu) \\ &\leq K^{(1)}(\mathfrak{I}_p^{\otimes k}) + 2\mu + \mu\log M + h(\mu). \end{aligned} \tag{5.76}$$

The first and the third of the above inequalities are from (5.74), while the second is from (5.75). The fourth is by definition of the quantum mutual information together with the fact, that the distribution $(K;\Lambda)$ does not depend on the chosen V. The last one results from observing, that $X\to(\Lambda,K)\to\Lambda$ is a Markov chain. The estimate in (5.76) is valid for each $p\in\mathcal{P}$. Minimization over all $p\in\mathcal{P}$ leads to

$$\begin{aligned} \log M &\leq \inf_{p\in\mathcal{P}} K^{(1)}(\mathfrak{I}_p^{\otimes k}) + 2\mu + \mu\log M + h(\mu) \\ &= K^{(1)}(\mathfrak{I}^{\otimes k}) + 2\mu + \mu\log M + h(\mu), \end{aligned}$$

where the equality above is by definition of the function $K^{(1)}_{\rightarrow}$. □

5.4 Discussion of regularity of compound cqq sources

This section is of twofold purpose. First, we point out that regularity issues have operational significance for forward secret-key distillation from tripartite compound sources. While for regular sources, there is no gap between the forward secret-key key distillation capacities with and without SMI, there may be serious differences in capacities, if the source is not regular. Second, we introduce a weaker notion of regularity than the one introduced in Definition 46, where we utilize notions from the theory of set-valued functions.

5.4.1 Operational significance of regularity conditions

We have seen in the previous section, that there is no difference between the forward secret key distillation capacities with and without SMI, as long as the source is regular in the sense of Definition 46. We admit, that there may be weaker notions of regularity which also exhibit this property (an example of such a condition is introduced in the next section). Regularity conditions seem somewhat technical on a first view. One can easily imagine large classes of sets of cqq density matrices, which are notoriously easy to process even in the case without sender knowledge, while being irregular. This feature is shared in a trivial way by all irregular sources having zero forward secret key distillation capacity under sender knowledge. The following example depicts the fact, that also in nontrivial cases irregularities may be of no consequences for the behaviour of the source regarding forward secret-key distillation.

Example 60. *Define for a finite alphabet $\mathcal{X}$, $A := \{p \in \mathfrak{P}(\mathcal{X}) : \forall x \in \mathcal{X} : p(x) \in \mathbb{Q}\}$, $V \in \mathcal{CQ}(\mathcal{X}, \mathcal{H}_B \otimes \mathcal{H}_E)$, and let $\mathcal{K}_B = \mathbb{C}^{\otimes 2}$ be the Hilbert space of an additional system assigned to the legitimate receiver. Define states*

$$\rho_a := \sum_{x \in \mathcal{X}} a(x) \left|x\right\rangle \left\langle x\right| \otimes V_{BE}(x) \otimes \left|e_a\right\rangle \left\langle e_a\right|$$

with $\{e_1, e_2\}$ being an orthonormal basis in $\mathcal{K}_B$, $e_a := e_1$ if $a \in A$ and $e_a = e_2$ if $a \in A^c$. The source defined by $\mathfrak{I} := \{\rho_a\}_{a \in \mathfrak{P}(\mathcal{X})}$ is not regular, but can be easily converted to a regular one by just discarding the systems on $\mathcal{K}_B$.

Beside the mentioned facts, the question of regularity in principle, bears strong operational significance. The next Theorem shows, that for irregular sources, the capacities with and without sender marginal state knowledge may be substantially different.

Theorem 61. *The equality*

$$K_{\rightarrow,SMI}(\mathfrak{I}) = K_{\rightarrow}(\mathfrak{I})$$

does not hold in general.

Proof. We construct an example of a set $\mathfrak{I}$ with

$$K_{\rightarrow,SMI}(\mathfrak{I}) = 1 \quad \text{and} \quad K_{\rightarrow}(\mathfrak{I}) = 0. \tag{5.77}$$

Let $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and $\mathcal{H}_B = \mathcal{H}_E = \mathbb{C}^2 \otimes \mathbb{C}^2$. We introduce classical-quantum channels $W_1, W_2 : \{0,1\} \to \mathcal{S}(\mathbb{C}^2 \otimes \mathbb{C}^2)$ by

$$\begin{aligned} W_1(x,y) = W_1(x) &:= \left|x\right\rangle \left\langle x\right| \otimes \Pi, \\ W_2(x,y) = W_2(y) &:= \Pi \otimes \left|y\right\rangle \left\langle y\right| \qquad\qquad ((x,y) \in \mathcal{X} \times \mathcal{Y}), \end{aligned}$$

where $\Pi := \frac{\mathbb{1}}{2}$ is the flat state on $\mathbb{C}^2$. We set

$$V_{1,B} = V_{2,E} = W_1, \quad V_{2,B} = V_{1,E} = W_2,$$

and define states

$$\rho_p := \begin{cases} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \frac{1}{4} |x\otimes y\rangle\langle x\otimes y| \otimes V_{1,B}(x) \otimes V_{1,E}(y) & \text{if } p = \pi \\ \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \frac{1}{2} p(y) |x\otimes y\rangle\langle x\otimes y| \otimes V_{2,B}(y) \otimes V_{2,E}(x) & \text{otherwise,} \end{cases}$$

where π denotes the equidistribution on $\{0,1\}$, i.e. $p(0) = p(1) = \frac{1}{2}$. Consider the set $\mathfrak{I} := \{\rho_p : p \in \mathfrak{P}(\mathcal{Y})\}$. We first show the left equality in (5.77). If we define stochastic matrices $P^{(1)}_{U|XY}$, $P^{(2)}_{U|XY} : \mathcal{X}\times\mathcal{Y} \to \mathcal{U} := \{0,1\}$ with entries

$$P^{(1)}_{U|XY}(u|x,y) := \delta_{xu}, \quad \text{and} \quad P^{(2)}_{U|XY}(u|x,y) := \delta_{yu} \qquad (x\in\mathcal{X}, y\in\mathcal{Y}, u\in\mathcal{U}),$$

and use the sender's preprocessings $P^{(1)}_{U|XY}$ for ρ_π and $P^{(2)}_{U|XY}$ for each $p \neq \pi$, we achieve the maximum in the capacity formula derived in Theorem 49. The corresponding states are

$$\begin{aligned} \hat{\rho}_\pi &:= \sum_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P^{(1)}_{U|XY}(u|x,y) \frac{1}{4} |u\rangle\langle u| \otimes V_{1,B}(x) \otimes V_{1,E}(y) \\ &= \sum_{u\in\mathcal{U}} \frac{1}{2} |u\rangle\langle u| \otimes |u\rangle\langle u| \otimes \Pi \otimes \Pi \otimes \Pi \\ \hat{\rho}_p &:= \sum_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P^{(2)}_{U|XY}(u|x,y) p(x) \frac{1}{2} |u\rangle\langle u| \otimes V_{2,B}(y) \otimes V_{2,E}(x) \\ &= \sum_{u\in\mathcal{U}} \frac{1}{2} |u\rangle\langle u| \otimes \Pi \otimes |u\rangle\langle u| \otimes \sigma_{p,E} \qquad (p\in\mathfrak{P}(\mathcal{Y})\setminus\{\pi\}), \end{aligned}$$

where $\sigma_{p,E} := \sum_{x\in\mathcal{X}} p(x) V_{2,E}(x)$. Note, that both of the above states contain perfect common randomness between the legitimate users without sharing any correlations to the eavesdropper, which is the optimum they can achieve, as is easily observed. It therefore holds

$$K_{\rightarrow,SMI}(\mathfrak{I}) = \log 2 = 1.$$

The situation is completely different, if no SMI is present. Let $\mu > 0$ be fixed and (T, D) an arbitrary (n, M, L, μ) forward secret key distillation protocol for $\mathfrak{I}$ without SMI. I.e. the inequalities

$$\Pr(K_p \neq K'_p) \leq \mu \tag{5.78}$$

and

$$\log M - H(K_p) + I(K, \Lambda E^n, \rho_{K\Lambda E^n,p}) \leq \mu \tag{5.79}$$

are satisfied for each $p \in \mathfrak{P}(\mathcal{Y})$. If we define the states

$$\tilde{\rho}_p := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x) \frac{1}{2} |x\otimes y\rangle\langle x\otimes y| \otimes V_{1,B}(x) \otimes V_{2,E}(x) \qquad (p\in\mathfrak{P}(\mathcal{Y})\setminus\{\pi\}),$$

the identity

$$I(K;\Lambda E^n,\rho_{K\Lambda E^n,p}) = I(K;\Lambda B^n,\tilde{\rho}_{K\Lambda B^n,p}) \tag{5.80}$$

is fulfilled by symmetry. Moreover, it holds

$$\|\rho_{K\Lambda B^n,\pi} - \tilde{\rho}_{K\Lambda B^n,p}\|_1 \leq \|\mathrm{tr}_{\mathcal{H}_E^{\otimes n}}\rho_\pi^{\otimes n} - \mathrm{tr}_{\mathcal{H}_E^{\otimes n}}\tilde{\rho}_p^{\otimes n}\|_1 = \|p^n - \pi^n\|_1 \leq n\|p-\pi\|_1 \tag{5.81}$$

where the first inequality is by c.p.t.p. monotonicity of the trace norm distance, and the second is by construction. Combining (5.80) and (5.81) with Fannes' inequality for the quantum mutual information, we yield

$$\begin{aligned} I(K;\Lambda E^n,\rho_{K\Lambda E^n,p}) &= I(K;\Lambda B^n,\tilde{\rho}_{K\Lambda B^n,p}) \\ &\leq I(K;\Lambda B^n,\rho_{K\Lambda B^n,\pi)}) + f(n\|p-\pi\|_1) \end{aligned} \tag{5.82}$$

for each $p \in \mathfrak{P}(\mathcal{Y}) \setminus \{\pi\}$, where f is a function with $f(a) > 0$ for all $a > 0$, and $f(a) \to 0, (a \to 0)$. Therefore, we have for each $p \neq \pi$

$$\begin{aligned} \log M \;&\leq\; H(K_p) - I(K;\Lambda E^n,\rho_{K\Lambda E^n,p}) + \mu \\ &\leq\; H(K_p) - I(K;\Lambda B^n,\rho_{K\Lambda B^n,\pi}) + \mu + f(n\|p-\pi\|_1) \\ &\leq\; H(K_p) - I\left(K_\pi;K'_\pi\right) + \mu + f(n\|p-\pi\|_1) \\ &\leq\; H(K_p) - H(K_\pi) + H\left(K_\pi|K'_\pi\right) + \mu + f(n\|p-\pi\|_1) \\ &\leq\; H(K_p) - H(K_\pi) + \mu\log M + h(\mu) + \mu + f(n\|p-\pi\|_1), \\ &\leq \mu\log M + h(\mu) + \mu + 2f(n\|p-\pi\|_1), \end{aligned} \tag{5.83}$$

where the first inequality is (5.79), the second is by (5.82), the third is by the quantum data processing inequality, the fifth by Fano's inequality together with (5.78), and the last is by Fannes' inequality. By taking the infimum over all p in the above inequality arrive at

$$\log M \;\leq \mu\log M + h(\mu) + \mu. \tag{5.84}$$

We conclude, that $R = 0$ is the only achievable forward secret key distillation rate. □

Remark 62. *The lack of sender knowledge can have worst consequences. A closer look at the example introduced to prove the preceding theorem shows, how different the situations with and without sender knowledge can be. With sender knowledge, we achieve capacity with zero error and security index for each blocklength, while all public forward communication needed is the information whether π is present or not. On the other hand, the bound (5.84) reveals, that no nonzero forward secret-key rate can be achieved even if nonzero asymptotically performance μ is allowed asymptotically!*

5.4.2 A weaker notion of regularity

In this section we show that a slightly weaker condition on the set $\mathfrak{I}$ of cqq density matrices generating the outputs of the compound source is sufficient for proving a version of Theorem

48. To formulate the corresponding assertion, we introduce some notions from the theory of set-valued maps, where we take the corresponding definitions from Chapter 11 in [Bor85]. In the following we denote for each given set Ω the power set of Ω (i.e. the family of subsets of Ω) by 2^{Ω}.
Let $f:\ \Theta \to 2^{\Omega}$ be a set-valued map. We define for each $E \subset \Omega$

$$f^{+}(E) := \{\theta \in \Theta:\ f(\theta) \subset E\}, \quad \text{and} \quad f^{-}(E) := \{\theta \in \Theta:\ f(\theta) \cap E \neq \varnothing\}. \tag{5.85}$$

Definition 63. *We call a set-valued map $f:\ \Theta \to 2^{\Omega}$*

1. upper hemi-continuous*, if for each $\theta \in \Theta$ the following is true. Whenever $\theta \in f^{+}(E)$ for an open set E, there is a neighbourhood $U(\theta)$ of θ with $U(\theta) \subset f^{+}(E)$.*
2. lower hemi-continuous*, if for each $\theta \in \Theta$ the following is true. Whenever $\theta \in f^{-}(E)$ for an open set E, there is a neighbourhood $U(\theta)$ of θ with $U(\theta) \subset f^{-}(E)$.*
3. continuous*, if f is both upper and lower hemi-continuous.*

We will always regard Θ and Ω being finite-dimensional. In this case, we obtain sequential characterizations of upper and lower hemi-continuity, if we assume the set-valued function to have only compact values.

Proposition 64. *Let $f:\Theta \to 2^{\Omega}$ be a set-valued map with $\Theta \subset \mathbb{R}^m$, $\Omega \subset \mathbb{R}^k$, and $f(\theta)$ compact for each $\theta \in \Theta$. It holds*

1. *f is* upper hemi-continuous *if and only if for each $\theta \in \Theta$, every sequence $(\theta_n)_{n\in\mathbb{N}}$ with $\theta_n \to \theta (n \to \infty)$ and $\omega_n \in f(\theta_n)$, $n \in \mathbb{N}$ there is a subsequence $\{\omega_{n_k}\}_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} \omega_k \in f(\theta)$.*
2. lower hemi-continuous*, if for each $\theta \in \Theta$ and sequence $\{\theta_n\}_{n\in\mathbb{N}} \subset \Theta$, and $\omega \in f(\theta)$ from $\lim_{n\to\infty} \theta_n = \theta$ it follows, that there is a sequence $\{\omega_n\}_{n\in\mathbb{N}}$ with $\omega_n \in f(\theta_n)$, $n \in \mathbb{N}$ and $\lim_{n\to\infty} \omega_n = \omega$.*

Proof. See [Bor85], Proposition 11.11. □

For our considerations the closed-graph characterization of upper hemi-continuity will be of utility.

Theorem 65. *Let $\Theta \subset \mathbb{R}^m$, $\Omega \subset \mathbb{R}^k$, $f:\Theta \to 2^{\Omega}$ be a set-valued map with Ω being compact. If the graph of f, i.e. the set*

$$Grf := \{(\theta,\omega) \in \Theta \times \Omega:\ \omega \in f(\theta)\} \tag{5.86}$$

is closed, then f is upper hemi-continuous.

Proof. See for example Proposition 11.9 in [Bor85] □

We need the following basic Lemma.

Lemma 66. *If a set-valued function is lower hemi-continuous, then its closure $\overline{f}$ (i.e. the function defined by closing the graph of f) is lower hemi-continuous as well.*

Proof. Assume, that there is a sequence $\{\theta_n\}_{n\in\mathbb{N}}$ with $\theta_n \to \theta$ and $\omega \in \overline{f}(\theta)$, such that no sequence $\{\omega_n\}_{n\in\mathbb{N}}$ exists with $\omega_n \in \overline{f}(\theta_n)$ for all $n \in \mathbb{N}$ and $\omega_n \to \omega$. If ω is in $f(\theta)$, such a sequence always exists by lower hemi-continuity of f. If ω is in $\overline{f}(\theta) \setminus f(\theta)$ the hypothesis is only true if ω is no point of accumulation of $\overline{f}(\theta)$, which contradicts the definition of $\overline{f}$. □

Definition 67. *We call a set $\mathfrak{I} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$* weakly regular*, if the set-valued map*

$$f_{AX} : \mathcal{P}_{\mathfrak{I}} \to 2^{\mathcal{S}_{cq}(\mathcal{H}_{AX})} \tag{5.87}$$

$$p \mapsto \mathfrak{I}_p^{AX} \tag{5.88}$$

is lower hemi-continuous for $X = B, E$.

Proposition 68. *Let $\mathfrak{I} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$ be a weakly regular set of cqq density matrices. It exists a regular set $\hat{\mathfrak{I}} \subset \mathcal{S}_{cqq}(\mathcal{H}_{ABE})$ with*

1. $\mathfrak{I} \subset \hat{\mathfrak{I}}$
2. $K_{\rightarrow}(\mathfrak{I}) \leq K_{\rightarrow}(\hat{\mathfrak{I}})$.
3. $\hat{\mathfrak{I}}$ *is regular*

Proof. Assume $\mathfrak{I}$ being parameterized as

$$\mathfrak{I} := \left\{ \rho_{(p,V)} : \sum_{x\in\mathcal{X}} p(x) \left|x\right\rangle \left\langle x\right| \otimes V(x) \right\}_{(p,V)\in S} \tag{5.89}$$

with

$$S := \bigcup_{p\in\mathfrak{P}_{\mathfrak{I}}} \{p\} \times \mathcal{V}_p \tag{5.90}$$

with sets $\mathcal{P}_{\mathfrak{I}} \in \mathfrak{P}(\mathcal{X})$, $\mathcal{V}_p \subset \mathcal{CQ}(\mathcal{X}, \mathcal{H}_{BE})$, $p \in \mathcal{P}_{\mathfrak{I}}$. We define $\hat{\mathfrak{I}}$ as the closure of $\mathfrak{I}$. Obviously, the first condition $\mathfrak{I} \subset \hat{\mathfrak{I}}$ stated in the proposition is fulfilled. We show that the two remaining conditions are also fulfilled. Assume, that (T, D) is an (n, M, L, μ) forward secret-key distillation protocol for $\mathfrak{I}$. Since the performance and security criteria in Definition 43 are defined in terms of functions being continuously dependent on the cqq density matrix, it is clear, that (T, D) is an (n, M, L, μ) forward secret-key distillation protocol also for $\hat{\mathfrak{I}}$, which directly implies, that also the second claim of the proposition is satisfied.
For validating the third claim we notice, that since $\hat{\mathfrak{I}}$ is closed, the corresponding set-valued functions f_{AB} and f_{AE} have closed graphs. Therefore both maps are upper hemi-continuous by Theorem 65. The hypothesis of $\mathfrak{I}$ being weakly regular, together with Lemma 66 ensures

us, that $\overline{f}_{AB}$ and $\overline{f}_{AE}$ are also lower hemi-continuous. Therefore, they are continuous. Since the set of sender marginal distributions $\mathcal{P}_{\hat{\mathfrak{J}}}$ deriving from $\hat{\mathfrak{J}}$ is a compact set, we infer, that f_{AB}, f_{AE} are uniformly continuous, which implies, that for each $\epsilon > 0$ we find a $\delta > 0$, such that the implication

$$\|p - q\|_1 < \delta \;\Rightarrow\; d_H(f_{AB}(p), f_{AB}(q)) + d_H(f_{AE}(p), f_{AE}(q)) < \epsilon \tag{5.91}$$

for each $p, q \in \mathcal{P}_{\hat{\mathfrak{J}}}$. Since

$$d_H(\hat{\mathfrak{J}}_p^{AB}, \hat{\mathfrak{J}}_q^{AB}) + d_H(\hat{\mathfrak{J}}_p^{AE}, \hat{\mathfrak{J}}_q^{AE}) \;=\; d_H(f_{AB}(p), f_{AB}(q)) + d_H(f_{AE}(p), f_{AE}(q)) \tag{5.92}$$

holds by definition, $\hat{\mathfrak{J}}$ is regular. □

Theorem 69. *Let $\mathfrak{I}$ be a weakly regular set of cqq density matrices on $\mathcal{H}_{ABE}$. It holds*

$$K_{\rightarrow}(\mathfrak{I}) = \lim_{k\to\infty} \frac{1}{k} K_{\rightarrow}^{(1)}(\mathfrak{I}^{\otimes k}), \tag{5.93}$$

Proof. We approximate $\mathfrak{I}$ by the set $\hat{\mathfrak{J}}$ as defined in the proof of Proposition 68. The first and second property of $\hat{\mathfrak{J}}$ in Proposition 68 together imply, that

$$K_{\rightarrow}(\mathfrak{I}) = K_{\rightarrow}(\hat{\mathfrak{J}}) = \lim_{k\to\infty} \frac{1}{k} K_{\rightarrow}^{(1)}(\hat{\mathfrak{J}}^{\otimes k}) \tag{5.94}$$

holds. The rightmost of the above inequalities is by application of Theorem 48 on $\hat{\mathfrak{J}}$, which is possible, because $\hat{\mathfrak{J}}$ is regular by Proposition 68. In fact, $d_H(\mathfrak{I}, \hat{\mathfrak{J}}) = 0$, and consequently $d_H(\mathfrak{I}^{\otimes k}, \hat{\mathfrak{J}}^{\otimes k}) = 0$ holds for each $k \in \mathbb{N}$. Therefore

$$K_{\rightarrow}^{(1)}(\mathfrak{I}^{\otimes k}) = K_{\rightarrow}^{(1)}(\hat{\mathfrak{J}}^{\otimes k}) \tag{5.95}$$

holds for each $k \in \mathbb{N}$ by continuity of $K^{(1)}$. We are done. □

5.5 Special case: Forward secret-key distillation capacity of a classical tripartite compound sources

Our results also cover the case of a completely classical tripartite source. Let (X, Y, Z) be a triple of classical random variables with distribution $P_{XYZ} \in \mathfrak{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$. The state of this classical system coherified to a Hilbert space $\mathcal{H}_X \otimes \mathcal{H}_Y \otimes \mathcal{H}_Z$ is represented by the density matrix

$$\rho := \sum_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} P_{XYZ}(x,y,z)\, |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes |z\rangle\langle z|\,. \tag{5.96}$$

Forward secret-key distillation for this kind of classical compound memoryless source was considered in the previous work [TBS16] done under collaboration of on of the authors of this

paper. Among other results obtained therein, it was derived a capacity formula for the case without sender marginal knowledge in case, that the set of sender marginal distributions deriving from the source is finite. Our results extend the capacity description also to the case of an arbitrary regular tripartite classical source. By coherifying each set $(X_s, Y_s, Z_s)_{s\in S}$ of triples into a mutually commuting set of density matrices as in (5.96), Theorem 48 directly leads to a version of Theorem 2 in [TBS16] where instead of assuming finite cardinality of the set of marginal states the assumption is that the compound source is regular. Theorem 49 in the same way provides a capacity formula for the case where the sender party perfectly knows the distribution of his/her part of the source. The reader may reply, that the definition of a (deterministic) classical decoding procedure is more restricting than that of a POVM in quantum theory, since the decoding sets are demanded to be pairwise non-intersecting. Our reasoning is not affected by this fact, because in case of pairwise commuting density operators, optimal decoding can always be achieved by using projection valued measures. Alternatively, on could replace Lemma 51 by a completely classical version.
We point out, that the need for a regularity condition as well as differences between the capacities with and without sender's knowledge of the marginal state are not effects of the quantum nature of the sources considered in this paper. The reader may note, that our counterexample given to prove Theorem 61 is essentially classical, since all density matrices involved pairwise commute.
Since there are stronger results known in classical information theory (especially regarding error exponents for coding of classical compound channels), a classical method of proof will lead to faster decrease of error with a potentially simplified proof.

5.6 Conclusion

We have considered the the task of secret key distillation under free forward classical communication for compound memoryless sources with classical legitimate sender and quantum legal receiver and eavesdropper outputs. We derived a capacity formula for all sources of this class which additionally exhibit a certain regularity condition.
We also discussed the situation, where the legitimate sender has perfect knowledge of the probability distribution governing his/her outputs. In this case, we were able to derive a capacity formula which equals the one given for the case without sender marginal information for regular sources, and moreover does hold for all non-regular sources.
As we have also seen, the capacities with and without sender marginal information differ at least for some non-regular sources. We admit, that the regularity conditions assumed in this paper may be somewhat weakened to determine the forward secret-key generation capacity without sender knowledge for a larger class of sources. We provided a further step in this direction by applying the general theory of set-valued maps to derive a slightly broader class of compound cqq sources with a general capacity description.
We leave open the more general case of proving a capacity theorem for forward secret-key distillation from compound sources where the generating set of density matrices may contain members being not in the class of cqq density matrices. In [DW05] such a theorem were proven

in case of a perfectly known source without restriction on the legitimate sender to be classical.
The strategy therein to prove a coding theorem was, to combine an achievability result for cqq sources with an optimization over instruments dephasing the sender's system to a classical one. Notice, that such a strategy in general does not apply to compound sources in a direct way as it did in [DW05] (at least in case that the sender does not have perfect marginal knowledge). In general, there is no control whether or not a dephasing operation leads to a non-regular compound cqq source.
However, approximation techniques presented here may lead to a better understanding of the secret-key distillation task even for tripartite quantum compound sources.
Another astonishing result from Ref.[DW05] is the close correspondence between the forward secret-key distillation and entanglement distillation tasks. It was demonstrated, that modifying forward secret-key distillations leads to one-way local operations and classical communications protocols suitable for proving the so-called hashing bound eventually determining the entanglement distillation capacity of bipartite memoryless quantum sources.
The authors of this paper are of the opinion, that following a similar strategy to derive the entanglement distillation capacity of bipartite compound memoryless quantum sources may be not successful in the same way as it is with perfect knowledge of the source. A closer look to the corresponding considerations in [Dev05] may underpin this opinion. Therein, an important part of the coding strategy was to apply a nondestructive measurement on the sender's marginal of the bipartite quantum state subject to entanglement distillation – a strongly state-dependent task, which can hardly be performed without sender marginal knowledge. For a generalization to the case of bipartite compound quantum case under assumption of SMI, the strategy may be feasible. We did not pursue this path, because the one-way entanglement distillation capacity of compound quantum sources is already known from [BBJ13] and [BJ14a].
This parallels a similar observation made for channel coding from [BN13]. Therein, it was argued, that the ingenuous way to derive entanglement generation codes for quantum channels from codes for secret message transmission over classical-quantum wiretap channels used in [DW05] for proving the quantum coding theorem leads to suboptimal results for compound channels if no sender state knowledge is assumed.

A Appendix A

A.1 Universal random constant composition codes for compound DMCQ Channels

In this section, we state and prove some results on compound DMcq channels we need within the proof of Lemma 51. For convenience of the reader, we first provide definitions, necessary to understand the subsequent arguments. Let $\mathcal{V} \subset CQ(\mathcal{X}, \mathcal{K})$ be a set of cq channels mapping a finite alphabet $\mathcal{X}$ to the set of density matrices on a Hilbert space $\mathcal{K}$. The *compound discrete memoryless classical quantum channel generated by* $\mathcal{V}$ (the *DMcqC* $\mathcal{V}$ for short) is given by the family $\{V^{\otimes n} : V \in \mathcal{V}\}_{n\in\mathbb{N}}$ of possible outputs. To catch up with the notation in [BB09], wie sometimes write $\mathcal{V} = \{V_s\}_{s\in S}$ assuming a suitable parameterization of $\mathcal{V}$ by an index set S. For given blocklength $n \in \mathbb{N}$, $M \in \mathbb{N}$, an (n, M)*-code for transmission of classical messages over* $\mathcal{V}$ is a family $\mathcal{C} := (u_m, D_m)_{m=1}^M$ with $u_m \in \mathcal{X}^n$, $D_m \in \mathcal{L}(\mathcal{K}^{\otimes n})$ for each $m \in [M]$, with the additional property, that for all $m \in [M]$

$$0 \le D_m \le \mathbb{1}_{\mathcal{K}}^{\otimes n}, \quad \text{and} \quad \sum_{m=1}^{M} D_m \le \mathbb{1}_{\mathcal{K}}^{\otimes n}$$

holds. For given (n, M)-code $\mathcal{C}$, $s \in S$, we define the *average error of transmission* by

$$\overline{e}(\mathcal{C}, V_s^{\otimes n}) := \frac{1}{M} \sum_{m=1}^{M} \operatorname{tr}(D_m^{\perp} V^{\otimes n}(u_m)),$$

where we allow ourselves to define $A^{\perp} := \mathbb{1}_{\mathcal{K}}^{\otimes n} - A$ (even if A is not a projector). The following two Propositions combined, immediately imply the proof of Lemma 51 stated in the text. The claim of the first one follows by careful modification of the proof of Theorem 5.10 in [BB09] and delivers for each large enough blocklength and each type of sequences a suitable random compound transmission code having superpolynomially decrease of error, universal regarding the channels in the compound set, as well as the appearing types.

Proposition 70. *Let $\mathcal{W} \subset CQ(\mathcal{X}, \mathcal{K})$ be an arbitrary set of cq channels. For each $\delta > 0$, there is a number $n_2 := n_2(\delta)$ such that for each $n > n_2$ and each type $\lambda \in \mathfrak{T}(n, \mathcal{X})$ there exists a random (n, M_λ)-code $\mathcal{C}_\lambda(U) := (U_m, D_m(U))_{m=1}^{M_\lambda}$ for $\mathcal{W}$ with the following properties*

1. *$U := (U_1, ..., U_{M_\lambda})$ is an i.i.d. sequence of random variables, each with values in $\mathcal{X}^n$ and distribution $\lambda^{\otimes n}$.*
2. *$M_\lambda \ge \exp\left\{n\left(\inf_{W\in\mathcal{W}} \chi(\lambda, W) - \delta\right)\right\}$.*

3. $\mathbb{E}\left[\sup_{W\in\mathcal{W}} \overline{e}(\mathcal{C}_\lambda, W^{\otimes n})\right] \leq 2^{-\sqrt[16]{n}c(\delta)}$

with a constant $c(\delta) > 0$.

Proof. The assertion is basically contained in the proof of Theorem 5.10 in [BB09] and follows by minor modifications of the argument given there. We assume the reader's familiarity with the arguments presented in [BB09] and restrict ourselves to indicate the steps of modification necessary to justify our claim.
We write $\mathcal{W} := \{W_t\}_{t\in T}$ with a suitable index set T. Let $n \in \mathbb{N}$, and consider a type $\lambda \in \mathfrak{T}(n, \mathcal{X})$. Assume, that

$$\inf_{t\in T} \chi(\lambda, W_t) - \delta > 0, \tag{A.1}$$

holds, since otherwise the claim is trivially fulfilled for λ. Choose an approximating set $\mathcal{W}_n := \{W'_t\}_{t\in T_n}$ for $\mathcal{W}$ as used in [BB09]. We will execute the suggestions given in Remark 5.11 in [BB09], and therefore choose the diameter of the partition of output states used to define $\mathcal{W}_n$ to be $2^{-\sqrt[16]{n}}$ instead of $\frac{1}{n^2}$. Define

$$\Omega_{\lambda,n} := \{\rho'_{t,\lambda} := \sum_{x\in\mathcal{X}} \lambda(x) |x\rangle\langle x| \otimes W'_t(x) \ : \ t \in T_n\},$$

$$\lambda := \sum_{x\in\mathcal{X}} \lambda(x) |x\rangle\langle x|, \ \text{ and } \ \sigma'_{t,\lambda} := \sum_{x\in\mathcal{X}} \lambda(x)\sigma'_t(x) \qquad (t \in T_n).$$

Note, that the properties of the set $\mathcal{W}_n$ in [BB09] (see Lemma 5.6 therein) together with the above definitions, implies the bound

$$\begin{aligned}\lambda_{\min}(\lambda \otimes \sigma'_{t,\lambda}) &\geq \min_{x\in\mathrm{supp}(\lambda)} \lambda(x) \cdot \frac{1}{d} \cdot 2^{-\sqrt[16]{n}} \\ &\geq \frac{2^{-\sqrt[16]{n}}}{n \cdot d}\end{aligned} \tag{A.2}$$

on the minimal eigenvalue $\lambda_{\min}(\lambda \otimes \sigma'_{t,\lambda})$ of $\lambda \otimes \sigma'_{t,\lambda}$. The last estimate above follows from the fact, that λ is a type of sequences in $\mathcal{X}^n$. Closely following the argument given in [BB09], we are ensured, that choosing $l_n = [\sqrt{n}]$, we find $a_n, b_n \in \mathbb{N}$, $0 \leq b_n < l_n$, with

$$n = a_n l_n + b_n,$$

and a PVM $\mathcal{M}_{l_n,\lambda} := \{P_{1,l_n,\lambda}, P_{2,l_n,\lambda}\}$, such that for all $t, s \in T_n$

$$\begin{aligned}S_{M_{l_n,\lambda}}(\rho'^{\otimes l_n}_{t,\lambda} \| (\lambda \otimes \sigma_{\lambda,s})^{\otimes l_n}) &\geq l_n(S(\Omega_{\lambda,n} \| \lambda \otimes \sigma'_{\lambda,s}) - \xi_{l_n}(\lambda \otimes \sigma'_{\lambda,s}) \\ &\geq l_n(\min_{t\in T_n} \chi(\lambda, W'_t) - \xi_{l_n}(\lambda \otimes \sigma'_{\lambda,s}))\end{aligned} \tag{A.3}$$

holds. Careful investigation of the function ξ_{l_n} in [BB09] using the type-independent bound in (A.2) shows that

$$\lim_{n\to\infty} \max_{\lambda\in\mathfrak{T}(n,\mathcal{X})} \max_{s\in T_n} \xi_{l_n}(\lambda\otimes\sigma'_{\lambda,s}) = 0 \tag{A.4}$$

holds. Introducing the refinement $\mathcal{Q}_{l_n,\lambda}$ of the PVM $\mathcal{M}_{l_n,\lambda}$, and the stochastic matrices $V_{t,\lambda}$, $t \in T_n$ generated by $Q_{l_n,\lambda}$ as in [BB09] (note, that these, also may depend on the chosen type λ), it holds

$$\frac{1}{l_n}\min_{t\in T_n} I(\lambda^{\otimes l_n}, V_{t,\lambda}) \geq \inf_{t\in T} \chi(\lambda, W_t) - n\cdot 2^{-\sqrt[16]{n}} C(d) - \xi_{l_n,\max},$$

where we used $\xi_{l_n,\max}$ defined by

$$\xi_{l_n,\max} := \max_{\lambda\in\mathfrak{T}(n,\mathcal{X})} \max_{s\in T_n} \xi_{l_n}(\lambda\otimes\sigma'_{\lambda,s}).$$

Since (A.4) holds, we find for each $0 < \eta < \delta$ a number $n_2(\eta) \in \mathbb{N}$ (independent of λ) such that

$$\frac{1}{l_n}\min_{t\in T_n} I(\lambda^{\otimes l_n}, V_{t,\lambda}) \geq \inf_{t\in T} \chi(\lambda, W_t) - \eta > 0 \tag{A.5}$$

is fulfilled for all $n > n_2$. Note, that the last inequality holds by (A.1). The bound above differs a bit from the one given in Eq. (30) in [BB09]. However, it will be sufficient for the following argument. Let

$$\Theta := \left\{\theta \in \mathbb{R} :\ 0 < \theta < \frac{\eta}{4}\right\}, \tag{A.6}$$

and

$$I_{n,\lambda} := \min_{t\in T_n} I(\lambda^{\otimes l_n}, V_{t,\lambda}).$$

Following the lines of [BB09] (always respecting dependencies on λ), we yield the bound

$$\Pr(i_\lambda^{a_n} \leq I_{n,\lambda} - 2l_n\theta) \leq \frac{1}{|T_n|}\sum_{t\in T_n} \Pr_{t,\lambda}(i_{t,\lambda}^{a_n} \leq I_n - l_n\theta) + |T_n| 2^{-a_n l_n \theta}.$$

Since $i_{t,\lambda}^{a_n}$ is a sum of i.i.d. random variables each with values in the interval

$$[-l_n d\sqrt[16]{n}, l_n d\sqrt[16]{n}],$$

and

$$I_{n,\lambda} \leq \mathbb{E}_{t,\lambda}(i_{t,\lambda}^{a_n})$$

holds for each $t \in T_n$, our counterpart to eq. (34) in [BB09],

$$\Pr(i_{t,\lambda}^{a_n} \leq I_n - l\theta) \leq e^{\frac{a_n\theta^2}{16\cdot\sqrt[16]{n}}}$$

is valid. By closely following the lines of [BB09] (having in mind our bounds) together with the choice $\theta = \frac{\eta}{4}$, we know, that there is a projection $P_{n,\lambda,\theta}$ with

$$\mathrm{tr}(\rho_\lambda^{(n)} P_{n,\lambda,\theta}) \geq 1 - e^{\frac{a_n \eta^2}{16^2 \sqrt[16]{n}}} - |T_n| 2^{-a_n l_n \frac{\eta}{4}}$$

and

$$\begin{aligned} \mathrm{tr}((\lambda^{\otimes n} \otimes \sigma_\lambda^{(n)}) P_{n,\lambda,\theta}) &\leq 2^{-a_n (I_n - 2 l_n \theta)} \\ &\leq 2^{-a_n l_n (\inf_{t \in T} \chi(\lambda, W_t) - \frac{3}{2}\eta)}, \end{aligned} \tag{A.7}$$

where the last of the above inequalities above holds by (A.5). Since $b < \sqrt{n}$,

$$\mathrm{tr}((\lambda^{\otimes n} \otimes \sigma_\lambda^{(n)}) P_{n,\lambda,\theta}) \leq 2^{-n(\inf_{t \in T} \chi(\lambda, W_t) - 2\eta)} \tag{A.8}$$

holds for each $n > n_2$, if n_2 is chosen large enough. Setting $\eta := \frac{\delta}{2}$, and using the bounds in (A.7) and (A.8) together with Theorem 1.1. stated in the Appendix of [BB09] (and leaving out the derandomization step leading to a deterministic code in the Proof of Theorem 5.10 in [BB09]), we conclude that there is an (n, M'_λ) random code $\mathcal{C}_\lambda(U) := (U_m, D_m(U))_{m=1}^{M'}$ for the average channel $W^{\otimes n} := \frac{1}{|T_n|} \sum_{t \in T_n} W'^{\otimes n}_t$ with $U = (U_1, \ldots, U_{M'})$ being a sequence of i.i.d. random variables each with values in $\mathcal{X}^n$ and generic distribution $\lambda^{\otimes n}$, such that

1. $M \geq \left\lfloor \exp\left\{ n \left(\inf_{t \in T} \chi(\lambda W_t) - \frac{3}{2}\delta \right) \right\} \right\rfloor$.
2. $\mathbb{E}\left[\overline{e}(\mathcal{C}_\lambda, W^{\otimes n}) \right] \leq 2^{-\sqrt[16]{n} \tilde{c}(\delta)}$

with a positive constant $\tilde{c}(\delta) > 0$. From this, we can conclude, that there is a number $n_0(\delta)$, such that for each $n > n_0(\delta)$ (independent of λ) $\mathcal{C}_\lambda$ is an (n, M) random code of the above stated properties, with 2. above replaced by

$$\mathbb{E}\left[\sup_{W \in \mathcal{W}} \overline{e}(\mathcal{C}_\lambda, W^{\otimes n}) \right] \leq 2^{-\sqrt[16]{n} c(\eta)},$$

where $c(\delta) := \frac{1}{2}\tilde{c}(\delta)$. □

The following proposition is a compound version of Proposition 2.5 in [Dev05]. It is proven by exactly the same strategy replacing the Holevo-Schumacher-Westmoreland codes for $DMcqC$ with perfectly known generic cq channel by the codes constructed in [BB09] together with the modifications done in Proposition 70 above.

Proposition 71 (cf. [Dev05], Prop. 2.5)**.** *Let $\mathcal{W} \subset CQ(\mathcal{X}, \mathcal{K})$ be an arbitrary set of cq channels, $n \in \mathbb{N}$ and $\lambda \in \mathfrak{T}(n, \mathcal{X})$ a type of sequences in $\mathcal{X}^n$. If there exists a random random (n, M')-classical message transmission code*

$$\mathcal{C}'(U) = (U_m, D_m(U))_{m=1}^{M'}$$

for the DMcqC $\mathcal{W}$ which has the properties

1. $U := (U_1, \dots, U_{M'})$ *is an i.i.d. sequence of random variables with values in* $\mathcal{X}^n$ *with generic distribution* λ^n
2. $\mathbb{E}\left[\sup_{W \in \mathcal{W}} \overline{e}(\mathcal{C}, W^{\otimes n})\right] \leq \mu$

with $\mu \in (0,1)$*, then there exists for each given* $\vartheta \in (0,1)$ *a random* (n, M)*-message transmission code*

$$\mathcal{C}(V) := (V_m, D_m(V))_{m=1}^{M}$$

having the properties

1. $V := (V_1, ..., V_M)$ *is an i.i.d. sequence of random variables, each equidistributed on* T_λ^n*,*
2. $M = \lfloor \vartheta \cdot (n+1)^{-|\mathcal{X}|} \cdot M' \rfloor$*,*
3. *and*

$$\mathbb{E}\left[\sup_{W \in \mathcal{W}} \overline{e}(\mathcal{C}', W^{\otimes n})\right] \leq \frac{2}{\vartheta}(n+1)^{|\mathcal{X}|}\mu + 2^{-M'(1-\vartheta)^2(n+1)^{-|\mathcal{X}|}/\ln 2}.$$

For provint the above assertion, we will make use of the following variant of the Chernov bound

Proposition 72. *Let* $n \in \mathbb{N}$*,* $\delta > 0$ *and* $X = (X_1, \dots, X_n)$ *be an i.i.d. sequence of random variables with* $0 \leq X_1, \dots, X_n \leq 1$ *and* $\mathbb{E}[X_i] = E$*,* $i \in [n]$*, then*

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq (1-\delta)E\right) \leq 2^{-n\delta^2 E^2/\ln 2}$$

Proof of Proposition 71. Define the event, that at least M codewords of a codebook u are λ-typical sequences by

$$A(M) := \left\{u = (u_1, \dots, u_{M'}) \in \mathcal{X}^{nM'} : \ |\{m : \ u_m \in T_\lambda^n\}| \geq M\right\}.$$

For an i.i.d. sequence $U = (U_1, \dots, U_{M'})$ as in the hypotheses of the proposition, it holds

$$\Pr\left(A(M)^c\right) = \Pr\left(\sum_{i=1}^{M'} \mathbb{1}_{T_\lambda^n}(U_m) < M\right).$$

Notice, that

$$\mathbb{E}[\mathbb{1}_{T_\lambda^n}(U_m)] = \lambda^n(T_\lambda^n) \geq (n+1)^{-|\mathcal{X}|}$$

holds. The rightmost inequality above holds by the fact, that among all typical sets of words in $\mathcal{X}^n$, T^n_λ has the largest probability w.r.t. λ^n.
We fix $\vartheta \in (0,1)$ and set $M := \lfloor \vartheta(n+1)^{-|\mathcal{X}|} M' \rfloor$. We obtain

$$\begin{aligned}
\Pr\left(A(M)^c\right) &\leq \Pr\left(\sum_{i=1}^{M'} \mathbb{1}_{T^n_\lambda}(U_m) < M\right) \\
&\leq \Pr\left(\frac{1}{M'}\sum_{i=1}^{M'} \mathbb{1}_{T^n_\lambda}(U_m) < \vartheta(n+1)^{-|\mathcal{X}|}\right) \\
&\leq \Pr\left(\frac{1}{M'}\sum_{i=1}^{M'} \mathbb{1}_{T^n_\lambda}(U_m) < \vartheta\mathbb{E}[U_1]\right) \\
&\leq \exp\left\{-M'(1-\vartheta)^2\mathbb{E}[U_1]/\ln 2)\right\} \qquad \text{(A.9)} \\
&\leq \exp\left\{-M'(1-\vartheta)^2(n+1)^{-|\mathcal{X}|})/\ln 2\right\} =: \tau.
\end{aligned}$$

Except the one in Eq. (A.9), which is by application of the bound in Proposition 72, all of the above estimates are by the preceding definitions and bounds.
Next, define a function $\varphi : \mathcal{X}^{nM'} \to (T^n_\lambda)^M$ by

$$\varphi(u) := \begin{cases} (v_1, \ldots, v_M) = v & \text{if } u \in A(M) \\ (\tilde{v}, \ldots \tilde{v}) & \text{otherwise,} \end{cases}$$

where $\tilde{v}$ is any word from T^n_λ. By symmetry, it holds

$$\lambda^{nM'}(\varphi^{-1}(v)) = \frac{\lambda^{nM'}(A(M'))}{|T^n_\lambda|},$$

i.e. the push forward measure $\lambda^{nM'} \circ \varphi^{-1}$ of $\lambda^{nM'}$ under φ is nearly equidistributed. Explicitely,

$$\left\| \pi^{\otimes M}_{T^n_\lambda} - \lambda^{nM'} \circ \varphi^{-1} \right\|_1 = |1 - \lambda^{nM'}(A(M))| = \lambda^{nM'}(A(M)^c) < \tau.$$

For each outcome $v = (v_1, \ldots, v_M) \in T^{nM}_\lambda$, we define an (n, M)-message transmission code $\mathcal{C}(v) := (v_{m'}, D_{m'})_{m'=1}^M$ as follows. Let

$$u_v := \operatorname*{argmin}_{u \in \varphi^{-1}(v)} \sup_{t \in T} \overline{e}(\mathcal{C}'(u), W_t^{\otimes n}), \qquad \text{(A.10)}$$

and $D_{m'} := D_m(u_v)$ for the respective m (i.e. $(u_{v,m}, D_m(u_v))$ is a pair of codeword and

decoding set in $\mathcal{C}'(u_v)$.) Then

$$\begin{aligned}
\sup_{t\in T} \overline{e}(\mathcal{C}(v), W_t^{\otimes n}) &= \sup_{t\in T} \frac{1}{M} \sum_{m=1}^{M} \operatorname{tr}\left(D_m(v)^{\perp} W_t^{\otimes n}(v_m)\right) \\
&\leq \sup_{t\in T} \frac{1}{M} \sum_{m=1}^{M'} \operatorname{tr}\left(D_m^{\perp}(u_v) W_t^{\otimes n}(u_{v,m})\right) \\
&\leq \frac{M'}{M} \sup_{t\in T} \overline{e}(\mathcal{C}'(u_v), W_t^{\otimes n}) \\
&= \frac{M'}{M} \min_{u\in\varphi^{-1}(v)} \sup_{t\in T} \overline{e}(\mathcal{C}'(u_v), W_t^{\otimes n}).
\end{aligned}$$

The last equality above is by our code definition from Eq. (A.10). To each $u \in A(M)^c$, we assign some valid (n, M)-code $\mathcal{C}(\varphi(u)) := (v_{0,m}, D_m)_{m=1}^{M}$ being of no further interest. Let $\hat{V} = \varphi(U)$ (which is not i.i.d. so far!), then $\mathcal{C}(\hat{V}_m, D_m(\hat{V}))_{m=1}^{M}$ is a random constant composition (n, M)-code with

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in T} \overline{e}(\mathcal{C}(\hat{V}), W_t^{\otimes n})\right] &= \sum_{v\in\varphi(A(M))} \lambda^{nM'}(\varphi^{-1}(v)) \sup_{t\in T} \overline{e}(\mathcal{C}(v), W_t^{\otimes n}) \\
&\quad + \sum_{v\in\varphi(A(M)^c)} \lambda^{nM'}(\varphi^{-1}(v)) \sup_{t\in T} \overline{e}(\mathcal{C}(v), W_t^{\otimes n}) \\
&< \sum_{v\in\varphi(A(M))} \lambda^{nM'}(\varphi^{-1}(v)) \sup_{t\in T} \overline{e}(\mathcal{C}(v), W_t^{\otimes n}) + \tau \\
&\leq \frac{M'}{M} \sum_{u\in A(M)} \lambda^{nM'}(u) \sup_{t\in T} \overline{e}(\mathcal{C}'(u), W_t^{\otimes n}) + \tau \\
&\leq \frac{M'}{M} \mathbb{E}\left[\sup_{t\in T} \overline{e}(\mathcal{C}'(u), W_t^{\otimes n})\right] + \tau \\
&\leq \frac{M'}{M}\mu + \tau.
\end{aligned}$$

Now, let $V = (V_1, \ldots, V_M)$ be a sequence of i.i.d. random variables each equidistributed on T_λ^n. And $\mathcal{C}(V) := (V_m, D_m(V))_{m=1}^{M}$. It holds

$$\begin{aligned}
\frac{M'}{M} \mathbb{E}\left[\sup_{t\in T} \overline{e}(\mathcal{C}'(u), W_t^{\otimes n})\right] &\leq \frac{M'}{M} \mathbb{E}\left[\sup_{t\in T} \overline{e}(\mathcal{C}'(u), W_t^{\otimes n})\right] \\
&\quad + \left\| \pi_{T_\lambda^n}^{\otimes M} - \lambda^{nM'} \circ \varphi^{-1} \right\|_1 \\
&< \frac{M'}{M}\mu + 2\tau.
\end{aligned}$$

Since

$$\frac{M'}{M} \leq \frac{2}{\vartheta}(n+1)^{|\mathcal{X}|},$$

we are done. □

A.2 Continuity bounds

For convenience of the reader, we state and prove several continuity properties of entropic quantities. Most of them follow straightforwardly from the Alicki-Fannes bound [AF04] for von Neumann entropies.

Theorem 73 ([AF04])**.** *Let $\rho, \sigma \in \mathcal{S}(\mathcal{K}_A \otimes \mathcal{K}_B)$ be states on $\mathcal{K}_A \otimes \mathcal{K}_B$ with $\|\rho - \sigma\| \leq \epsilon$. It holds*

$$|S(A|B,\rho) - S(A|B,\sigma)| \leq 4\epsilon \log \dim \mathcal{K}_A + 2h(\epsilon),$$

where $h(x) := -x \log x - (1-x)\log(1-x)$ is the binary Shannon entropy of $(x, 1-x)$.

The following bound is easily derived from Theorem 73.

Lemma 74. *Let $p, q \in \mathfrak{P}(\mathcal{Y})$ be probability distributions with $\|p - q\|_1 \leq \epsilon$. For each cq-channel $V \in \mathcal{CQ}(\mathcal{Y}, \mathcal{K})$, it holds*

$$|\chi(p,V) - \chi(q,V)| \leq 6\epsilon \log \dim \mathcal{K} + 2h(\epsilon).$$

Lemma 75. *Let $\mathcal{Q}, \mathcal{Q}' \subset \mathfrak{P}(\mathcal{Y})$ be probability distributions with $d_H(\mathcal{Q}, \mathcal{Q}') \leq \epsilon$. For each set $\mathcal{V} \subset \mathcal{CQ}(\mathcal{Y}, \mathcal{K}_B \otimes \mathcal{K}_E)$*

$$\begin{aligned}&\left|\inf_{q\in\mathcal{Q}}\left(\inf_{V\in\mathcal{V}}\chi(p,V_B) - \sup_{V\in\mathcal{V}}\chi(q,V_E)\right) - \inf_{q\in\mathcal{Q}'}\left(\inf_{V\in\mathcal{V}}\chi(p,V_B) - \sup_{V\in\mathcal{V}}\chi(q,V_E)\right)\right| \\ &\leq 6\epsilon \log \dim \mathcal{K}_{BE} + 4h(\epsilon).\end{aligned}$$

Lemma 76. *Let $\mathfrak{I}, \mathfrak{J} \subset \mathcal{S}_{cqq}(\mathcal{Y}, \mathcal{K}_X)$ of cqq density matrices and $d_H(\mathfrak{I}, \mathfrak{J}) \leq \Gamma$, and with stochastic matrices $P_{U|Y} : \mathcal{Y} \to \mathfrak{P}(\mathcal{U})$ and $P_{T|U} : \mathcal{U} \to \mathfrak{P}(\mathcal{T})$*

$$\tilde{\rho} := \sum_{t\in\mathcal{T}} \sum_{u\in\mathcal{U}} \sum_{y\in\mathcal{Y}} P_{T|U}(t|u) P_{U|Y}(u|y) p(y) \, |u\rangle\langle u| \otimes |t\rangle\langle t| \otimes V(y)$$

if

$$\rho := \sum_{y\in\mathcal{Y}} p(y) \, |y\rangle\langle y| \otimes V(y).$$

Then, the inequalities

$$\begin{aligned}\inf_{\rho\in\mathfrak{J}} I(U;X|T,\tilde{\rho}) &\geq \inf_{\rho\in\mathfrak{J}} I(U;X|T,\tilde{\rho}) - 8\Delta \log(|\mathcal{U}| \cdot \dim \mathcal{K}) - 6h(\Delta) \\ \sup_{\rho\in\mathfrak{J}} I(U;X|T,\tilde{\rho}) &\leq \sup_{\rho\in\mathfrak{J}} I(U;X|T,\tilde{\rho}) + 8\Delta \log(|\mathcal{U}| \cdot \dim \mathcal{K}) + 6h(\Delta)\end{aligned}$$

are valid.

Proof. It holds for any two states $\rho, \sigma \in \mathcal{S}_{cqq}(\mathcal{Y}, \mathcal{K}_X)$

$$\|\tilde{\rho} - \tilde{\sigma}\|_1 \leq \|\rho - \sigma\|_1. \tag{A.11}$$

Note, that for each cqq density matrix ρ, it holds

$$I(U;X|T,\tilde{\rho}) = S(U|T,\tilde{\rho}) + S(X|T,\tilde{\rho}) - S(UX|\tilde{\rho}).$$

by definition of the quantum mutual information. If ρ, σ fulfill $\|\rho - \sigma\|_1 \leq \delta$, then

$$|I(U;X|T,\tilde{\rho}) - I(U;X|T,\tilde{\sigma})| \leq 8\delta \log(|\mathcal{U}| \cdot \dim \mathcal{K}) + 6h(\delta) \tag{A.12}$$

holds by (A.11) and application of Lemma 73. From (A.12) and the assumptions, we directly infer the claims. □

A.3 Proof of Eq. (5.61)

Let $\lambda \in \hat{\mathfrak{T}}$ and $p \in \mathcal{P}_\lambda$. We define for each $q \in \mathcal{P}_\lambda$ the set

$$\mathfrak{I}^B_{(p,q)} := \left\{ \sum_{y \in \mathcal{Y}} p(y) \left|y\right\rangle \left\langle y\right| \otimes V_B : \; V \in \mathcal{V}_q \right\}.$$

Observe the relations

$$\hat{\mathfrak{I}}^B_{p,\lambda} = \bigcup_{q \in \mathcal{P}_\lambda} \mathfrak{I}^B_{(p,q)}, \quad \text{and} \quad \mathfrak{I}^B_p = \mathfrak{I}^B_{(p,p)}.$$

assume $V \in \mathcal{V}_p$, $V' \in \mathcal{V}_q$. It holds

$$\begin{aligned}\left\| \sum_{y \in \mathcal{Y}} p(y) \left|y\right\rangle \left\langle y\right| \otimes V(y) - \sum_{y \in \mathcal{Y}} p(y) \left|y\right\rangle \left\langle y\right| \otimes V'(y) \right\|_1 &= \sum_{y \in \mathcal{Y}} \|p(y)V(y) - p(y)V'(y)\|_1 \\ &\leq \sum_{y \in \mathcal{Y}} \|p(y)V(y) - q(y)V'(y)\|_1 + \|p - q\|_1,\end{aligned}$$

from which we directly infer

$$d_H(\mathfrak{I}^B_p, \mathfrak{I}_{(p,q)}) \leq d_H(\mathfrak{I}^B_p, \mathfrak{I}^B_q) + \|p - q\|_1.$$

Using these facts together with the first claim of Lemma 76, we obtain

$$\begin{aligned}\inf_{\rho \in \hat{\mathfrak{I}}_{p,\lambda}} I(U_\lambda; B|T_\lambda, \tilde{\rho}) &= \inf_{\sigma \in \hat{\mathfrak{I}}^B_{p,\lambda}} I(U_\lambda; B|T_\lambda, \tilde{\sigma}) \\ &= \inf_{q \in \mathcal{P}_\lambda} \inf_{\sigma \in \mathfrak{I}^B_{(p,q)}} I(U_\lambda; B|T_\lambda, \tilde{\sigma}) \\ &\geq \inf_{\sigma \in \mathfrak{I}^B_p} I(U_\lambda; B|T_\lambda, \tilde{\sigma}) - 8\Delta \log(|\mathcal{U}| \cdot \dim \mathcal{K}_B) - 6h(\Delta) \\ &= \inf_{\rho \in \mathfrak{I}_p} I(U_\lambda; B|T_\lambda, \tilde{\rho}) - 8\Delta \log(|\mathcal{U}| \cdot \dim \mathcal{K}_B) - 6h(\Delta)\end{aligned} \tag{A.13}$$

Applying a similar reasoning leads us to

$$\sup_{\rho\in\hat{\mathfrak{J}}_{p,\lambda}} I(U_\lambda;E|T_\lambda,\tilde{\rho}) \leq \sup_{\rho\in\mathfrak{J}_p} I(U_\lambda;E|T_\lambda,\tilde{\rho}) - 8\Delta\log(|\mathcal{U}|\cdot\dim\mathcal{K}_E) - 6h(\Delta). \tag{A.14}$$

Combination of (A.13) and (A.14) for all $p \in \mathcal{P}_\lambda$ yields the desired bound

$$\begin{aligned}&\inf_{\rho\in\hat{\mathfrak{J}}_{p,\lambda}} I(U_\lambda;B|T_\lambda,\tilde{\rho}) - \sup_{\rho\in\hat{\mathfrak{J}}_{p,\lambda}} I(U_\lambda;E|T_\lambda,\tilde{\rho})\\ &\geq \inf_{\rho\in\mathfrak{J}_p} I(U_\lambda;B|T_\lambda,\tilde{\rho}) - \sup_{\rho\in\mathfrak{J}_p} I(U_\lambda;E|T_\lambda,\tilde{\rho}) - 16\Delta\log(|\mathcal{U}|\cdot\dim\mathcal{K}_{BE}) - 12h(\Delta).\end{aligned}$$

B Appendix B

B.1 Proof of the bound in Eq. (3.36)

Let $\eta > 0$ be fixed and p, q probability distributions on $[d]$, such that

$$|H(p) - H(q)| \geq \eta \tag{B.1}$$

holds. It is well known, that the Shannon entropy is uniformly continuous in the variation distance (see e.g. [CK11]), it holds

$$|H(p) - H(q)| \leq f(\|p - q\|_1) \tag{B.2}$$

with a strictly monotonically increasing function f. Therefore, (B.1) and (B.2) lead to

$$0 < 2c_3 := \frac{1}{2\ln 2} f^{-1}(\eta)^2 \leq \frac{1}{2\ln 2}\|p - q\|_1^2 \leq D(p\|q), \tag{B.3}$$

where the rightmost inequality is Pinsker's inequality $D(p\|q) \geq \frac{1}{2\ln 2}\|p - q\|_1^2$. Since p and q where arbitrary probability distributions on $[d]$ with entropy distance bounded below by η, for each $i \in [N]$, the bound in (3.36) is valid for each $i \in [N]$.

Bibliography

[AB07] R. Ahlswede and V. Blinovsky. „Classical Capacity of Classical-Quantum Arbitrarily Varying Channels“. In: *Information Theory, IEEE Transactions on* **53**:2 (2007), pp. 526–533.

[Abe+09] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. „The mother of all protocols: restructuring quantum information's family tree“. In: *Proc. R. Soc. Lond. A* **465**: (2009), pp. 2537–2563.

[AC93] R. Ahlswede and I. Csiszar. „Common randomness in information theory and cryptography, Part I: Secret sharing“. In: *Information Theory, IEEE Transactions on* **39**:4 (1993), pp. 1121–1132.

[AF04] R. Alicki and M. Fannes. „Continuity of quantum conditional information“. In: *J. Phys. A.: Mathematical and General* **37**: (2004), L55–L57(3).

[Ahl+12] R. Ahlswede, I. Bjelaković, H. Boche, and J. Nötzel. „Quantum Capacity under Adversarial Quantum Noise: Arbitrarily Varying Quantum Channels“. English. In: *Communications in Mathematical Physics* (2012), pp. 1–54.

[Ahl78] R. Ahlswede. „Elimination of correlation in random codes for arbitrarily varying channels“. In: *Z. für Wahrscheinlichkeitstheorie und verw. Gebiete* **44**: (1978), pp. 159–175.

[Ahl80] R. Ahlswede. „Coloring hypergraphs: A new approach to multi-user source coding II“. In: *Journ. of Combinatorics, Information and System Sciences* **5**:3 (1980), pp. 220–268.

[Ahl86] R. Ahlswede. „Arbitrarily varying channels with states sequence known to the sender“. In: *Information Theory, IEEE Transactions on* **32**:5 (1986), pp. 621–629.

[AL70] H. Araki and E. H. Lieb. „Entropy Inequalities“. In: *Comm. Math. Phys.* **18**: (1970), pp. 160–170.

[AW02] R. Ahlswede and A. Winter. „Strong converse for identification via quantum channels“. In: *Information Theory, IEEE Transactions on* **48**:3 (Mar. 2002), pp. 569–579.

[BB09] I. Bjelaković and H. Boche. „Classical Capacities of Compound and Averaged quantum Channels“. In: *Information Theory, IEEE Transactions on* **55**: (2009), pp. 3360–3374.

[BBJ13] I. Bjelaković, H. Boche, and G. Janßen. „Universal quantum state merging“. In: *J. Math. Phys.* **54**:3 (2013), p. 032204.

[BBN08] I. Bjelaković, H. Boche, and J. Nötzel. „Quantum capacity of a class of compound channels“. In: *Phys. Rev. A* **78**:4 (Oct. 2008), p. 042331.

[BBN09] I. Bjelaković, H. Boche, and J. Nötzel. „Entanglement transmission and generation under channel uncertainty: Universal quantum channel coding“. In: *Comm. Math. Phys.* **292**: (Nov. 2009), pp. 55–97. eprint: `0811.4588`.

[Ben+02] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal. „Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem“. In: *Information Theory, IEEE Transactions on* **48**:10 (Oct. 2002), pp. 2637–2655.

[Ben+93] C. H. Bennett et al. „Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels“. In: *Phys. Rev. Lett.* **70**:13 (Mar. 1993), pp. 1895–1899.

[BJ14a] H. Boche and G. Janssen. „Resource Cost Results for Entanglement Distillation and State Merging under Source Uncertainties“. In: *Proceedings of the 2014 IEEE International Symposium on Information Theory*. 2014, pp. 721–725.

[BJ14b] H. Boche and G. Janssen. „Resource cost results for one-way entanglement distillation and state merging of compound and arbitrarily varying quantum sources“. In: *J. Math. Phys* **55**:8 (2014), p. 082208.

[Bje+05] I. Bjelakovic et al. „A quantum version of Sanov's theorem“. In: *Comm. Math. Phys.* **260**:3 (2005), pp. 659–671.

[BN13] H. Boche and J. Noetzel. „Arbitrarily small amounts of correlation for arbitrarily varying quantum channels“. In: *J. Math. Phys* **54**: (2013), p. 112202.

[Bor85] K. C. Border. *Fixed Point Theorems with Applications to Economics and Game Theory*. Cambridge University Press, 1985.

[Chi+14] E. Chitambar et al. „Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask)“. In: *Comm. Math. Phys.* **328**:1 (May 2014), pp. 303–326.

[CK11] I. Csiszár and J. Körner. *Information Theory - Coding Theorems for Discrete Memoryless Systems, 2nd ed.* Cambridge University Press, 2011.

[CM06] M. Christandl and G. Mitchinson. „The Spectra of Quantum States and the Kronecker Coefficients of the Symmetric Group“. In: *Comm. Math. Phys.* **261**: (2006), pp. 789–797.

[CN04] I. Csiszar and P. Narayan. „Secrecy capacities for multiple terminals“. In: *IEEE Transactions on Information Theory* **50**:12 (Dec. 2004), pp. 3047–3061.

[Dev05] I. Devetak. „The Private Classical Capacity and Quantum Capacity of a Quantum Channel“. In: *IEEE Transactions on Information Theory* **51**:1 (2005), pp. 44–55.

[DH76] W. Diffie and M. E. Hellman. „New Directions in Cryptography“. In: *IEEE Trans. Inf. Th.* **22**: (1976), pp. 644–654.

[DHW04] I. Devetak, A. W. Harrow, and A. Winter. „A family of quantum protocols“. In: *Phys. Rev. Lett.* **93,** (2004), p. 230504. eprint: `quant-ph/0308044`.

[DHW08] I. Devetak, A. W. Harrow, and A. Winter. „A Resource Framework for Quantum Shannon Theory“. In: *IEEE Trans. Inf. Th.* **54**:10 (2008), pp. 4587–4618. eprint: quant-ph/0512015.

[DL70] E. Davies and J. Lewis. „An operational approach to quantum probability“. In: *Comm. Math. Phys.* **17**: (1970), pp. 239–260.

[Dür+99] W. Dür, H.-J. Briegel, J. I. Cirac, and P. Zoller. „Quantum repeaters based on entanglement purification“. In: *Phys. Rev. A* **59**:1 (Jan. 1999), pp. 169–181.

[DW05] I. Devetak and A. Winter. „Distillation of secret key and entanglement from quantum states“. In: *Proc. R. Soc. Lond. A* **461**: (2005), pp. 207–235. eprint: quant-ph/0306078.

[Fan73] M. Fannes. „A Continuity Property of the Entropy Density for Spin Lattice Systems“. In: *Comm. Math. Phys.* **31**: (1973), pp. 291–294.

[Fek23] M. Fekete. „Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten“. In: *Mathematische Zeitschrift* **17**: (1923), pp. 228–249.

[GPW05] B. Groisman, S. Popescu, and A. Winter. „Quantum, classical, and total amount of correlations in a quantum state“. In: *Phys. Rev. A* **72**:3 (Sept. 2005), p. 032317.

[Hol73] A. S. Holevo. „Bounds for the quantity of information transmissted by a quantum communication channel“. In: *Probl. Inf. Trans.* **9**: (1973), pp. 177–183.

[HOW05] M. Horodecki, J. Oppenheim, and A. Winter. „Partial quantum Information“. In: *Nature* **436**: (2005), pp. 673–676.

[HOW07] M. Horodecki, J. Oppenheim, and A. Winter. „Quantum State Merging and Negative Information“. In: *Comm. Math. Phys.* **269**:1 (2007), pp. 107–136. eprint: arXiv:quant-ph/0512247.

[HZ12] T. Heinosaari and M. Ziman. *The Mathematical Language of Quantum Theory*. Cambridge University Press, 2012.

[Jan10] G. Janssen. „Universelles Quanten-State Merging“. Diploma Thesis (in german). Technical University of Berlin, 2010.

[Joz+98] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki. „Universal Quantum Information Compression“. In: *Phys. Rev. Lett.* **81**:8 (Aug. 1998), pp. 1714–1717.

[Joz94] R. Jozsa. „Fidelity for Mixed Quantum States“. In: *Journal of Modern Optics* **41**:12 (1994), pp. 2315–2323.

[Key02] M. Keyl. „Fundamentals of Quantum Information Theory“. In: *Phys. Rep.* **369,** (2002), no.5, 431–548. eprint: quant-ph/0202122.

[KW01] M. Keyl and R. F. Werner. „Estimating the spectrum of a density operator“. In: *Phys. Rev. A* **64**: (5 Oct. 2001), p. 052311.

[LS08] D. Leung and G. Smith. „Continuity of quantum channel capacities“. In: (Oct. 2008). eprint: 0810.4931.

[Mau93] U. M. Maurer. „Secret key agreement by public discussion from common information“. In: *IEEE Transactions on Information Theory* **39**:3 (May 1993), pp. 733–742.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[ON07] T. Ogawa and H. Nagaoka. „Making Good Codes for Classical-Quantum Channel Coding via Quantum Hypothesis Testing“. In: *Information Theory, IEEE Transactions on* **53**:6 (June 2007), pp. 2261–2266.

[San57] I. N. Sanov. „On the probability of large deviations of random variables“. In: *Mat. Sbornik* **42**: (1957), pp. 11–44.

[Sch95] B. Schumacher. „Quantum coding“. In: *Phys. Rev. A* **51**:4 (Apr. 1995), pp. 2738–2747.

[Sim96] B. Simon. *Representations of Finite and Compact Groups*. Ed. by L. E. S. James E. Humphreys. Graduate Studies in Mathematics. Vol. 10. American Mathematical Society, 1996.

[TBS16] N. Tavangaran, H. Boche, and R. F. Schaefer. „Secret-Key Generation Using Compound Sources and One-Way Public Communication“. In: *arxiv.org* (2016).

[Uhl76] A. Uhlmann. „The 'transition probability' in the state space of a *-algebra“. In: *Rep. Math. Phys.* **9**: (1976), pp. 273–279.

[Ver26] G. S. Vernam. „Cipher printing telegraphy systems for secret wire and ration telegraphic communications“. In: *J. Amer. Inst. Elec. Eng.* 5 (1926), pp. 109–115.

[Web94] R. Webster. *Convexity*. Oxford University Press, 1994.

[Wil13] M. M. Wilde. *Quantum Information Theory*. Camb, 2013.

[Win99] A. Winter. „Coding Theorem and Strong Converse for Quantum Channels“. In: *IEEE Transactions on Information Theory* **45**:7 (1999), pp. 2481–2485.

[WTS07] R. Wilson, D. Tse, and R. A. Scholz. „Channel identification: Secret sharing using reciprocity in ultrawideband channels“. In: *IEEE Trans. Inf. Forensics and Security* 2 (2007).

[YHD08] J. Yard, P. Hayden, and I. Devetak. „Capacity theorems for quantum multiple-access channels: classical-quantum and quantum-quantum capacity regions“. In: *Information Theory, IEEE Transactions on* **54**:7 (July 2008), pp. 3091–3113. eprint: `quant-ph/0501045`.