Diss. ETH No. 16242

Security of Quantum Key Distribution

A dissertation submitted to

SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Natural Sciences

presented by

Renato Renner
Dipl. Phys. ETH

born December 11, 1974, in Lucerne
citizen of Lucerne, LU, and Andermatt, UR, Switzerland

accepted on the recommendation of

Prof. Dr. Ueli Maurer, examiner
Dr. Charles H. Bennett, co-examiner

September 2005

to Sophie and Jill

Acknowledgments

The research leading to this thesis was carried out under the supervision of Ueli Maurer. I would like to thank him for his encouragement and support. His lucid way of thinking and his persistent quest for getting the “right view” had a major impact on my research. His door was always open to discuss both scientific and non-scientific problems.

I also owe many thanks to Charles Bennett for investing his time in studying my work and for being the co-examiner of this thesis.

During the last few years, I had a strong collaboration with Stefan Wolf. It was always a pleasure working with him, both in Zurich and during my visits to Montreal. He also introduced me to the research group of Nicolas Gisin at the University of Geneva and, with that, to the field of quantum cryptography.

I have undoubtedly benefited a lot from the stimulating discussions with my collaborators. In particular, I would like to thank Matthias Christandl, Robert König, and Barbara Kraus for their valuable and unique ideas. Moreover, I am grateful to all members of the cryptography research group at ETH Zurich for their support (even though they never got me to play table-top football). Particular thanks go to my officemate Thomas Holenstein who gave me much valuable advice on both research and TeX related questions.

I am also very grateful to Nicolas Gisin and the members of his research group for many interesting and enjoyable discussions on quantum cryptography. It was always a pleasure to visit them in Geneva.

This thesis has improved substantially by the valuable comments of those who read preliminary versions of it. In particular, I would like to thank Robert König and Christopher Portmann for their helpful suggestions and proofreading. (The remaining errors and omissions are entirely the author’s responsibility.)

Finally and most importantly, I would like to thank my family for their love and support. This thesis is dedicated to my wife Sophie and my daughter Jill.

This research was partially supported by the Swiss National Science Foundation, grant No. 2000-66716.01/1.

Abstract

Quantum information theory is an area of physics which studies both fundamental and applied issues in quantum mechanics from an information-theoretic viewpoint. The underlying techniques are, however, often restricted to the analysis of systems which satisfy a certain *independence condition*. For example, it is assumed that an experiment can be repeated independently many times or that a large physical system consists of many virtually independent parts. Unfortunately, such assumptions are not always justified. This is particularly the case for practical applications—e.g., in (quantum) cryptography—where parts of a system might have an arbitrary and unknown behavior.

We propose an approach which allows to study general physical systems for which the above mentioned independence condition does not necessarily hold. It is based on an extension of various information-theoretic notions. For example, we introduce new uncertainty measures, called *smooth min- and max-entropy*, which are generalizations of the von Neumann entropy. Furthermore, we develop a quantum version of de Finetti's representation theorem, as described below.

Consider a physical system consisting of n parts. These might, for instance, be the outcomes of n runs of a physical experiment. Moreover, assume that the joint state of this n-partite system can be extended to an $(n+k)$-partite state which is symmetric under permutations of its parts (for some $k \gg 1$). The *de Finetti representation theorem* then says that the original n-partite state is, in a certain sense, close to a mixture of product states. Independence thus follows (approximatively) from a symmetry condition. This symmetry condition can easily be met in many natural situations. For example, it holds for the joint state of n parts which are chosen at random from an arbitrary $(n+k)$-partite system.

As an application of these techniques, we prove the security of *quantum key distribution (QKD)*, i.e., secret key agreement by communication over a quantum channel. In particular, we show that, in order to analyze QKD protocols, it is generally sufficient to consider so-called *collective attacks*, where the adversary is restricted to applying the same operation to each particle sent over the quantum channel separately. The proof is generic and thus applies to known protocols such as *BB84* and *B92* (where better bounds

on the secret-key rate and on the the maximum tolerated noise level of the quantum channel are obtained) as well as to *continuous variable* schemes (where no full security proof has been known). Furthermore, the security holds with respect to a strong so-called *universally composable* definition. This implies that the keys generated by a QKD protocol can safely be used in any application, e.g., for *one-time pad* encryption—which, remarkably, is not the case for most of the standard definitions.

Zusammenfassung

Quanteninformationstheorie ist ein Gebiet der Physik, das sich sowohl mit fundamentalen als auch angewandten Fragen innerhalb der Quantenmechanik beschäftigt und diese aus einem informationstheoretischen Gesichtspunkt betrachtet. Die dabei verwendeten Techniken sind jedoch oft darauf beschränkt, Systeme zu analyzieren, welche eine gewisse *Unabhängigkeitsbedingung* erfüllen. Beispielsweise wird angenommen, dass ein Experiment viele Male unabhängig wiederholt werden kann, oder dass ein grosses physikalisches System aus vielen nahezu unabhängigen Teilen besteht. Leider sind solche Annahmen nicht immer gerechtfertigt. Dies gilt insbesondere für praktische Anwendungen wie z.B. innerhalb der (Quanten)Kryptographie, wo Teile eines Systems ein beliebiges und unbekanntes Verhalten aufweisen können.

Wir stellen einen Ansatz vor, welcher es erlaubt, allgemeine physikalische Systeme zu studieren, für die keine solche Unabhängigkeitsbedingung gilt. Er basiert auf einer Erweiterung verschiedener informationstheoretischer Konzepte. Zum Beispiel führen wir neue Entropiemasse ein, genannt *Smooth Min-Entropy* und *Smooth Max-Entropy*, welche die von Neumann-Entropie verallgemeinern. Zudem entwickeln wir eine quantenmechanische Version des Darstellungssatzes von de Finetti, die wir im folgenden beschreiben.

Wir betrachten ein physikalisches System, welches aus n Teilen besteht. Diese könnten beispielsweise durch n Wiederholungen eines physikalischen Experiments entstanden sein. Weiter nehmen wir an, dass der Gesamtzustand dieses n-teiligen Systems zu einem $(n + k)$-teiligen Zustand erweitert werden kann, welcher symmetrisch ist unter Vertauschungen der Teilsysteme (für $k \gg 1$). Der *Darstellungssatz von de Finetti* besagt dann, dass der Zustand des ursprünglichen n-teiligen Systems in einem gewissen Sinn nahe an einer Mischung von Produktzuständen ist. Die Unabhängigkeit der Teilsysteme folgt also näherungsweise aus einer Symmetriebedingung. Diese ist in vielen natürlichen Situationen einfach zu erfüllen. So gilt sie etwa für ein System von n Teilen welche zufällig aus einem $(n + k)$-teiligen System ausgewählt worden sind.

Als Anwendung dieser Techniken beweisen wir die Sicherheit von *Quantum Key Distribution (QKD)*, d.h., der Schlüsselverteilung über Quanten-

kanäle. Insbesondere zeigen wir, dass es zur Analyse von QKD-Protokollen im allgemeinen genügt, sogenannte *kollektive* Attacken zu betrachten, bei denen der Gegner darauf beschränkt ist, jedes über den Quantenkanal gesendete Teilchen gleich zu behandeln. Der Beweis ist generisch und daher sowohl auf bekannte Protokolle wie *BB84* und *B92* (für welche bessere Grenzen an die Schlüsselrate und den maximal tolerierten Geräuschpegel des Quantenkanals folgen) als auch auf *Continuous-Variable*-Protokolle (für welche kein vollständiger Sicherheitsbeweis bekannt war) anwendbar. Dabei ist Sicherheit gemäss einer sogenannten *universally composable* Definition garantiert. Das bedeutet, dass die durch ein QKD-Protokoll erzeugten Schlüssel in jeder denkbaren Anwendung verwendet werden dürfen, so z.B. für *One-Time-Pad*-Verschlüsselung — was bemerkenswerterweise für die meisten gebräuchlichen Definitionen nicht zutrifft.

Contents

Chapter 1

Introduction

1.1 Motivation

What is needed to establish a secret key between two spatially separated parties? Clearly, this question is of immediate interest for practical cryptographic applications such as secure message transmission.[1] More importantly, however, it is related to fundamental problems in (classical and quantum) information theory. Is information physical? Is classical information distinct from quantum information? In fact, it turns out that the possibility of secret key agreement (over insecure channels) strongly depends on the physical properties of information and that there is indeed a fundamental difference between classical and quantum information.

In this thesis, we address several basic question of quantum information theory: What does secrecy mean in a quantum world? (Chapter 2) How can knowledge and uncertainty be quantified? (Chapter 3) What is the role of symmetry? (Chapter 4) Can any type of randomness be transformed into uniform randomness? (Chapter 5) As we shall see, the answers to these questions allow us to treat the problem of secret key agreement in a very natural way (Chapters 6 and 7).

1.2 Quantum key distribution: general facts

Cryptographic setting

We consider a setting where two distant parties, traditionally called *Alice* and *Bob*, want to establish a common *secret key*, i.e., a string of random bits which is unknown to an adversary, *Eve*. Throughout this thesis, we focus on *information-theoretic security*, which is actually the strongest rea-

[1] For example, using one-time pad encryption [Ver26], the problem of secretly exchanging ℓ message bits reduces to the problem of distributing a secret key consisting of ℓ bits.

sonable notion of security.[2] It guarantees that an adversary does not get any information correlated to the key, except with negligible probability.

For the following, we assume that Alice and Bob already have at hand some means to exchange classical messages in an *authentic* way.[3] In fact, only relatively weak resources are needed to turn a completely insecure communication channel into an authentic channel. For example, Alice and Bob might invoke an authentication protocol (see, e.g., [Sti91, GN93]) for which they need a short[4] initial key. Actually, as shown in [RW03, RW04], it is even sufficient for Alice and Bob to start with only weakly correlated and partially secret information (instead of a short secret key).

Key agreement by quantum communication

Under the sole assumption that Alice and Bob are connected by a classical authentic communication channel, secret communication—and thus also the generation of a secret key—is impossible [Sha49, Mau93]. This changes dramatically when quantum mechanics comes into the game. Bennett and Brassard [BB84] (see also [Wie83]) were the first to propose a *quantum key distribution (QKD)* scheme which uses communication over a (completely insecure) quantum channel (in addition to the classical authentic channel). The scheme is commonly known as the *BB84 protocol*.

Quantum key distribution is generally based on the impossibility to observe a quantum mechanical system without changing its state. An adversary trying to wiretap the quantum communication between Alice and Bob would thus inevitably leave traces which can be detected. A quantum key distribution protocol thus achieves the following type of security: As long as the adversary is passive, it generates an (arbitrarily long) secret key. On the other hand, if the adversary tampers with the quantum channel, the protocol recognizes the attack and aborts the computation of the key.[5] (Note that this is actually the best one can hope for: As the quantum channel is completely insecure, an adversary might always interrupt the quantum communication between Alice and Bob, in which case it is impossible to generate a secret key.)

[2] An example of a weaker level of security is *computational security*, where one only requires that it is *difficult* (i.e., time-consuming, but not impossible) for an adversary to compute information on the key.

[3] *Authentic* means that, upon receiving a message, Bob can verify whether the message was indeed sent by Alice, and vice-versa.

[4] The length of the key only grows logarithmically in the length of the message to be authenticated.

[5] More precisely, it is guaranteed that the protocol does not abort as long as the adversary is passive (this is called *robustness*). Moreover, for any attack on the quantum channel, the probability that the protocol does not abort *and* the adversary gets information on the generated key is negligible (see Section 6.1.3 for details).

An example: the BB84 protocol

To illustrate the main principle of quantum key distribution, let us have a closer look at the *BB84 protocol*. It uses an encoding of classical bits in *qubits*, i.e., two-level quantum systems[6]. The encoding is with respect to one of two different orthogonal bases, called the *rectilinear* and the *diagonal basis*.[7] These two bases are *mutually unbiased*, that is, a measurement in one of the bases reveals no information on a bit encoded with respect to the other basis.

In the first step of the protocol, Alice chooses N random bits $X_1, \ldots, X_N$, encodes each of these bits into qubits using at random[8] either the rectilinear or the diagonal basis, and transmits them to Bob (using the quantum channel). Bob measures each of the qubits he receives with respect to—a random choice of—either the rectilinear or the diagonal basis to obtain classical bits Y_i. The pair of classical bitstrings $X = (X_1, \ldots, X_N)$ and $Y = (Y_1, \ldots, Y_N)$ held by Alice and Bob after this step is called the *raw key* pair.

The remaining part of the protocol is purely classical (in particular, Alice and Bob only communicate classically). First, Alice and Bob apply a *sifting* step, where they announce their choices of bases used for the encoding and the measurement, respectively. They discard all bits of their raw key for which the encoding and measurement bases are not compatible. Then Alice and Bob proceed with a *parameter estimation* step. They compare some (small) randomly chosen set of bits of their raw key in order to get a guess for the *error rate*, i.e., the fraction of positions i in which X_i and Y_i disagree. If the error rate is too large—which might indicate the presence of an adversary—Alice and Bob abort the protocol.

Let X' and Y' be the remaining parts of the raw keys (i.e., the bits of X and Y that have neither been discarded in the sifting step nor used for parameter estimation). These strings are now used for the actual computation of the final key. In an *information reconciliation* step, Alice sends certain error correcting information on X' to Bob.[9] This, together with Y', allows him to compute a guess for X'. (Note that, because of the parameter estimation step, it is guaranteed that X' and Y' only differ in a limited number of positions.) In the final step of the protocol, called *privacy amplification*, Alice and Bob use *two-universal hashing*[10] to turn the (generally only partially secret) string X' into a shorter but secure key.

---

[6] For example, the classical bits might be encoded into the spin orientation of particles.

[7] See Section 7.2.1 for a definition.

[8] In the original proposal of the BB84 protocol, Alice and Bob choose the two bases with equal probabilities. However, as pointed out in [LCA05], the efficiency of the protocol is increased if they select one of the two bases with probability almost one. In this case, the choices of Alice and Bob will coincide with high probability, which means that the number of bits to be discarded in the sifting step is small.

[9] The information reconciliation step might also be interactive.

[10] See Section 5.4 for a definition of two-universality.

The security of the BB84 protocol is based on the fact that an adversary, ignorant of the actual encoding bases used by Alice, cannot gain information about the encoded bits without disturbing the qubits sent over the quantum channel. If the disturbance is too large, Alice and Bob will observe a high error rate and abort the protocol in the parameter estimation step. On the other hand, if the disturbance is below a certain threshold, then the strings X' and Y' held by Alice and Bob are sufficiently correlated and secret in order to distill a secret key.

In order to prove security, one thus needs to quantify the amount of information that an adversary has on the raw key, given the disturbance measured by Alice and Bob. It is a main goal of this thesis to develop the information-theoretic techniques which are needed for this analysis. (See also Section 1.6.3 for a sketch of the security proof.)

Alternative protocols

Since the invention of quantum cryptography, a considerable effort has been taken to get a better understanding of its theoretical foundations as well as to make it more practical. In the course of this research, a large variety of alternative QKD protocols has been proposed. Some of them are very efficient with respect to the *secret-key rate*, i.e., the number of key bits generated per channel use [Bru98, BPG99]. Others are designed to cope with high channel noise or noise in the detector, which makes them more suitable for practical implementations [SARG04].

The structure of these protocols is mostly very similar to the BB84 protocol described above. For example, the *six-state protocol* proposed in [Bru98, BPG99] uses *three* different bases for the encoding (i.e., *six* different states), but otherwise is identical to the BB84 protocol. On the other hand, the *B92 protocol* [Ben92] is based on an encoding with respect to only *two* non-orthogonal states.

QKD over noisy channels

Any realistic quantum channel is subject to intrinsic noise. Alice and Bob will thus observe errors even if the adversary is passive. However, as these errors are not distinguishable from errors caused by an attack, the distribution of a secret key can only be successful if the noise level of the channel is sufficiently low.

As an example, consider the BB84 protocol described above. In the parameter estimation step, Alice and Bob compute a guess for the error rate and abort the protocol if it exceeds a certain threshold. Hence, the scheme only generates a key if the noise level of the channel is below this threshold.

The amount of noise tolerated by a QKD scheme is an important measure for its practicability. In fact, in an implementation, the level of noise inevitably depends on the distance between Alice and Bob (i.e., the length of the optical fiber, for an implementation based on photons). To characterize the efficiency of QKD schemes, one thus often considers the relation between the channel noise and the secret-key rate (see plots in Chapter 7). Typically, the secret-key rate decreases with increasing noise level and becomes zero as soon as the noise reaches a certain bound, called the *maximum tolerated channel noise*.

Quantum key distribution and distillation

Assume that Alice and Bob have access to some correlated quantum systems (e.g., predistributed pairs of entangled particles). A *quantum key distillation* protocol allows them to transform this correlation into a common secret key, while using only classical authentic communication.

As explained below, a quantum key *distribution* (QKD) protocol can generally be transformed into a key *distillation* protocol in such a way that security of the latter implies security of the first. This is very convenient for security proofs, as key distillation only involves quantum states (instead of quantum channels) which are easier to analyze (see [Eke91, BBM92]).

The connection between key distillation and key distribution protocols is based on the following observation: Let X be a classical value chosen according to a distribution P_X and let $|\phi^x\rangle$ be a quantum encoding of X. This situation could now equivalently be obtained by the following two-step process: (i) prepare a bipartite quantum state $|\Psi\rangle := \sum_x \sqrt{P_X(x)}|x\rangle \otimes |\phi^x\rangle$, where $\{|x\rangle\}_x$ is some orthonormal basis of the first subsystem; (ii) measure the first part of $|\Psi\rangle$ with respect to the basis $\{|x\rangle\}_x$. In fact, it is easy to verify that the outcome X is distributed according to P_X and that the remaining quantum system contains the correct encoding of X.

To illustrate how this observation applies to QKD, consider a protocol where Alice uses the quantum channel to transmit an encoding $|\phi^x\rangle$ of some randomly chosen value X to Bob (as, e.g., in the first step of the BB84 protocol described above). According to the above discussion, this can equivalently be achieved as follows:[11] First, Alice locally prepares the bipartite state $|\Psi\rangle$ defined above, keeps the first half of it, and sends the second half over the quantum channel to Bob. Second, Alice measures the quantum system she kept to get the classical value X. (Such a protocol is sometimes called an *entanglement-based* scheme.)

Note that, after the use of the quantum channel—but before the measurement—Alice and Bob share some (generally entangled) quantum state.

[11] More generally, any arbitrary protocol step can be replaced by a coherent quantum operation followed by some measurement.

The remaining part of the key distribution protocol is thus actually a quantum key distillation protocol. Hence, if this key distillation protocol is secure (for any predistributed entanglement) then the original quantum key distribution protocol is secure (for any arbitrary attack of Eve).

1.3 Contributions

This thesis makes two different types of contributions. First, we introduce various concepts and prove results which are of general interest in quantum information theory and cryptography.[12] These contributions are summarized in Section 1.3.1 below. Second, we apply our techniques to QKD in order to derive a general security criterion. Some aspects and implications of this result are discussed in Section 1.3.2.

1.3.1 New notions in quantum information theory

Smooth min- and max-entropies as generalizations of von Neumann entropy

The von Neumann entropy, as a measure for the uncertainty on the state of a quantum system, plays an important role in quantum information theory. This is mainly due to the fact that it characterizes fundamental information-theoretic tasks such as *randomness extraction* or *data compression*. For example, the von Neumann entropy of a source emitting quantum states can be interpreted as the minimum space needed to encode these states such that they can later be reconstructed with arbitrarily small error. However, any such interpretation of the von Neumann entropy only holds asymptotically in situations where a certain underlying experiment is repeated many times independently. For the above example, this means that the encoding is over many (sufficiently independent) outputs of the source.

In the context of cryptography, where an adversary might corrupt parts of a system in an arbitrary way, this independence can often not be guaranteed. The von Neumann entropy is thus usually not an appropriate measure—e.g., to quantify the uncertainty of an adversary—unless we put some severe restrictions on her capabilities (e.g., that her attack consists of many independent repetitions of the same action).

In this thesis, we introduce two entropy measures, called *smooth min-* and *max-entropy*, which can be seen as generalizations of the von Neumann entropy. While smooth min-entropy quantifies the amount of uniform

[12] For example, our result on privacy amplification against quantum adversaries is not only useful to prove the security of QKD. It has also found interesting applications within other fields of cryptography, as for instance in the context of multi-party computation (see, e.g., [DFSS05] for a result on bit commitment).

randomness that can be *extracted* from a quantum system, the smooth max-entropy corresponds to the length of an optimal *encoding* of the system's state. Unlike the von Neumann entropy, however, this characterization applies to arbitrary situations—including those for which there is no underlying independently repeated experiment.

In the special case of many *independent repetitions* (that is, if the system's state is described by a density operator which has product form), smooth min- and max-entropy both reduce to the von Neumann entropy, as expected. Moreover, smooth min- and max-entropy inherit most of the properties known from the von Neumann entropy, as for example the strong subadditivity. (We refer to Section 1.5 for a summary of these results.) On the other hand, because the von Neumann entropy is a special case of smooth min- and max-entropy, its properties follow directly from the corresponding properties of the smooth min- or max-entropy. Interestingly, some of the proofs are surprisingly easy in this general case. For example, the strong subadditivity of the smooth min-entropy follows by a very short argument (cf. Lemma 3.1.7 and Lemma 3.2.7). Note that this immediately gives a simple proof for the strong subadditivity of the von Neumann entropy.

De Finetti representation theorem for finite symmetric quantum states

An n-partite density operator ρ_n is said to be N*-exchangeable*, for $N \geq n$, if it is the partial state (i.e., $\rho_n = \mathrm{tr}_k(\rho_N)$) of an N-partite density operator ρ_N which is invariant under permutations of the subsystems. Moreover, ρ_n is *infinitely-exchangeable* if it is N-exchangeable for all $N \geq n$. The *quantum de Finetti representation theorem* [HM76] (which is the quantum version of a theorem in probability theory named after its inventor Bruno de Finetti[13]) makes a fundamental statement on such symmetric operators.[14] Namely, it says that any infinitely-exchangeable operator ρ_n can be written as a convex combination (i.e., a *mixture*) of product operators,

$$\rho_n = \int_\sigma \sigma^{\otimes n} \nu(\sigma) \ .$$

We generalize the quantum de Finetti representation theorem for *infinitely* exchangeable operators to the *finite* case.[15] More precisely, we show that the above formula still holds approximatively if ρ_n is only N-exchangeable for, some finite N which is sufficiently larger than n. (We refer to Section 1.5 below for a more detailed description of this statement.)

[13] See [MC93] for a collection of de Finetti's original papers.

[14] See [CFS02] for a nice proof of the quantum de Finetti theorem based on its classical analogue.

[15] The result presented in this thesis is different from the one proposed in a previous paper [KR05] (see Section 1.5 for more details).

The de Finetti representation theorem turns out to be a useful tool in quantum information theory. In fact, symmetric (and exchangeable) states play an important role in many applications. For example, the operator describing the joint state of n particles selected at random from a set of N particles is N-exchangeable. Hence, according to our finite version of the de Finetti representation theorem, the analysis of such states can be reduced to the analysis of product states—which is often much easier than the general case. Following this idea, we will use the finite de Finetti representation theorem to argue that, for proving the security of a QKD scheme against arbitrary attacks, it suffices to consider attacks that have a certain product structure (so-called *collective attacks*, cf. Section 1.3.2).

Universal security of keys in a quantum world

In quantum cryptography, the security of a secret key S is typically defined with respect to the classical information W that an adversary might obtain when measuring her quantum system $\mathcal{H}_E$. More precisely, S is said to be secure if, for any measurement of the adversary's system $\mathcal{H}_E$, the resulting outcome W gives virtually no information on S. Although this definition looks quite strong, we shall see that it is not sufficient for many applications, e.g., if the key S is used for one-time pad encryption (see Section 2.2).

We propose a security definition which overcomes this problem. Roughly speaking, we say that a key S is *ε-secure* if, except with probability ε, S is equal to a *perfect key* which is uniformly distributed and completely independent of the adversary's quantum system. In particular, our security definition is *universal* in the sense that an ε-secure key can safely be used in any application, except with probability ε.[16]

Security of privacy amplification against quantum adversaries

Let X be a classical random variable on which an adversary has some partial information. *Privacy amplification* is the art of transforming this partially secure X into a fully secure key S, and has been studied extensively for the case where the adversary's information is purely classical. It has been shown [BBR88, ILL89, BBCM95] that it is always possible to generate an ℓ-bit key S which is secure against any adversary whose uncertainty on X—measured in terms of the *collision entropy*[17]—is sufficiently larger than ℓ.

We generalize this classical privacy amplification theorem to include *quantum* adversaries who might hold information on X encoded in the state

[16] Hence, our security definition fits into general frameworks concerned with the universal security of quantum protocols, as proposed by Ben-Or and Mayers [BOM04] and Unruh [Unr04] (see Section 2.2 for more details).

[17] The *collision entropy*, also called *Rényi entropy of order two*, of a probability distribution P_X is the negative binary logarithm of its collision probability $\sum_x P_X(x)^2$.

of a quantum system. We show that, similar to the classical result, X can be transformed into a key of length ℓ which is secure[18] if the uncertainty of the adversary on X—this time measured in terms of the *smooth min-entropy*—is at least roughly ℓ. Because the smooth min-entropy is generally larger than the collision entropy, this also implies the above classical result.

Our privacy amplification theorem is optimal with respect to the maximum length ℓ of the extractable secret key—i.e., smooth min-entropy completely characterizes the number of secret key bits that can be generated from a partially secret string (up to some small constant). This also improves our previous results [KMR05, RK05] which are only optimal in certain special cases.[19]

1.3.2 Properties and implications of the security result

We provide a simple and general[20] criterion for the security of QKD against *any* attack allowed by the laws of quantum physics. The following is a summary of the most important properties and consequences of this result. (For a more detailed description of the security criterion and a proof sketch, we refer to Section 1.6 below.)

Coherent attacks are not stronger than collective attacks

An adversary might in principle apply an arbitrary operation on the quantum states exchanged between Alice and Bob. In the case of the most general, so-called *coherent attacks*, this operation could involve all subsystems (particles) simultaneously, which makes it (seemingly) difficult to analyze. One thus often considers a restricted class of attacks, called *collective attacks* [BM97b, BM97a], where the adversary is assumed to apply the same transformation to each of the subsystems that is sent over the channel.[21] A natural and long-standing open question in this context is whether security against collective attacks implies full security (see, e.g., [BBB$^+$02]). Our

[18] We prove security according to the strong definition proposed in Section 2.2—i.e., the security is *universal*.

[19] The result proven in [RK05] is optimal if the density operator describing the initial string together with the adversary's quantum information has product form.

[20] The security criterion is *general* in the sense that it applies to virtually all known protocols. Note that this stands in contrast to previous security proofs, which are mostly designed for specific protocols.

[21] An even more restricted type of attacks are the so-called *individual attacks* where, additionally, the adversary is supposed to apply some fixed measurement operation to each of the subsystems sent through the channel. In particular, this measurement cannot depend on the classical information that Alice and Bob exchange for error correction and privacy amplification. As shown in [BMS96], such individual attacks are generally weaker than collective attacks. Hence, security against individual attacks does not imply full security.

result immediately answers this question in the positive, that is, coherent attacks cannot be more powerful than collective attacks.[22]

Security of practical implementations

Because of technical limitations, practical implementations of QKD are subject to many imperfections. In addition to noisy channels, these might include faulty sources[23] or detector losses. Because of its generality, our security criterion can be used for the analysis of such practical settings.[24]

Keys generated by QKD can safely be used in applications

The security result holds with respect to a so-called *universal* security definition. This guarantees that the key generated by a QKD protocol can safely be used in applications such as for one-time pad encryption. (As mentioned above, this is not necessarily the case for many of the standard security definitions.)

Improved bounds on the efficiency of concrete protocols

Our security result applies to protocols which could not be analyzed with previously known techniques (e.g., a reduction to entanglement purification schemes, as proposed in [SP00]). In particular, it allows to compute the key rates for new variants of known protocols.[25] For example, we propose an improved version of the six-state protocol and show that it is more efficient than previous variants. Moreover, we derive new bounds on the maximum tolerated channel noise of the BB84 or the six-state protocol with one-way post-processing.

Explicit bounds on the security of finite keys

The security criterion gives explicit (non-asymptotic) bounds on the secrecy and the length of keys generated from any (finite) number of invocations of the quantum channel. Moreover, it applies to schemes which use arbitrary (not necessarily optimal) subprotocols for information reconciliation. This is in contrast to most known security results which—with a few exceptions[26]—only hold asymptotically for large key sizes and for asymptotically

[22] This statement holds for virtually any QKD protocol; the only requirement is that the protocol is symmetric under permutations of the channel uses (see Section 1.6 for more details).

[23] For example, it is difficult to design sources that emit perfect single-photon pulses.

[24] As there is no restriction on the structure of the underlying Hilbert space, the security criterion applies to any modeling of the physical system which is used for the quantum communication between Alice and Bob.

[25] E.g., we will analyze protocols that use an alternative method for the processing of the raw key.

[26] See, e.g., [ILM01] for a nice and very careful explicit analysis of the BB84 protocol.

optimal information reconciliation.

1.4 Related work

The techniques developed in this thesis are partly motivated by ideas known from classical information theory and, in particular, cryptography (e.g., classical de Finetti-style theorems, privacy amplification against classical adversaries, or universally composable security). For a discussion of these notions and their relation to our results we refer to Section 1.3. In the following, we rather focus on work related to the security of QKD.

Since Bennett and Brassard proposed the first QKD protocol in 1984 [BB84], it took more than a decade until Mayers [May96] proved that the scheme is secure against arbitrary attacks.[27] This result was followed by various alternative proofs (see, e.g., [CRE04] or [LCA05] for an overview).

One of the most popular proof techniques was proposed by Shor and Preskill [SP00], based on ideas of Lo and Chau [LC99]. It uses a connection between key distribution and entanglement purification [BBP+96] pointed out by Ekert [Eke91] (see also [BBM92]). The proof technique of Shor and Preskill was later refined and applied to other protocols (see, e.g., [GL03, TKI03]).

In [CRE04], we have presented a general method for proving the security of QKD which does not rely on entanglement purification. Instead, it is based on a result on the security of privacy amplification in the context of quantum adversaries [KMR05, RK05]. Later, this method has been extended and applied to prove the security of new variants of the BB84 and the six-state protocol [RGK05, KGR05].[28] The security proof given in this thesis is based on ideas developed in these papers.

Our new approach for proving the security of QKD has already found various applications. For example, it is used for the analysis of protocols based on continuous systems as well as to improve the analysis of known (practical) protocols exploiting the fact that an adversary cannot control the noise in the physical devices owned by Alice and Bob (see, e.g., [Gro05, NA05, Lo05]).

1.5 Outline of the thesis

The following is a brief summary of the main results obtained in each chapter.

[27] See also [May01] for an improved version of Mayers' proof.

[28] In [RGK05, KGR05] we use an alternative technique (different from the quantum de Finetti theorem) to show that collective attacks are equivalent to coherent attacks for certain QKD protocols.

Chapter 2: Preliminaries

The first part of this chapter (Section 2.1) is concerned with the representation of physical (cryptographic) systems as mathematical objects. We briefly review the density operator formalism which is used to describe quantum mechanical systems. Moreover, we present some variant of this formalism which is useful when dealing with physical systems that consist of both classical and quantum parts.

The second part of Chapter 2 (Section 2.2) is devoted to the security definition for secret keys. We first argue that many of the widely used definitions are problematic—in the sense that they do not imply the security of applications such as one-time pad encryption. Then, as a solution to this problem, we introduce a so-called *universal* security definition for secret keys and discuss its properties.

Chapter 3: Smooth min- and max-entropy

This chapter introduces and studies *smooth min-entropy* $H^{\varepsilon}_{\min}$ and *smooth max-entropy* $H^{\varepsilon}_{\max}$, which both are entropy measures for density operators. We first discuss some basic properties (Sections 3.1 and 3.2) which are actually very similar to those of the von Neumann entropy (Theorem 3.2.12). For example, the smooth min-entropy is *strongly subadditive*, that is,[29]

$$H^{\varepsilon}_{\min}(A|BC) \leq H^{\varepsilon}_{\min}(A|B) \ , \tag{1.1}$$

and it obeys an inequality which can be interpreted as a *chain rule*,

$$H^{\varepsilon}_{\min}(AB|C) \leq H^{\varepsilon}_{\min}(A|BC) + H_{\max}(B) \ . \tag{1.2}$$

Moreover, if the states in the subsystems $\mathcal{H}_A$ and $\mathcal{H}_C$ are independent conditioned on a classical value Y then

$$H^{\varepsilon}_{\min}(AY|C) \geq H^{\varepsilon}_{\min}(Y|C) + H_{\min}(A|Y) \ . \tag{1.3}$$

The second part of Chapter 3 (Section 3.3) treats the special case where the density operators have product form. In this case, smooth min- and max-entropy both reduce to the von Neumann entropy. Formally, the smooth min-entropy $H^{\varepsilon}_{\min}(A^n|B^n)$ of a product state $\rho_{A^nB^n} = \sigma_{AB}^{\otimes n}$ satisfies

$$\lim_{n\to\infty} \frac{1}{n} H^{\varepsilon}_{\min}(A^n|B^n) = H(A|B) \ , \tag{1.4}$$

where $H(A|B) = H(\sigma_{AB}) - H(\sigma_B)$ is the (conditional) von Neumann entropy evaluated for the operator σ_{AB} (cf. Theorem 3.3.6 and Corollary 3.3.7).

[29] We use a slightly simplified notation in this summary. For example, we write $H^{\varepsilon}_{\min}(A|B)$ to denote the smooth min-entropy of a state ρ_{AB} given the second subsystem (instead of $H^{\varepsilon}_{\min}(\rho_{AB}|B)$ which is used in the technical part).

Chapter 4: Symmetric states

This chapter is concerned with *symmetric* states, that is, states on n-fold product system $\mathcal{H}^{\otimes n}$ that are invariant under permutations of the subsystems. We first show that any permutation-invariant density operator has a symmetric purification, which allows us to restrict our attention to the analysis of *pure* symmetric states (Section 4.2).

The main result of this section is a finite version of the *quantum de Finetti representation theorem* (Section 4.3). It says that symmetric states can be approximated by a convex combination of states which have "almost" product form (cf. Theorem 4.3.2). Formally, if ρ_{n+k} is a permutation-invariant operator on $N = n + k$ subsystems $\mathcal{H}$, then the partial state ρ_n on $\mathcal{H}^{\otimes n}$ (obtained by tracing over k subsystems) is approximated by a mixture of operators ρ_n^σ, i.e.,

$$\rho_n \approx \int_\sigma \rho_n^\sigma \nu(\sigma) \ , \tag{1.5}$$

where the integral ranges over all density operators σ on *one single* subsystem $\mathcal{H}$ and ν is some probability measure on these operators. Roughly speaking, the states ρ_n^σ are superpositions of states which, on at least $n - r$ subsystems, for some small r, have product form $\sigma^{\otimes n-r}$. Moreover, the distance[30] between the left and the right hand side of the approximation (1.5) decreases exponentially fast in r and k.[31]

The properties of the states ρ_n^σ occurring in the convex combination (1.5) are similar to those of perfect product states $\sigma^{\otimes n}$. The main result of Section 4.4 can be seen as a generalization of (1.4). It states that, for a state $\rho_{A^n B^n}^{\sigma_{AB}}$ which has almost product form $\sigma_{AB}^{\otimes n}$ (in the sense defined above, where σ_{AB} is a bipartite operator on $\mathcal{H}_A \otimes \mathcal{H}_B$) the smooth min-entropy is given by

$$\lim_{n\to\infty} \frac{1}{n} H_{\min}^\varepsilon(A^n|B^n) = H(A|B) \tag{1.6}$$

(see Theorem 4.4.1[32]).

Analogously, in Section 4.5, we show that states ρ_n^σ which have almost product form $\sigma^{\otimes n}$ lead to similar statistics as perfect product states $\sigma^{\otimes n}$ if they are measured with respect to a product measurement. Formally, let P_Z be the distribution of the outcomes when measuring σ with respect to

[30] The distance is measured with respect to the L_1-distance, as defined in Section 2.1.4.

[31] Note that this version of the finite quantum de Finetti representation theorem—although the same in spirit—is distinct from the the one proposed in [KR05]: In [KR05], the decomposition is with respect to *perfect* n-fold product states $\sigma^{\otimes n}$—instead of states ρ_n^σ which are products on only $n - r$ subsystems—but the approximation is not exponential.

[32] Note that Theorem 4.4.1 only implies one direction ($\geq$). The other direction ($\leq$) follows from a similar argument for the smooth max-entropy, which is an upper bound on the smooth min-entropy.

a POVM $\mathcal{M}$. Moreover, let $\lambda_{\mathbf{z}}$ be the statistics (i.e., the frequency distribution) of the outcomes $\mathbf{z} = (z_1, \ldots, z_n)$ of the product measurement $\mathcal{M}^{\otimes n}$ applied to ρ_n^σ. Then

$$\lim_{n \to \infty} \lambda_{\mathbf{z}} = P_Z \tag{1.7}$$

(cf. Theorem 4.5.2).

Chapter 5: Privacy amplification

This chapter is on privacy amplification in the context of quantum adversaries. The main result is an explicit expression for the secrecy of a key S which is computed from an only partially secure string X by *two-universal hashing*[33] (Theorem 5.5.1 and Corollary 5.6.1). The result implies that the key S is secure under the sole condition that its length ℓ is bounded by

$$\ell \lessapprox H_{\min}^{\varepsilon}(X|E) \tag{1.8}$$

where $H_{\min}^{\varepsilon}(X|E)$ denotes the smooth min-entropy of X given the adversary's initial information.

Chapter 6: Security of QKD

This chapter is devoted to the statement and proof of our main result on the security of QKD. In particular, it contains an expression for the key rate for a general class of protocols in terms of simple entropic quantities (Theorem 6.5.1 and Corollary 6.5.2). (We refer to Section 1.6 for an overview on this result and its proof.)

Chapter 7: Examples

As an illustration, we apply the general result of Chapter 6 to specific types of QKD protocols. The focus is on schemes which are based on two-level systems. In particular, we analyze different versions of the six-state QKD protocol and compute explicit values for their rates (see Plots 7.1–7.5).

1.6 Outline of the security analysis of QKD

The following is a summary of our main result on the security of quantum key distillation which—according to the discussion in Section 1.2—also implies the security of quantum key distribution. Moreover, we give a sketch of the security proof, which is based on the technical results summarized in Section 1.5 above. (For a complete description of the security result and the full proof, we refer to Chapter 6.)

[33] That is, S is the output $f(X)$ of a function f which is randomly chosen from a so-called *two-universal* family of hash functions (see Section 5.4 for a definition).

1.6.1 Protocol

We start with a brief characterization of the general type of quantum key distillation protocols to which our security proof applies. For this, we assume that Alice and Bob start with N bipartite quantum systems $\mathcal{H}_A \otimes \mathcal{H}_B$ (describing, e.g., pairs of entangled particles). The protocol then runs through the following steps in order to transform this initial entanglement between Alice and Bob into a common secret key.

- *Parameter estimation:* Alice and Bob sacrifice some small number, say m, subsystems $\mathcal{H}_A \otimes \mathcal{H}_B$ in order to estimate their average correlation. For this, they both apply measurements with respect to different bases and publicly announce the outcomes (using the authentic classical communication channel). Depending on the resulting statistics, they either decide to proceed with the computation of the key or to abort the protocol.

- *Measurement:* Alice and Bob both apply measurements to their parts of the remaining subsystems $\mathcal{H}_A \otimes \mathcal{H}_B$ to obtain a pair of *raw keys*. (Note that these raw keys are generally only weakly correlated and partially secure.)

- *Block-wise processing:* Alice and Bob might[34] further process their raw key pair in order to improve its correlation or secrecy. We assume that this processing acts on n blocks of size b individually. For example, Alice and Bob might invoke a so-called *advantage distillation* protocol (see Section 7.1.3) whose purpose is to single out blocks of the raw key that are highly correlated. We denote by X^n and Y^n the strings held by Alice and Bob after this step.

- *Information reconciliation:* The purpose of this step is to transform the (possibly only weakly correlated) pair of strings X^n and Y^n into a pair of identical strings. Typically, Alice sends certain error correcting information on X^n to Bob which allows him to compute a guess $\hat{X}^n$ of X^n.

- *Privacy amplification:* Alice and Bob use two-universal hashing to transform their strings X^n and $\hat{X}^n$ into secret keys of length ℓ.

Additionally, we assume that the action of the protocol is invariant under permutations of the N input systems. This does not restrict the generality of our results, because any protocol can easily be turned into a permutation-invariant one: Before starting with the parameter estimation, Alice and Bob simply have to (publicly) agree on a random permutation which they use to reorder their subsystems (see Section 1.6.3 below for more details).

[34] In many protocols, this step is omitted, i.e., Alice and Bob directly proceed with information reconciliation.

1.6.2 Security criterion

The security of a key distillation scheme depends on the actual choice of various protocol parameters which we define in the following:

- Γ is the set of states on *single* subsystems which are not filtered by the *parameter estimation* subprotocol: More precisely, Γ contains all density operators σ_{AB} such that, when starting with the product state $\rho_{A^N B^N} := \sigma_{AB}^{\otimes N}$, the protocol does *not* abort.
- $\mathcal{E}_{XY\bar{E} \leftarrow A^b B^b E^b}$ is the CPM[35] on b subsystems which describes the *measurement* together with the *block-wise processing* on blocks of size b.
- n is the *number of blocks* of size b that are used for the actual computation of the key (i.e., the number of blocks of subsystems that are left after the parameter estimation step).
- ℓ denotes the length of the final key generated in the *privacy amplification* step.

In addition, the security of the scheme depends on the efficiency of the *information reconciliation* subprotocol, i.e., the amount of information that is leaked to Eve. However, for this summary, we assume that Alice and Bob use an *optimal*[36] information reconciliation protocol. In this case, the leakage is roughly equal to the entropy of X^n given Y^n.[37]

We are now ready to formulate a general security criterion for quantum key distillation (cf. Theorem 6.5.1): The scheme described above is secure (for any initial state) if[38]

$$\frac{\ell}{n} \lessapprox \min_{\sigma_{AB} \in \Gamma} H(X|E) - H(X|Y) \ , \tag{1.9}$$

where the minimum ranges over all states σ_{AB} contained in the set Γ defined above and where $H(X|E)$ and $H(X|Y)$ are the (conditional) von Neumann entropies of

$$\sigma_{XYE} := \mathcal{E}_{XY\bar{E} \leftarrow A^b B^b E^b}(\sigma_{ABE}^{\otimes b})$$

where σ_{ABE} is a purification of σ_{AB}. Note that, because the operators σ_{AB} are on *single* subsystems, formula (1.9) is usually fairly easy to evaluate for concrete protocols (cf. Chapter 7).

Typically, the number m of subsystem that are sacrificed for parameter estimation is small compared to the total number N of initial subsystems.

---

[35] See Section 2.1.1 for a definition of completely positive maps (CPM).

[36] In Section 6.3, we show that optimal information reconciliation protocols exist.

[37] We refer to Chapter 6 for the general result which deals with arbitrary—not necessarily optimal—information reconciliation schemes.

[38] The approximation $\lessapprox$ in (1.9) indicates that the criterion holds asymptotically for increasing n. We refer to Chapter 6 for a non-asymptotic result.

Hence, the number n of blocks of size b that can be used for the actual computation of the key is roughly given by $n \approx \frac{N}{b}$.[39] The criterion (1.9) can thus be turned into an expression for the *key rate* of the protocol (i.e., the number of key bits generated per channel use):

$$\text{rate} = \frac{1}{b} \min_{\sigma_{AB} \in \Gamma} H(X|E) - H(X|Y) \ .$$

1.6.3 Security proof

We need to show that, for *any* initial state shared by Alice and Bob, the probability that the protocol generates an insecure key is negligible.[40] Roughly speaking, the proof consists of two parts. In the first (Steps 1–2) we argue that we can restrict our analysis to a much smaller set of initial states, namely those that have (almost) product form. In the second part (Steps 3–5) we show that for each such state either of the following holds: (i) there is not sufficient correlation between Alice and Bob in which case the protocol aborts during the parameter estimation or (ii) a measurement applied to the state generates an outcome with sufficient entropy such that the key computed from it is secure.

Step 1: Restriction to permutation-invariant initial states

As we assumed that the protocol is invariant under permutations of the input systems, we can equivalently think of a protocol which starts with the following symmetrization step: Alice chooses a permutation π at random and announces it to Bob, using the (insecure) classical communication channel. Then Alice and Bob both permute the order of their N subsystems according to π. Obviously, the state $\rho_{A^N B^N}$ of Alice and Bob's system after this symmetrization step (averaged over all choices of π) is invariant under permutations.

Because the state $\rho_{A^N B^N}$ is invariant under permutations, it has a purification $\rho_{A^N B^N E^N}$ (with an auxiliary system $\mathcal{H}_E^{\otimes N}$) which is symmetric as well (cf. Lemma 4.2.2). As the pure state $\rho_{A^N B^N E^N}$ cannot be correlated with anything else (cf. Section 2.1.2) we can assume without loss of generality that the knowledge of a potential adversary is fully described by the auxiliary system.

[39] This is also true for QKD protocols with a sifting step (where Alice and Bob discard the subsystems for which they have used incompatible encoding and decoding bases). In fact, as mentioned in Section 1.2, if Alice and Bob choose one of the bases with probability close to one, the fraction of positions lost in the sifting step is small.

[40] Note that the protocol might abort if the initial state held by Alice and Bob is not sufficiently correlated.

Step 2: Restriction to (almost) product states

Because $\rho_{A^N B^N E^N}$ is invariant under permutations, it is, according to our finite version of the de Finetti representation theorem approximated by a mixture of states which have "almost" product form $\sigma_{ABE}^{\otimes N}$—in the sense described by formula (1.5).

Step 3: Smooth min-entropy of Alice and Bob's raw keys

Assume for the moment that the joint initial state $\rho_{A^N B^N E^N}$ held by Alice, Bob, and Eve has perfect product form $\sigma_{ABE}^{\otimes N}$. As Alice and Bob's measurement operation (including the block-wise processing) $\mathcal{E}_{XY \leftarrow AB}$ acts on n blocks of size b individually, the density operator $\rho_{X^n Y^n E^n}$ which describes the situation before the information reconciliation step is given by

$$\rho_{X^n Y^n E^n} = (\mathcal{E}_{XY \leftarrow AB} \otimes \mathrm{id}_E)^{\otimes n}(\rho_{A^{bn} B^{bn} E^{bn}}) \ ,$$

where X^n and Y^n is Alice and Bob's raw key, respectively. Consequently, $\rho_{X^n Y^n E^n}$ is the product of operators of the form

$$\sigma_{XYE} = (\mathcal{E}_{XY \leftarrow AB} \otimes \mathrm{id}_E)(\sigma_{ABE}^{\otimes b}) \ . \tag{1.10}$$

By (1.4), the smooth min-entropy of $\rho_{X^n E^n}$ is approximated in terms of the von Neumann entropy of σ_{XE}, i.e.,

$$H_{\min}^{\varepsilon}(X^n | E^n) \gtrapprox nH(X|E) \ . \tag{1.11}$$

Using (1.6), this argument can easily be generalized to states $\rho_{A^N B^N E^N}$ which have almost product form.

Step 4: Smooth min-entropy after information reconciliation

In the information reconciliation step, Alice sends error correcting information C about X^n to Bob, using the authentic classical communication channel. Eve might wiretap this communication which generally decreases the smooth min-entropy of X^n from her point of view.

As mentioned above, we assume for this summary that the information reconciliation subprotocol is optimal with respect to the amount of information leaked to Eve. It follows from classical coding theory that the number of bits that Alice has to send to Bob in order to allow him to compute her value X^n is given by the Shannon entropy of X^n conditioned on Bob's knowledge Y^n. Formally, if $\rho_{X^n Y^n}$ has product form $\sigma_{XY}^{\otimes n}$ then the communication C satisfies

$$H_{\max}(C) - H_{\min}(C|X^n) \approx nH(X|Y) \ , \tag{1.12}$$

where $H(X|Y)$ is the Shannon entropy of X given Y, evaluated for the probability distribution defined by σ_{XY}. (Note that the entropy difference on the left hand side can be interpreted as a measure for the information that C gives on X^n.)

Let us now compute a lower bound on the smooth min-entropy of X^n given Eve's knowledge after the information reconciliation step. By the chain rule (1.2), we have

$$H^\varepsilon_{\min}(X^n|CE^n) \geq H^\varepsilon_{\min}(X^n C|E^n) - H_{\max}(C) \ .$$

Moreover, because C is computed from X^n, we can apply inequality (1.3), i.e.,

$$H^\varepsilon_{\min}(X^n C|E^n) \geq H^\varepsilon_{\min}(X^n|E^n) + H_{\min}(C|X^n) \ .$$

Combining this with (1.12) gives

$$\begin{aligned} H^\varepsilon_{\min}(X^n|CE^n) &\geq H^\varepsilon_{\min}(X^n|E^n) - \big(H_{\max}(C) - H_{\min}(C|X^n)\big) \\ &\approx H^\varepsilon_{\min}(X^n|E^n) - nH(X|Y) \ . \end{aligned}$$

Finally, using the approximation (1.11) for $H^\varepsilon_{\min}(X^n|E^n)$, we conclude

$$H^\varepsilon_{\min}(X^n|CE^n) \gtrapprox nH(X|E) - nH(X|Y) \ . \tag{1.13}$$

Step 5: Security of the key generated by privacy amplification

To argue that the key generated in the final privacy amplification step is secure, we apply criterion (1.8). Because the adversary has access to both the quantum system and the classical communication C, this security criterion reads

$$\ell \lessapprox H^\varepsilon_{\min}(X^n|E^n C) \tag{1.14}$$

where ℓ is the length of the key.

As shown in Step 2, the state $\rho_{A^N B^N E^N}$ has almost product form $\sigma_{ABE}^{\otimes N}$. Hence, according to (1.7), the statistics obtained by Alice and Bob in the parameter estimation step corresponds to the statistics that they would obtain if they started with a perfect product state $\sigma_{AB}^{\otimes N}$. We conclude that, by the definition of the set Γ, the protocol aborts whenever $\sigma_{AB} \notin \Gamma$.

To bound the smooth min-entropy of the string held by Alice before privacy amplification, it thus suffices to evaluate (1.13) for all states σ_{AB} contained in Γ. Formally,

$$\frac{1}{n} H^\varepsilon_{\min}(X^n|CE^n) \gtrapprox \min_{\sigma_{ABE}} H(X|E) - H(X|Y) \ .$$

where the minimum is over all (pure) states σ_{ABE} such that $\sigma_{AB} \in \Gamma$ and where $H(X|E)$ and $H(X|Y)$ are the entropies of the state σ_{XYE} given by (1.10). Combining this with criterion (1.14) concludes the proof.

Chapter 2

Preliminaries

2.1 Representation of physical systems

2.1.1 Density operators, measurements, and operations

Quantum mechanics, like any other physical theory, allows us to make certain predictions about the behavior of physical systems. These are, however, not deterministic—a system's initial state merely determines a probability distribution over all possible outcomes of an observation.[1]

Mathematically, the state of a quantum mechanical system with d degrees of freedom is represented by a normalized nonnegative[2] operator ρ, called *density operator*, on a d-dimensional Hilbert space $\mathcal{H}$. The normalization is with respect to the trace norm, i.e., $\|\rho\|_1 = \mathrm{tr}(\rho) = 1$. In the following, we denote by $\mathcal{P}(\mathcal{H})$ the set of nonnegative operators on $\mathcal{H}$, i.e., ρ is a density operator on $\mathcal{H}$ if and only if $\rho \in \mathcal{P}(\mathcal{H})$ and $\mathrm{tr}(\rho) = 1$.

Any observation of a quantum system corresponds to a *measurement* and is represented mathematically as a *positive operator valued measure (POVM)*, i.e., a family $\mathcal{M} = \{M_w\}_{w \in \mathcal{W}}$ of nonnegative operators such that $\sum_{w \in \mathcal{W}} M_w = \mathrm{id}_{\mathcal{H}}$. The theory of quantum mechanics postulates that the probability distribution P_W of the outcomes when measuring a system in state ρ with respect to $\mathcal{M}$ is given by $P_W(w) := \mathrm{tr}(M_w \rho)$.

Consider a physical system whose state $\bar{\rho}^z$ depends on the value z of a classical random variable Z with distribution P_Z. For an observer which is ignorant of the value of Z, the state ρ of the system is given by the convex

[1] With his famous statement "Gott würfelt nicht," Einstein expressed his doubts about the completeness of such a non-deterministic theory.

[2] An operator ρ on $\mathcal{H}$ is *nonnegative* if it is hermitian and has nonnegative eigenvalues.

combination[3]

$$\rho = \sum_{z \in \mathcal{Z}} P_Z(z) \bar{\rho}^z . \tag{2.1}$$

The decomposition (2.1) of a density operator ρ is generally not unique. Consider for example the *fully mixed state* defined by $\rho := \frac{1}{\dim(\mathcal{H})} \mathrm{id}_{\mathcal{H}}$. In the case of a two-level system, ρ might represent a photon which is polarized horizontally or vertically with equal probabilities; but the same operator ρ might also represent a photon which is polarized according to one of the two diagonal directions with equal probabilities. In fact, the two settings cannot be distinguished by any measurement.

A physical process is most generally described by a linear mapping $\mathcal{E}$, called a *quantum operation*, which takes the system's initial state ρ to its final state ρ'.[4] Mathematically, a quantum operation $\mathcal{E}$ is a *completely positive map (CPM)*[5] from the set of hermitian operators on a Hilbert space $\mathcal{H}$ to the set of hermitian operators on another Hilbert space $\mathcal{H}'$. Additionally, in order to ensure that the image $\mathcal{E}(\rho)$ of a density operator ρ is again a density operator, $\mathcal{E}$ must be *trace-preserving*, i.e., $\mathrm{tr}(\mathcal{E}(\rho)) = \mathrm{tr}(\rho)$, for any $\rho \in \mathcal{P}(\mathcal{H})$. It can be shown that any CPM $\mathcal{E}$ can be written as

$$\mathcal{E}(\rho) = \sum_{w \in \mathcal{W}} E_w \rho E_w^\dagger \tag{2.2}$$

where $\{E_w\}_{w \in \mathcal{W}}$ is a family of linear operators from $\mathcal{H}$ to $\mathcal{H}'$. On the other hand, any mapping of the form (2.2) is a CPM.[6] Moreover, it is trace-preserving if and only if $\sum_{w \in \mathcal{W}} E_w^\dagger E_w = \mathrm{id}_{\mathcal{H}}$.

As we have seen, the state of a quantum system might depend on some classical event Ω (e.g., that Z takes a certain value z). In this context, it is often convenient to represent both the probability $\Pr[\Omega]$ of Ω and the state $\bar{\rho}^\Omega$ of the system conditioned on Ω as one single mathematical object, namely the nonnegative operator $\rho^\Omega := \Pr[\Omega] \cdot \bar{\rho}^\Omega$.[7] For this reason, we formulate most statements on quantum states in terms of general (not necessarily normalized) nonnegative operators. Similarly, we often consider general (not necessarily trace-preserving) CPMs $\mathcal{E}$. The quantity $\mathrm{tr}(\mathcal{E}(\rho))$ can then be

[3] Because a measurement is a linear mapping from the set of density operators to the set of probability distributions, this is consistent with the above description. In particular, the distribution of the outcomes resulting from a measurement of ρ is the convex combination of the distributions obtained from measurements of $\bar{\rho}^z$.

[4] A measurement can be seen as a special case of a quantum operation where the outcome is classical (see Section 2.1.3).

[5] *Complete positivity* means that any extension $\mathcal{E} \otimes \mathrm{id}$ of the map $\mathcal{E}$, where id is the identity map on the set of hermitian operators on some auxiliary Hilbert space $\mathcal{H}''$, maps nonnegative operators to nonnegative operators. Formally, $(\mathcal{E} \otimes \mathrm{id})(\rho) \in \mathcal{P}(\mathcal{H}' \otimes \mathcal{H}'')$ for any $\rho \in \mathcal{P}(\mathcal{H} \otimes \mathcal{H}'')$.

[6] This is in fact a direct consequence of Lemma B.5.1.

[7] The probability of the event Ω is then equal to the trace of $\bar{\rho}^\Omega$, i.e., $\Pr[\Omega] = \mathrm{tr}(\rho^\Omega)$, and the system's state conditioned on Ω is $\bar{\rho}^\Omega = \frac{1}{\Pr[\Omega]} \rho^\Omega$.

interpreted as the probability that the process represented by $\mathcal{E}$ occurs when starting with a system in state ρ.

2.1.2 Product systems and purifications

To analyze complex physical systems, it is often convenient to consider a partitioning into a number of subsystems. This is particularly useful if one is interested in the study of operations that act on the parts of the system individually.[8] Mathematically, the partition of a quantum system into subsystems induces a product structure on the underlying Hilbert space. For example, the state of a bipartite system is represented as a density operator ρ_{AB} on a product space $\mathcal{H}_A \otimes \mathcal{H}_B$. The state of one part of a product system is then obtained by taking the corresponding partial trace of the overall state, e.g., $\rho_A = \mathrm{tr}_B(\rho_{AB})$ for the first part of a bipartite system.

A density operator ρ on $\mathcal{H}$ is said to be *pure* if it has rank[9] one, that is, $\rho = |\theta\rangle\langle\theta|$, for some $|\theta\rangle \in \mathcal{H}$. If it is normalized, ρ is a projector[10] onto $|\theta\rangle$. A pure density operator can only be decomposed trivially, i.e., for any decomposition of the form (2.2), $\bar{\rho}^z = \rho$ holds for all $z \in \mathcal{Z}$. According to the above interpretation, one could say that a pure state contains no classical randomness, that is, it cannot be correlated with any other system.

The fact that a pure state cannot be correlated with the environment plays a crucial role in cryptography. It implies, for example, that the randomness obtained from the measurement of a pure state is independent of any other system and thus guaranteed to be secret. More generally, let ρ_A be an arbitrary operator on $\mathcal{H}_A$ and let ρ_{AE} be a *purification* of ρ_A, i.e., ρ_{AE} is a pure state on a product system $\mathcal{H}_A \otimes \mathcal{H}_E$ such that $\mathrm{tr}_E(\rho_{AE}) = \rho_A$. Then, because ρ_{AE} is uncorrelated with any other system, the partial system $\mathcal{H}_E$ comprises everything that might possibly be correlated with the system $\mathcal{H}_A$ (including the knowledge of a potential adversary).

2.1.3 Quantum and classical systems

Consider a classical random variable Z with distribution P_Z on some set $\mathcal{Z}$. In a quantum world, it is useful to view Z as a special case of a quantum system. For this, one might think of the classical values $z \in \mathcal{Z}$ as being represented by orthogonal[11] states $|z\rangle$ on some Hilbert space $\mathcal{H}_{\mathcal{Z}}$. The state

---

[8] This is typically the case in the context of cryptography, where various parties control separated subsystems.

[9] The *rank* of a hermitian operator S, denoted $\mathrm{rank}(S)$, is the dimension of the *support* $\mathrm{supp}(S)$, i.e., the space spanned by the eigenvectors of S with nonzero eigenvalues.

[10] A hermitian operator P is said to be a *projector* if $PP = P$.

[11] The orthogonality of the states $|z\rangle$ guarantees that they can be distinguished perfectly, as this is the case for classical values.

ρ_Z of the quantum system is then defined by

$$\rho_Z = \sum_{z \in \mathcal{Z}} P_Z(z) |z\rangle\langle z| \ . \tag{2.3}$$

We say that ρ_Z is the *operator representation* of the classical distribution P_Z *(with respect to the basis* $\{|z\rangle\}_{z\in\mathcal{Z}}$*)*.[12]

On the other hand, any operator ρ_Z can be written in the form (2.3) where $P_Z(z)$ are the *eigenvalues* of ρ_Z and $|z\rangle$ are the corresponding *eigenvectors*. The right hand side of (2.3) is called the *spectral decomposition* of ρ_Z. Moreover, we say that P_Z is the *probability distribution defined by* ρ_Z.

This notion can be extended to hybrid settings where the state $\bar{\rho}_A^z$ of a quantum system $\mathcal{H}_A$ depends on the value z of a classical random variable Z (see, e.g., [DW05]). The joint state of the system is then given by

$$\rho_{AZ} = \sum_{z \in \mathcal{Z}} \rho_A^z \otimes |z\rangle\langle z| \ , \tag{2.4}$$

where $\rho_A^z := P_Z(z)\bar{\rho}_A^z$.

We can also go in the other direction: If a density operator has the form (2.4), for some basis $\{|z\rangle\}_{z\in\mathcal{Z}}$, then the first subsystem can be interpreted as the representation of a classical random variable Z. This motivates the following definition: An operator $\rho_{AZ} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_Z)$ is said to be *classical with respect to* $\{|z\rangle\}_{z\in\mathcal{Z}}$ if there exists a family $\{\rho_A^z\}_{z\in\mathcal{Z}}$ of operators on $\mathcal{H}_A$, called *(non-normalized) conditional operators*, such that ρ_{AZ} can be written in the form (2.4). Moreover, we say that ρ_{AZ} is *classical on* $\mathcal{H}_Z$ if there exists a basis $\{|z\rangle\}_{z\in\mathcal{Z}}$ of $\mathcal{H}_Z$ such that ρ_{AZ} is classical with respect to $\{|z\rangle\}_{z\in\mathcal{Z}}$.[13]

A similar definition can be used to characterize quantum operations (i.e., CPMs) whose outcomes are partly classical: A CPM $\mathcal{E}$ from $\mathcal{H}$ to $\mathcal{H}_A \otimes \mathcal{H}_Z$ is said to be *classical with respect to* $\{|z\rangle\}_{z\in\mathcal{Z}}$ (or simply *classical on* $\mathcal{H}_Z$) if it can be written as

$$\mathcal{E}(\sigma) = \sum_{z \in \mathcal{Z}} \mathcal{E}^z(\sigma) \otimes |z\rangle\langle z| \ ,$$

where, for any $z \in \mathcal{Z}$, $\mathcal{E}^z$ is a CPM from $\mathcal{H}$ to $\mathcal{H}_A$. Note that a measurement on $\mathcal{H}$ with outcomes in $\mathcal{Z}$ can be seen as a CPM from $\mathcal{H}$ to $\mathcal{H}_Z$ which is classical on $\mathcal{H}_Z$.

---

[12] This definition can easily be generalized to multi-partite nonnegative (not necessarily normalized) functions (e.g., $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, where $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ denotes the set of nonnegative functions on $\mathcal{X} \times \mathcal{Y}$) in which case one gets nonnegative operators on product systems (e.g., $\rho_{XY} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_Y)$).

[13] The operators ρ_A^z, for $z \in \mathcal{Z}$, are uniquely defined by ρ_{AZ} and the basis $\{|z\rangle\}_{z\in\mathcal{Z}}$. Moreover, because ρ_{AZ} is nonnegative, the operators ρ_A^z, for $z \in \mathcal{Z}$, are also nonnegative.

2.1.4 Distance between states

Intuitively, we say that two states of a physical system are *similar* if any observation of them leads to identical results, except with small probability. For two operators $\rho, \rho' \in \mathcal{P}(\mathcal{H})$ representing the state of a quantum system, this notion of similarity is captured by the L_1-*distance*, i.e., the trace norm[14] $\|\rho-\rho'\|_1$ of the difference between ρ and ρ'.[15] The L_1-distance for operators can be seen as the quantum version of the L_1-distance for probability distributions [16] (or, more generally, nonnegative functions), which is defined by $\|P - P'\|_1 := \sum_z |P(z) - P'(z)|$, for $P, P' \in \mathcal{P}(\mathcal{Z})$. In particular, if ρ and ρ' are operator representations of probability distributions P and P', respectively, then the L_1-distance between ρ and ρ' is equal to the L_1-distance between P and P'.

Under the action of a quantum operation, the L_1-distance between two density operators ρ and ρ' cannot increase (cf. Lemma A.2.1). Because any measurement can be seen as a quantum operation, this immediately implies that the distance $\|P - P'\|_1$ between the distributions P and P' obtained from (identical) measurements of two density operators ρ and ρ', respectively, is bounded by $\|\rho - \rho'\|_1$.

The following proposition provides a very simple interpretation of the L_1-distance: If two probability distributions P and P' have L_1-distance at most 2ε, then the two settings described by P and P', respectively, cannot differ with probability more than ε.

Proposition 2.1.1. *Let $P, P' \in \mathcal{P}(\mathcal{X})$ be probability distributions. Then there exists a joint distribution $P_{XX'}$ such that P and P' are the marginals of $P_{XX'}$ (i.e., $P = P_X$, $P' = P_{X'}$) and, for (x, x') chosen according to $P_{XX'}$,*

$$\Pr_{(x,x')}[x \neq x'] \leq \frac{1}{2}\|P - P'\|_1 .$$

In particular, if the L_1-distance between two states is bounded by 2ε, then they cannot be distinguished with probability more than ε.

2.2 Universal security of secret keys

Cryptographic primitives (e.g., a secret key or an authentic communication channel) are often used as components within a more complex system. It is thus natural to require that the security of a cryptographic scheme is not compromised when it is employed as part of another system. This requirement is captured by the notion of *universal security*. Roughly speaking,

[14] The *trace norm* $\|S\|_1$ of a hermitian operator S on $\mathcal{H}$ is defined by $\|S\|_1 := \mathrm{tr}(|S|)$.

[15] The L_1-distance between two operators is closely related to the *trace distance*, which is usually defined with an additional factor $\frac{1}{2}$.

[16] The L_1-distance between classical probability distributions is also known as *variational distance* or *statistical distance* (which are often defined with an additional factor $\frac{1}{2}$).

we say that a cryptographic primitive is *universally secure* if it is secure in *any* arbitrary context. For example, the universal security of a secret key S implies that any bit of S remains secret even if some other part of S is given to an adversary.

In the past few years, universal security has attracted a lot of interest and led to important new definitions and proofs (see, e.g., the so-called *universal composability* framework of Canetti [Can01] or Pfitzmann and Waidner [PW00]). Recently, Ben-Or and Mayers [BOM04] and Unruh [Unr04] have generalized Canetti's notion of universal composability to the quantum world.

Universal security definitions are usually based on the idea of characterizing the security of a *real* cryptographic scheme by its distance to an *ideal* system which (by definition) is perfectly secure. For instance, a secret key S is said to be secure if it is close to a *perfect key* U, i.e., a uniformly distributed string which is independent of the adversary's information. As we shall see, such a definition immediately implies that any cryptosystem which is proven secure when using a perfect key U remains secure when U is replaced by the (real) key S.

2.2.1 Standard security definitions are not universal

Unfortunately, many security definitions that are commonly used in quantum cryptography are not universal. For instance, the security of the key S generated by a QKD scheme is typically defined in terms of the mutual information $I(S;W)$ between S and the classical outcome W of a measurement of the adversary's system (see, e.g., [LC99, SP00, NC00, GL03, LCA05] and also the discussion in [BOHL$^+$05] and [RK05]). Formally, S is said to be secure if, for some small ε,

$$\max_W I(S;W) \leq \varepsilon , \qquad (2.5)$$

where the maximum ranges over all measurements on the adversary's system with output W. Such a definition—although it looks reasonable—does, however, not guarantee that the key S can safely be used in applications. Roughly speaking, the reason for this flaw is that criterion (2.5) does not account for the fact that an adversary might wait with the measurement of her system until she learns parts of the key. (We also refer to [RK05] and [BOHL$^+$05] for a more detailed discussion and an analysis of existing security definitions with respect to this concern.[17])

[17] Note that the conclusions in [BOHL$^+$05] are somewhat different to ours: It is shown that existing privacy conditions of the form (2.5) *do* imply universal security, which seems to contradict the counterexample sketched below. However, the result of [BOHL$^+$05] only holds if the parameter ε in (2.5) is exponentially small in the key size, which is not the case for most of the existing protocols. (In fact, the security parameter ε can only be made exponentially small at the expense of decreasing the key rate substantially.)

Let us illustrate this potential problem with a concrete example: Assume that we would like to use an n-bit key $S = (S_1, \ldots, S_n)$ as a one-time pad to encrypt an n-bit message $M = (M_1, \ldots, M_n)$.[18] Furthermore, assume that an adversary is interested in the nth bit M_n of the message, but already knows the first $n-1$ bits $M_1, \ldots, M_{n-1}$. Upon observing the ciphertext, the adversary can easily determine[19] the first $n-1$ bits of S. Hence, in order to guarantee the secrecy of the nth message bit M_n, we need to ensure that the adversary still has no information on the nth key bit S_n, even though she already knows all previous key bits $S_1, \ldots, S_{n-1}$. This requirement, however, is not implied by the above definition. Indeed, for any arbitrary $\varepsilon > 0$ and n depending on ε, it is relatively easy to construct examples which satisfy (2.5) whereas an adversary—once she knows the first $n-1$ bits of the key—can determine the nth bit S_n with certainty. For an explicit construction and analysis of such examples, we refer to [Bar05].[20]

2.2.2 A universal security definition

Consider a key S distributed according to P_S and let ρ_E^s be the state of the adversary's system given that S takes the value s, for any element s of the *key space* $\mathcal{S}$. According to the discussion in Section 2.1.3, the joint state of the classical key S and the adversary's quantum system can be represented by the density operator

$$\rho_{SE} := \sum_{s \in \mathcal{S}} P_S(s) |s\rangle\langle s| \otimes \rho_E^s ,$$

where $\{|s\rangle\}_{s \in \mathcal{S}}$ is an orthonormal basis of some Hilbert space $\mathcal{H}_S$. We say that S is *ε-secure with respect to $\mathcal{H}_E$* if

$$\frac{1}{2} \big\| \rho_{SE} - \rho_U \otimes \rho_E \big\|_1 \leq \varepsilon , \tag{2.6}$$

where $\rho_U = \sum_{s \in \mathcal{S}} \frac{1}{|\mathcal{S}|} |s\rangle\langle s|$ is the fully mixed state on $\mathcal{H}_S$.

The universal security of a key S satisfying this definition follows from a simple argument: Criterion (2.6) guarantees that the *real* situation described by ρ_{SE} is ε-close—with respect to the L_1-distance—to an *ideal* situation where S is replaced by a perfect key U which is uniformly distributed and independent of the state of the system $\mathcal{H}_E$. Moreover, since the L_1-distance cannot increase when applying a quantum operation (cf. Lemma A.2.1), this also holds for any further evolution of the world (where, e.g., the key is used as part of a larger cryptographic system). In fact, it follows from the

[18] That is, the ciphertext $C = (C_1, \ldots, C_n)$ is the bit-wise XOR of S and M, i.e., $C_i = S_i \oplus M_i$.

[19] Note that $S_i = M_i \oplus C_i$.

[20] This phenomenon has also been studied in other contexts (see, e.g., [DHL$^+$04, HLSW04]) where it is called as *locking of classical correlation*.

discussion in Section 2.1.4 that an ε-secure key can be considered *identical* to an *ideal* (perfect) key—except with probability ε.[21] In particular, an ε-secure key is secure within any reasonable framework providing universal composability (e.g., [BOM04] or [Unr04]).[22]

The security of a key according to (2.6) also implies security with respect to most of the standard security definitions in quantum cryptography. For example, if S is ε-secure with respect to $\mathcal{H}_E$ then the mutual information between S and the outcome of any measurement applied to the adversary's system is small (whereas the converse is often not true, as discussed above). In particular, if the adversary is purely classical, (2.6) reduces to a classical security definition which has been proposed in the context of information-theoretically secure key agreement (see, e,g., [DM04]).

[21] For this statement to hold, it is crucial that the criterion (2.6) is formulated in terms of the L_1-distance (instead of other distance measures such as the fidelity).

[22] These frameworks are usually based on the so-called *simulatability paradigm*. That is, a real cryptosystem is said to be *as secure as* an ideal cryptosystem if any attack to the real scheme can be *simulated* by an attack to the ideal scheme (see also [MRH04]). It is easy to see that our security criterion is compatible with this paradigm: Consider a (real) key agreement protocol and assume that, for any possible attack of the adversary, the final key satisfies (2.6). The adversary's quantum state after the attack is then almost independent of the key, that is, the adversary could simulate virtually all her information without even interacting with the cryptosystem. The real key agreement protocol is thus *as secure as* an ideal key agreement scheme which, by definition, does not leak any information at all.

Chapter 3

(Smooth) Min- and Max-Entropy

Entropy measures are indispensable tools in classical and quantum information theory. They quantify *randomness*, that is, the uncertainty that an observer has on the state of a (quantum) physical system. In this chapter, we introduce two entropic quantities, called *smooth min-entropy* and *smooth max-entropy*. As we shall see, these are useful to characterize randomness with respect to fundamental information-theoretic tasks such as the extraction of uniform randomness or data compression.[1] Moreover, smooth min- and max-entropies have natural properties which are similar to those known from the von Neumann entropy and its classical special case, the Shannon entropy[2] (Sections 3.1 and 3.2). In fact, for product states, smooth min- and max-entropy are asymptotically equal to the von Neumann entropy (Section 3.3).

Smooth min- and max-entropies are actually families of entropy measures parameterized by some nonnegative real number ε, called *smoothness*. In applications, the smoothness is related to the error probability of certain information-theoretic tasks and is thus typically chosen to be small. We first consider the "non-smooth" special case where $\varepsilon = 0$ (Section 3.1). This is the basis for the general definition where the smoothness ε is arbitrary (Section 3.2).

We will introduce a *conditional* version of smooth min- and max-entropy. It is defined for bipartite operators ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ and measures the uncertainty on the state of the subsystem $\mathcal{H}_A$ given access to the subsystem $\mathcal{H}_B$. Unlike the *conditional von Neumann entropy* $H(A|B) := H(\rho_{AB}) - H(\rho_B)$,

[1] *Randomness extraction* is actually privacy amplification and is the topic of Chapter 5. *Data compression* is closely related to information reconciliation which is treated in Section 6.3.

[2] The *Shannon entropy* of a probability distribution P is defined by $H(P) := -\sum_x P(x) \log P(x)$, where log denotes the binary logarithm. Similarly, the *von Neumann entropy* of a density operator ρ is $H(\rho) := -\mathrm{tr}(\rho \log \rho)$.

however, it cannot be written as a difference between two "unconditional" entropy measures.

To illustrate our definition of (conditional) min- and max-entropy, let us, as an analogy, consider an alternative formulation of the conditional von Neumann entropy $H(A|B)$. Let

$$H(\rho_{AB}|\sigma_B) := -\mathrm{tr}\big(\rho_{AB}(\log \rho_{AB} - \log \mathrm{id}_A \otimes \sigma_B)\big) , \tag{3.1}$$

for some state σ_B on $\mathcal{H}_B$. This quantity can be rewritten as

$$H(\rho_{AB}|\sigma_B) = H(\rho_{AB}) - H(\rho_B) - D(\rho_B \| \sigma_B) ,$$

where $D(\rho_B\|\sigma_B)$ is the relative entropy[3] of ρ_B to σ_B. Because $D(\rho_B\|\sigma_B)$ cannot be negative, this expression takes its maximum for $\sigma_B = \rho_B$, in which case it is equal to $H(A|B)$. We thus have

$$H(A|B) = \sup_{\sigma_B} H(\rho_{AB}|\sigma_B) , \tag{3.2}$$

where the supremum ranges over all density operators σ_B on $\mathcal{H}_B$.

The definitions of (smooth) min- and max-entropies are inspired by this approach. We first introduce a quantity which corresponds to (3.1) (cf. Definitions 3.1.1 and 3.2.1) and then define our entropy measures by a formula of the form (3.2) (Definitions 3.1.2 and 3.2.2).

3.1 Min- and max-entropy

This section introduce a "non-smooth" version of min- and max-entropy. It is the basis for the considerations in Section 3.2, where these entropy measures are generalized. The focus is on min-entropy, which is used extensively in the remaining part of the thesis. However, most of the properties derived in the following also hold for max-entropy.

3.1.1 Definition of min- and max-entropy

Definition 3.1.1. Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$. The *min-entropy of ρ_{AB} relative to σ_B* is

$$H_{\min}(\rho_{AB}|\sigma_B) := -\log \lambda$$

where λ is the minimum real number such that $\lambda \cdot \mathrm{id}_A \otimes \sigma_B - \rho_{AB}$ is non-negative. The *max-entropy of ρ_{AB} relative to σ_B* is

$$H_{\max}(\rho_{AB}|\sigma_B) := \log \mathrm{tr}\big((\mathrm{id}_A \otimes \sigma_B)\rho^0_{AB}\big)$$

where ρ^0_{AB} denotes the projector onto the support of ρ_{AB}.

[3]The *relative entropy* $D(\rho\|\sigma)$ is defined by $D(\rho\|\sigma) := \mathrm{tr}(\rho \log \rho) - \mathrm{tr}(\rho \log \sigma)$.

Definition 3.1.2. Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$. The *min-entropy* and the *max-entropy of ρ_{AB} given $\mathcal{H}_B$* are

$$\begin{aligned} H_{\min}(\rho_{AB}|B) &:= \sup_{\sigma_B} H_{\min}(\rho_{AB}|\sigma_B) \\ H_{\max}(\rho_{AB}|B) &:= \sup_{\sigma_B} H_{\max}(\rho_{AB}|\sigma_B) \ , \end{aligned}$$

respectively, where the supremum ranges over all $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ with $\mathrm{tr}(\sigma_B) = 1$.

Remark 3.1.3. It follows from Lemma B.5.3 that the min-entropy of ρ_{AB} relative to σ_B, for σ_B invertible, can be written as

$$H_{\min}(\rho_{AB}|\sigma_B) = -\log \lambda_{\max}\big((\mathrm{id}_A \otimes \sigma_B^{-1/2})\rho_{AB}(\mathrm{id}_A \otimes \sigma_B^{-1/2})\big) \ ,$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of the argument.

If $\mathcal{H}_B$ is the trivial space $\mathbb{C}$, we simply write $H_{\min}(\rho_A)$ and $H_{\max}(\rho_A)$ to denote the min- and the max-entropy of ρ_A, respectively. In particular,

$$\begin{aligned} H_{\min}(\rho_A) &= -\log \lambda_{\max}(\rho_A) \\ H_{\max}(\rho_A) &= \log \mathrm{rank}(\rho_A) \ . \end{aligned}$$

The classical analogue

The above definitions can be specialized canonically to classical probability distributions.[4] More precisely, for $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $Q_Y \in \mathcal{P}(\mathcal{Y})$, we have

$$\begin{aligned} H_{\min}(P_{XY}|Q_Y) &:= H_{\min}(\rho_{XY}|\sigma_Y) \\ H_{\max}(P_{XY}|Q_Y) &:= H_{\max}(\rho_{XY}|\sigma_Y) \end{aligned}$$

where ρ_{XY} and σ_Y are the operator representations of P_{XY} and Q_Y, respectively (cf. Section 2.1.3).

Remark 3.1.4. Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $Q_Y \in \mathcal{P}(\mathcal{Y})$. Then[5]

$$\begin{aligned} H_{\min}(P_{XY}|Q_Y) &= -\log \max_{y \in \mathrm{supp}(Q_Y)} \max_{x \in \mathcal{X}} \frac{P_{XY}(x,y)}{Q_Y(y)} \\ H_{\max}(P_{XY}|Q_Y) &= \log \sum_{y \in \mathcal{Y}} Q_Y(y) \cdot \big|\mathrm{supp}(P_X^y)\big| \ , \end{aligned}$$

where P_X^y denotes the function $P_X^y : x \mapsto P_{XY}(x,y)$. In particular,

$$H_{\max}(P_{XY}|Y) = \log \max_{y \in \mathcal{Y}} \big|\mathrm{supp}(P_X^y)\big| \ .$$

---

[4] Similarly, the Shannon entropy can be seen as the classical special case of the von Neumann entropy.

[5] The *support* of a nonnegative function $f \in \mathcal{P}(\mathcal{X})$, denoted $\mathrm{supp}(f)$, is the set of values $x \in \mathcal{X}$ such that $f(x) > 0$.

3.1.2 Basic properties of min- and max-entropy

Min-entropy cannot be larger than max-entropy

The following lemma gives a relation between min- and max-entropy. It implies that, for a density operator ρ_{AB}, the min-entropy cannot be larger than the max-entropy.

Lemma 3.1.5. *Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$. Then*

$$H_{\min}(\rho_{AB}|\sigma_B) + \log \operatorname{tr}(\rho_{AB}) \leq H_{\max}(\rho_{AB}|\sigma_B) \ .$$

Proof. Let ρ^0_{AB} be the projector onto the support of ρ_{AB} and let $\lambda \geq 0$ such that $H_{\min}(\rho_{AB}|\sigma_B) = -\log \lambda$, i.e., $\lambda \cdot \mathrm{id}_A \otimes \sigma_B - \rho_{AB}$ is nonnegative. Using the fact that the trace of the product of two nonnegative operators is nonnegative (Lemma B.5.2), we have

$$\operatorname{tr}\bigl(\lambda \cdot (\mathrm{id}_A \otimes \sigma_B)\rho^0_{AB}\bigr) - \operatorname{tr}(\rho_{AB}) = \operatorname{tr}\bigl((\lambda \cdot \mathrm{id}_A \otimes \sigma_B - \rho_{AB})\rho^0_{AB}\bigr) \geq 0 \ .$$

Hence,

$$\log \operatorname{tr}\bigl((\mathrm{id}_A \otimes \sigma_B)\rho^0_{AB}\bigr) \geq \log \operatorname{tr}(\rho_{AB}) - \log \lambda \ .$$

The assertion then follows by the definition of the max-entropy and the choice of λ. □

Additivity of min- and max-entropy

The von Neumann entropy of a state which consists of two independent parts is equal to the sum of the entropies of each part, i.e., $H(\rho_A \otimes \rho_{A'}) = H(\rho_A) + H(\rho_{A'})$. This also holds for min- and max-entropy.

Lemma 3.1.6. *Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ and, similarly, $\rho_{A'B'} \in \mathcal{P}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$, $\sigma_{B'} \in \mathcal{P}(\mathcal{H}_{B'})$. Then*

$$H_{\min}(\rho_{AB} \otimes \rho_{A'B'}|\sigma_B \otimes \sigma_{B'}) = H_{\min}(\rho_{AB}|\sigma_B) + H_{\min}(\rho_{A'B'}|\sigma_{B'})$$
$$H_{\max}(\rho_{AB} \otimes \rho_{A'B'}|\sigma_B \otimes \sigma_{B'}) = H_{\max}(\rho_{AB}|\sigma_B) + H_{\max}(\rho_{A'B'}|\sigma_{B'}) \ .$$

Proof. The statement follows immediately from Definition 3.1.1. □

Strong subadditivity

The von Neumann entropy is subadditive, i.e., $H(A|BC) \leq H(A|B)$, which means that the entropy cannot increase when conditioning on an additional subsystem. This property can be generalized to min- and max-entropy.

Lemma 3.1.7. *Let $\rho_{ABC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ and $\sigma_{BC} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_C)$. Then*

$$H_{\min}(\rho_{ABC}|\sigma_{BC}) \leq H_{\min}(\rho_{AB}|\sigma_B)$$
$$H_{\max}(\rho_{ABC}|\sigma_{BC}) \leq H_{\max}(\rho_{AB}|\sigma_B) \ .$$

Note that, for min-entropy, the statement follows directly from the more general fact that the entropy cannot decrease under certain quantum operations (cf. Lemma 3.1.12).

Proof. Let $\lambda \geq 0$ such that $-\log \lambda = H_{\min}(\rho_{ABC}|\sigma_{BC})$, i.e., $\lambda \cdot \mathrm{id}_A \otimes \sigma_{BC} - \rho_{ABC}$ is nonnegative. Because the operator obtained by taking the partial trace of a nonnegative operator is nonnegative, $\lambda \cdot \mathrm{id}_A \otimes \sigma_B - \rho_{AB}$ is also nonnegative. This immediately implies $-\log \lambda \leq H_{\min}(\rho_{AB}|\sigma_B)$ and thus concludes the proof of the statement for min-entropy.

To show that the assertion also holds for max-entropy, let ρ^0_{AB} and ρ^0_{ABC} be the projectors on the support of ρ_{AB} and ρ_{ABC}, respectively. Because the support of ρ_{ABC} is contained in the tensor product of the support of ρ_{AB} and $\mathcal{H}_C$ (cf. Lemma B.4.1), the operator $\rho^0_{AB} \otimes \mathrm{id}_C - \rho^0_{ABC}$ is nonnegative. Moreover, because the trace of the product of two nonnegative operators is nonnegative (cf. Lemma B.5.2), we find

$$\begin{aligned} &\mathrm{tr}\big((\mathrm{id}_A \otimes \sigma_B)\rho^0_{AB}\big) - \mathrm{tr}\big((\mathrm{id}_A \otimes \sigma_{BC})\rho^0_{ABC}\big) \\ &\qquad = \mathrm{tr}\big((\mathrm{id}_A \otimes \sigma_{BC})(\rho^0_{AB} \otimes \mathrm{id}_C - \rho^0_{ABC})\big) \geq 0 \ . \end{aligned}$$

The assertion then follows by the definition of the max-entropy. □

Note that the strong subadditivity of the max-entropy together with Lemma 3.1.5 implies that $H_{\min}(\rho_{AB}|\sigma_B) \leq H_{\max}(\rho_A)$, for density operators ρ_{AB} and σ_B.

Conditioning on classical information

The min- and max-entropies of states which are partially classical can be expressed in terms of the min- and max-entropies of the corresponding conditional operators (see Section 2.1.3).

Lemma 3.1.8. *Let $\rho_{ABZ} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_Z)$ and $\sigma_{BZ} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_Z)$ be classical with respect to an orthonormal basis $\{|z\rangle\}_{z\in\mathcal{Z}}$ of $\mathcal{H}_Z$, and let ρ^z_{AB} and σ^z_B be the corresponding (non-normalized) conditional operators. Then*

$$\begin{aligned} H_{\min}(\rho_{ABZ}|\sigma_{BZ}) &= \inf_{z\in\mathcal{Z}} H_{\min}(\rho^z_{AB}|\sigma^z_B) \\ H_{\max}(\rho_{ABZ}|\sigma_{BZ}) &= \log \sum_{z\in\mathcal{Z}} 2^{H_{\max}(\rho^z_{AB}|\sigma^z_B)} \ . \end{aligned}$$

Proof. Because the vectors $|z\rangle$ are mutually orthogonal, the equivalence

$$\begin{aligned} &\lambda \cdot \mathrm{id}_A \otimes \sigma_{BZ} - \rho_{ABZ} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_Z) \\ &\qquad \Longleftrightarrow \ \forall z \in \mathcal{Z} : \ \lambda \cdot \mathrm{id}_A \otimes \sigma^z_B - \rho^z_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B) \end{aligned} \tag{3.3}$$

holds for any $\lambda \geq 0$. The assertion for the min-entropy then follows from the fact that the negative logarithm of the minimum λ satisfying the left hand side and the right hand side of (3.3) are equal to the quantities $H_{\min}(\rho_{ABZ}|\sigma_{BZ})$ and $\inf_{z\in\mathcal{Z}} H_{\min}(\rho^z_{AB}|\sigma^z_B)$, respectively.

To prove the statement for the max-entropy, let ρ^0_{ABZ} and $(\rho^z_{AB})^0$, for $z \in \mathcal{Z}$, be projectors onto the support of ρ_{ABZ} and ρ^z_{AB}, respectively. Because the vectors $|z\rangle$ are mutually orthogonal, we have

$$\rho^0_{ABZ} = \sum_{z\in\mathcal{Z}} (\rho^z_{AB})^0 \otimes |z\rangle\langle z| \ ,$$

and thus

$$\mathrm{tr}\big((\mathrm{id}_A \otimes \sigma_{BZ})\rho^0_{ABZ}\big) = \sum_{z\in\mathcal{Z}} \mathrm{tr}\big((\mathrm{id}_A \otimes \sigma^z_B)(\rho^z_{AB})^0\big) \ .$$

The assertion then follows by the definition of the max-entropy. □

Classical subsystems have nonnegative min-entropy

Similarly to the conditional von Neumann entropy, the min- and max-entropies of entangled systems can generally be negative. This is, however, not the case for the entropy of a classical subsystem. Lemma 3.1.9 below implies that

$$H_{\min}(\rho_{XC}|\rho_C) \geq 0 \ ,$$

for any density operator ρ_{XC} which is classical on the first subsystem[6]. By Lemma 3.1.5, the same holds for max-entropy.

Lemma 3.1.9. *Let $\rho_{XBC} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ be classical on $\mathcal{H}_X$ and let $\sigma_C \in \mathcal{P}(\mathcal{H}_C)$. Then*

$$H_{\min}(\rho_{XBC}|\sigma_C) \geq H_{\min}(\rho_{BC}|\sigma_C) \ .$$

Proof. Let $\lambda \geq 0$ such that $-\log\lambda = H_{\min}(\rho_{BC}|\sigma_C)$. Because ρ_{XBC} is classical on $\mathcal{H}_X$, there exists an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ and a family $\{\rho^x_{BC}\}_{x\in\mathcal{X}}$ of operators on $\mathcal{H}_B \otimes \mathcal{H}_C$ such that $\rho_{XBC} = \sum_{x\in\mathcal{X}} |x\rangle\langle x| \otimes \rho^x_{BC}$. By the definition of λ, the operator

$$\lambda \cdot \mathrm{id}_B \otimes \sigma_C - \sum_{x\in\mathcal{X}} \rho^x_{BC} = \lambda \cdot \mathrm{id}_B \otimes \sigma_C - \rho_{BC}$$

is nonnegative. Hence, for any $x \in \mathcal{X}$, the operator $\lambda \cdot \mathrm{id}_B \otimes \sigma_C - \rho^x_{BC}$ must also be nonnegative. This implies that the operator

$$\lambda \cdot \mathrm{id}_{XB} \otimes \sigma_C - \rho_{XBC} = \sum_{x\in\mathcal{X}} \lambda \cdot |x\rangle\langle x| \otimes \mathrm{id}_B \otimes \sigma_C - |x\rangle\langle x| \otimes \rho^x_{BC}$$

is nonnegative as well. We thus have $-\log\lambda \leq H_{\min}(\rho_{XBC}|\sigma_C)$, from which the assertion follows. □

[6] To see this, let $\mathcal{H}_B$ be the trivial space $\mathbb{C}$ and set $\sigma_C = \rho_C$.

3.1.3 Chain rules for min-entropy

The chain rule for the von Neumann entropy reads $H(AB|C) = H(A|BC) + H(B|C)$. In particular, since $H(B|C)$ cannot be larger than $H(B)$, we have $H(AB|C) \leq H(A|BC) + H(B)$. The following lemma implies that a similar statement holds for min-entropy, namely,

$$H_{\min}(\rho_{ABC}|C) \leq H_{\min}(\rho_{ABC}|BC) + H_{\max}(\rho_B) .$$

Lemma 3.1.10. *Let $\rho_{ABC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, $\sigma_C \in \mathcal{P}(\mathcal{H}_C)$, and let $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ be the fully mixed state on the support of ρ_B. Then*

$$H_{\min}(\rho_{ABC}|\sigma_C) = H_{\min}(\rho_{ABC}|\sigma_B \otimes \sigma_C) + H_{\max}(\rho_B) .$$

Proof. Let $\mathcal{H}_{B'} := \mathrm{supp}(\rho_B)$ be the support of ρ_B and let $\lambda \geq 0$. The operator σ_B can then be written as $\sigma_B = \frac{1}{\mathrm{rank}(\rho_B)} \mathrm{id}_{B'}$, where $\mathrm{id}_{B'}$ is the identity on $\mathcal{H}_{B'}$. Hence, because the support of ρ_{ABC} is contained in $\mathcal{H}_A \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_C$ (cf. Lemma B.4.1), the operator $\lambda \cdot \mathrm{id}_A \otimes \sigma_B \otimes \sigma_C - \rho_{ABC}$ is nonnegative if and only if the operator $\lambda \cdot \frac{1}{\mathrm{rank}(\rho_B)} \cdot \mathrm{id}_A \otimes \mathrm{id}_B \otimes \sigma_C - \rho_{ABC}$ is nonnegative. The assertion thus follows from the definition of the min-entropy and the fact that $H_{\max}(\rho_B) = \log \mathrm{rank}(\rho_B)$. □

Data processing

Let A, Y, and C be random variables such that $A \leftrightarrow Y \leftrightarrow C$ is a *Markov chain*, i.e., the conditional probability distributions $P_{AC|Y=y}$ have product form $P_{A|Y=y} \times P_{C|Y=y}$. The uncertainty on A given Y is then equal to the uncertainty on A given Y and C, that is, in terms of Shannon entropy, $H(A|Y) = H(A|YC)$. Hence, by the chain rule, we get the equality $H(AY|C) = H(Y|C) + H(A|Y)$.

The same equality also holds for quantum states ρ_{AYC} on $\mathcal{H}_A \otimes \mathcal{H}_Y \otimes \mathcal{H}_C$ which are classical on $\mathcal{H}_Y$ and where, analogously to the Markov condition, the conditional density operators $\bar{\rho}^y_{AC}$ have product form, i.e., $\bar{\rho}^y_{AC} = \bar{\rho}^y_A \otimes \bar{\rho}^y_C$. The following lemma generalizes this statement to min-entropy.

Lemma 3.1.11. *Let $\rho_{AYC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_Y \otimes \mathcal{H}_C)$ be classical with respect to an orthonormal basis $\{|y\rangle\}_{y \in \mathcal{Y}}$ of $\mathcal{H}_Y$ such that the corresponding conditional operators ρ^y_{AC}, for any $y \in \mathcal{Y}$, have product form and let $\sigma_C \in \mathcal{P}(\mathcal{H}_C)$. Then*

$$H_{\min}(\rho_{AYC}|\sigma_C) \geq H_{\min}(\rho_{YC}|\sigma_C) + H_{\min}(\rho_{AY}|\rho_Y) .$$

Proof. For any $y \in \mathcal{Y}$, let $p_y := \mathrm{tr}(\rho^y_{AC})$ and let $\bar{\rho}^y_{AC} := \frac{1}{p_y} \rho^y_{AC}$ be the normalization of ρ^y_{AC}. The operator ρ_{AYC} can then be written as

$$\rho_{AYC} = \sum_{y \in \mathcal{Y}} p_y \cdot \bar{\rho}^y_A \otimes |y\rangle\langle y| \otimes \bar{\rho}^y_C .$$

Let $\lambda, \lambda' \geq 0$ such that $-\log \lambda = H_{\min}(\rho_{YC}|\sigma_C)$, $-\log \lambda' = H_{\min}(\rho_{AY}|\rho_Y)$. Because the vectors $|y\rangle$ are mutually orthogonal, it follows immediately from the definition of the min-entropy that the operators $\lambda \cdot \sigma_C - p_y \cdot \bar{\rho}^y_C$ and $\lambda' \cdot \mathrm{id}_A - \bar{\rho}^y_A$ are nonnegative, for any $y \in \mathcal{Y}$. Consequently, the operator

$$\lambda \cdot \lambda' \cdot \mathrm{id}_A \otimes \mathrm{id}_Y \otimes \sigma_C - \rho_{AYC} = \sum_{y \in \mathcal{Y}} \lambda \cdot \lambda' \cdot \mathrm{id}_A \otimes |y\rangle\langle y| \otimes \sigma_C - p_y \cdot \bar{\rho}^y_A \otimes |y\rangle\langle y| \otimes \bar{\rho}^y_C$$

is nonnegative as well. This implies

$$H_{\min}(\rho_{AYC}|\sigma_C) \geq -\log(\lambda \cdot \lambda') = -\log \lambda - \log \lambda'$$

from which the assertion follows by the definition of λ and λ'. □

3.1.4 Quantum operations can only increase min-entropy

The min-entropy can only increase when applying quantum operations. Because the partial trace is a quantum operation, this general statement also implies the first assertion of Lemma 3.1.7 (strong subadditivity).

Lemma 3.1.12. *Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$, $\tilde{\sigma}_{B'} \in \mathcal{P}(\mathcal{H}_{B'})$ and let $\mathcal{E}$ be a CPM from $\mathcal{H}_A \otimes \mathcal{H}_B$ to $\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$ such that $\mathrm{id}_{A'} \otimes \tilde{\sigma}_{B'} - \mathcal{E}(\mathrm{id}_A \otimes \sigma_B)$ is nonnegative. Then, for $\tilde{\rho}_{A'B'} := \mathcal{E}(\rho_{AB})$,*

$$H_{\min}(\tilde{\rho}_{A'B'}|\tilde{\sigma}_{B'}) \geq H_{\min}(\rho_{AB}|\sigma_B) \ .$$

Proof. Let $\lambda \geq 0$ such that $-\log \lambda = H_{\min}(\rho_{AB}|\sigma_B)$, that is, the operator $\lambda \cdot \mathrm{id}_A \otimes \sigma_B - \rho_{AB}$ is nonnegative. Because $\mathcal{E}$ is a quantum operation, the operator $\lambda \cdot \mathcal{E}(\mathrm{id}_A \otimes \sigma_B) - \mathcal{E}(\rho_{AB})$ is also nonnegative. Combining this with the assumption that $\mathrm{id}_{A'} \otimes \tilde{\sigma}_{B'} - \mathcal{E}(\mathrm{id}_A \otimes \sigma_B)$ is nonnegative, we conclude that the operator

$$\lambda \cdot \mathrm{id}_{A'} \otimes \tilde{\sigma}_{B'} - \tilde{\rho}_{A'B'} = \lambda\big(\mathrm{id}_{A'} \otimes \tilde{\sigma}_{B'} - \mathcal{E}(\mathrm{id}_A \otimes \sigma_B)\big) + \lambda \cdot \mathcal{E}(\mathrm{id}_A \otimes \sigma_B) - \tilde{\rho}_{A'B'}$$

is also nonnegative. The assertion then follows by the definition of the min-entropy. □

3.1.5 Min-entropy of superpositions

Let $\{|x\rangle\}_{x \in \mathcal{X}}$ be an orthonormal basis on $\mathcal{H}_X$, let $\{|\psi^x\rangle\}_{x \in \mathcal{X}}$ be a family of vectors on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$, and define

$$\rho_{ABE} := |\psi\rangle\langle\psi| \qquad \text{where } |\psi\rangle := \sum_{x \in \mathcal{X}} |\psi^x\rangle \tag{3.4}$$

$$\tilde{\rho}_{ABEX} := \sum_{x \in \mathcal{X}} |\psi^x\rangle\langle\psi^x| \otimes |x\rangle\langle x| \ . \tag{3.5}$$

Note that, if the states $|\psi^x\rangle$ are orthogonal then $\tilde{\rho}_{ABEX}$ can be seen as the state resulting from an orthogonal measurement of ρ_{ABE} with respect to the projectors along $|\psi^x\rangle$. While ρ_{ABE} is a *superposition* (linear combination) of vectors $|\psi^x\rangle$, $\tilde{\rho}_{ABE}$ is a *mixture* of vectors $|\psi^x\rangle$. The following lemma gives a lower bound on the min-entropy of ρ_{ABE} in terms of the min-entropy of $\tilde{\rho}_{ABE}$.

Lemma 3.1.13. *Let ρ_{ABE} and $\tilde{\rho}_{ABEX}$ be defined by* (3.4) *and* (3.5)*, respectively, and let $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$. Then*

$$H_{\min}(\rho_{AB}|\sigma_B) \geq H_{\min}(\tilde{\rho}_{AB}|\sigma_B) - H_{\max}(\tilde{\rho}_X) \ .$$

Proof. Assume without loss of generality that, for all $x \in \mathcal{X}$, $|\psi^x\rangle$ is not the zero vector. This implies $H_{\max}(\tilde{\rho}_X) = \log|\mathcal{X}|$. Moreover, let $\lambda \geq 0$ such that $-\log\lambda = H_{\min}(\tilde{\rho}_{AB}|\sigma_B)$. It then suffices to show that the operator

$$\lambda \cdot |\mathcal{X}| \cdot \mathrm{id}_A \otimes \sigma_B - \rho_{AB} \tag{3.6}$$

is nonnegative.

Let $|\theta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. By linearity, we have

$$\langle\theta|\rho_{AB}|\theta\rangle = \langle\theta|\mathrm{tr}_E(|\psi\rangle\langle\psi|)|\theta\rangle = \sum_{x,x' \in \mathcal{X}} \langle\theta|\mathrm{tr}_E(|\psi^x\rangle\langle\psi^{x'}|)|\theta\rangle \ . \tag{3.7}$$

Let $\{|z\rangle\}_{z\in\mathcal{Z}}$ be an orthonormal basis of $\mathcal{H}_E$ and define $|\theta, z\rangle := |\theta\rangle \otimes |z\rangle$. Then, by the Cauchy-Schwartz inequality, for any $x, x' \in \mathcal{X}$,

$$\begin{aligned}
|\langle\theta|\mathrm{tr}_E(|\psi^x\rangle\langle\psi^{x'}|)|\theta\rangle| &= \Big|\sum_{z\in\mathcal{Z}} \langle\theta, z|\psi^x\rangle\langle\psi^{x'}|\theta, z\rangle\Big| \\
&\leq \sqrt{\sum_{z\in\mathcal{Z}} \langle\theta, z|\psi^x\rangle\langle\psi^x|\theta, z\rangle}\sqrt{\sum_{z\in\mathcal{Z}} \langle\theta, z|\psi^{x'}\rangle\langle\psi^{x'}|\theta, z\rangle} \\
&= \sqrt{\langle\theta|\mathrm{tr}_E(|\psi^x\rangle\langle\psi^x|)|\theta\rangle\langle\theta|\mathrm{tr}_E(|\psi^{x'}\rangle\langle\psi^{x'}|)|\theta\rangle} \ .
\end{aligned}$$

Combining this with (3.7) and using Jensen's inequality, we find

$$\begin{aligned}
\langle\theta|\rho_{AB}|\theta\rangle &\leq \sum_{x,x'\in\mathcal{X}} \sqrt{\langle\theta|\mathrm{tr}_E(|\psi^x\rangle\langle\psi^x|)|\theta\rangle\langle\theta|\mathrm{tr}_E(|\psi^{x'}\rangle\langle\psi^{x'}|)|\theta\rangle} \\
&\leq |\mathcal{X}|\sqrt{\sum_{x,x'\in\mathcal{X}} \langle\theta|\mathrm{tr}_E(|\psi^x\rangle\langle\psi^x|)|\theta\rangle\langle\theta|\mathrm{tr}_E(|\psi^{x'}\rangle\langle\psi^{x'}|)|\theta\rangle} \\
&= |\mathcal{X}| \sum_{x\in\mathcal{X}} \langle\theta|\mathrm{tr}_E(|\psi^x\rangle\langle\psi^x|)|\theta\rangle \\
&= |\mathcal{X}| \cdot \langle\theta|\tilde{\rho}_{AB}|\theta\rangle \ .
\end{aligned}$$

By the choice of λ, the operator $\lambda \cdot \mathrm{id}_A \otimes \sigma_B - \tilde{\rho}_{AB}$ is nonnegative. Hence $\langle\theta|\tilde{\rho}_{AB}|\theta\rangle \leq \lambda\langle\theta|\mathrm{id}_A \otimes \sigma_B|\theta\rangle$ and thus, by the above inequality, $\langle\theta|\rho_{AB}|\theta\rangle \leq \lambda \cdot |\mathcal{X}| \cdot \langle\theta|\mathrm{id}_A \otimes \sigma_B|\theta\rangle$. Because this is true for any vector $|\theta\rangle$, we conclude that the operator defined by (3.6) is nonnegative. □

Lemma 3.1.14. *Let ρ_{ABE}, $\tilde{\rho}_{ABEX}$ be defined by* (3.4) *and* (3.5), *respectively, and let $\sigma_{BX} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_{\mathcal{X}})$. Then*

$$H_{\min}(\rho_{AB}|\sigma_B) \geq H_{\min}(\tilde{\rho}_{ABX}|\sigma_{BX}) - H_{\max}(\tilde{\rho}_X) \ .$$

Proof. The assertion follows from Lemma 3.1.13 together with Lemma 3.1.7. □

3.2 Smooth min- and max-entropy

The min-entropy and the max-entropy, as defined in the previous section, are discontinuous in the sense that a slight modification of the system's state might have a large impact on its entropy. To illustrate this, consider for example a classical random variable X on the set $\{0, \ldots, n-1\}$ which takes the values 0 and 1 with probability almost one half, i.e., $P_X(0) = P_X(1) = \frac{1-\varepsilon}{2}$, for some small $\varepsilon > 0$, whereas the other values have equal probabilities, i.e., $P_X(x) = \frac{\varepsilon}{n-2}$, for all $x > 1$. Then, by the definition of the max-entropy, $H_{\max}(P_X) = \log n$. On the other hand, if we slightly change the probability distribution P_X to some probability distribution $\bar{P}_X$ such that $\bar{P}_X(x) = 0$, for all $x > 1$, then $H_{\max}(\bar{P}_X) = 1$. In particular, for n large, $H_{\max}(P_X) \gg H_{\max}(\bar{P}_X)$, while $\|P_X - \bar{P}_X\|_1 \leq \varepsilon$.

We will see later (cf. Section 6.3) that the max-entropy $H_{\max}(P_X)$ can be interpreted as the minimum number of bits needed to encode X in such a way that its value can be recovered from the encoding without errors. The above example is consistent with this interpretation. Indeed, while we need at least $\log n$ bits to store a value X distributed according to P_X, one single bit is sufficient to store a value distributed according to $\bar{P}_X$. However, for most applications, we allow some small error probability. For example, we might want to encode X in such a way that its value can be recovered with probability $1 - \varepsilon$. Obviously, in this case, one single bit is sufficient to store X even if it is distributed according to P_X.

The example illustrates that, given some probability distribution P_X, one might be interested in the maximum (or minimum) entropy of any distribution $\bar{P}_X$ which is close to P_X. This idea is captured by the notion of smooth min- and max-entropy.

3.2.1 Definition of smooth min- and max-entropy

The definition of smooth min- and max-entropy is based on the "non-smooth" version (Definition 3.1.1).

Definition 3.2.1. Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$, and $\varepsilon \geq 0$. The *ε-smooth min-entropy* and the *ε-smooth max-entropy of ρ_{AB} relative to σ_B*

are

$$H^{\varepsilon}_{\min}(\rho_{AB}|\sigma_B) := \sup_{\bar{\rho}_{AB}} H_{\min}(\bar{\rho}_{AB}|\sigma_B)$$
$$H^{\varepsilon}_{\max}(\rho_{AB}|\sigma_B) := \inf_{\bar{\rho}_{AB}} H_{\max}(\bar{\rho}_{AB}|\sigma_B) ,$$

where the supremum and infimum ranges over the set $\mathcal{B}^{\varepsilon}(\rho_{AB})$ of all operators $\bar{\rho}_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\|\bar{\rho}_{AB} - \rho_{AB}\|_1 \leq \mathrm{tr}(\rho_{AB}) \cdot \varepsilon$ and $\mathrm{tr}(\bar{\rho}_{AB}) \leq \mathrm{tr}(\rho_{AB})$.

Definition 3.2.2. Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and let $\varepsilon \geq 0$. The *ε-smooth min-entropy* and the *ε-smooth max-entropy of ρ_{AB} given $\mathcal{H}_B$* are

$$H^{\varepsilon}_{\min}(\rho_{AB}|B) := \sup_{\sigma_B} H^{\varepsilon}_{\min}(\rho_{AB}|\sigma_B)$$
$$H^{\varepsilon}_{\max}(\rho_{AB}|B) := \sup_{\sigma_B} H^{\varepsilon}_{\max}(\rho_{AB}|\sigma_B) ,$$

where the supremum ranges over all $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ with $\mathrm{tr}(\sigma_B) = 1$.

Note that, similar to the description in Section 3.1, these definitions can be specialized to classical probability distributions.

Evaluating the suprema and infima

Remark 3.2.3. If the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ has finite dimension, then the set of operators $\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\mathcal{H}_A \otimes \mathcal{H}_B)$ as well as the set of operators $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ with $\mathrm{tr}(\sigma_B) = 1$ is compact. Hence, the infima and suprema in the above definitions can be replaced by minima and maxima, respectively.

Remark 3.2.4. The supremum in the definition of the smooth min-entropy $H^{\varepsilon}_{\min}(\rho_{AB}|\sigma_B)$ (Definition 3.2.1) can be restricted to the set of operators $\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})$ with $\mathrm{supp}(\bar{\rho}_{AB}) \subseteq \mathrm{supp}(\rho_A) \otimes \mathrm{supp}(\sigma_B)$.

Additionally, to compute $H^{\varepsilon}_{\min}(\rho_{ABZ}|\sigma_{BZ})$ where ρ_{ABZ} and σ_{BZ} are classical with respect to an orthonormal basis $\{|z\rangle\}_{z\in\mathcal{Z}}$ on a subsystem $\mathcal{H}_Z$, it is sufficient to take the supremum over operators $\bar{\rho}_{ABZ} \in \mathcal{B}^{\varepsilon}(\rho_{ABZ})$ which are classical with respect to $\{|z\rangle\}_{z\in\mathcal{Z}}$.

Similarly, to compute $H^{\varepsilon}_{\min}(\rho_{XAB}|\sigma_B)$ where ρ_{XAB} is classical on a subsystem $\mathcal{H}_X$, the supremum can be restricted to states $\bar{\rho}_{XAB} \in \mathcal{B}^{\varepsilon}(\rho_{XAB})$ which are classical on $\mathcal{H}_X$.

Proof. For the first statement, we show that any operator $\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})$ can be transformed to an operator $\mathcal{E}(\bar{\rho}_{AB}) \in \mathcal{B}^{\varepsilon}(\rho_{AB})$ which has at least the same amount of min-entropy as $\bar{\rho}_{AB}$ and, additionally, has support on $\mathrm{supp}(\rho_A) \otimes \mathrm{supp}(\sigma_B)$.

Let $\mathcal{E}$ be the operation on $\mathcal{H}_A \otimes \mathcal{H}_B$ defined by

$$\mathcal{E}(\bar{\rho}_{AB}) := (\rho^0_A \otimes \mathrm{id}_B)\bar{\rho}_{AB}(\rho^0_A \otimes \mathrm{id}_B) .$$

Because the operator $\mathrm{id}_A \otimes \sigma_B - \mathcal{E}(\mathrm{id}_A \otimes \sigma_B)$ is nonnegative, Lemma 3.1.12 implies that the min-entropy can only increase under the action of $\mathcal{E}$. Moreover, $\mathrm{supp}(\rho_{AB}) \subseteq \mathrm{supp}(\rho_A) \otimes \mathcal{H}_B$ (cf. Lemma B.4.1) and thus $\mathcal{E}(\rho_{AB}) = \rho_{AB}$. Because $\mathcal{E}$ is a projection, the L_1-distance cannot increase under the action of $\mathcal{E}$ (cf. Lemma A.2.1), i.e.,

$$\big\|\mathcal{E}(\bar{\rho}_{AB}) - \rho_{AB}\big\|_1 = \big\|\mathcal{E}(\bar{\rho}_{AB} - \rho_{AB})\big\|_1 \leq \big\|\bar{\rho}_{AB} - \rho_{AB}\big\|_1 \leq \mathrm{tr}(\rho_{AB}) \cdot \varepsilon .$$

We thus have $\bar{\rho}_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})$. The assertion then follows because we can assume that $\mathrm{supp}(\bar{\rho}_{AB})$ is contained in $\mathcal{H}_A \otimes \mathrm{supp}(\sigma_B)$ (otherwise, the min-entropy is arbitrarily negative and the statement is trivial) and thus $\mathrm{supp}(\mathcal{E}(\bar{\rho}_{AB})) \subseteq \mathrm{supp}(\rho_A) \otimes \mathrm{supp}(\sigma_B)$.

The statements for ρ_{ABZ} and ρ_{XAB} are proven similarly. □

Remark 3.2.5. Let $\rho_{ABZ} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_Z)$ be classical with respect to an orthonormal basis $\{|z\rangle\}_{z\in\mathcal{Z}}$ of $\mathcal{H}_Z$. Then the supremum in the definition of the min-entropy $H^\varepsilon_{\min}(\rho_{ABZ}|BZ)$ can be restricted to operators $\sigma_{BZ} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_Z)$ which are classical with respect to $\{|z\rangle\}_{z\in\mathcal{Z}}$.

Proof. We show that for any $\bar{\rho}'_{ABZ} \in \mathcal{B}^\varepsilon(\rho_{ABZ})$ and $\sigma'_{BZ} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_Z)$ with $\mathrm{tr}(\sigma'_{BZ}) = 1$ there exists $\bar{\rho}_{ABZ} \in \mathcal{B}^\varepsilon(\rho_{ABZ})$ and $\sigma_{BZ} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_Z)$ with $\mathrm{tr}(\sigma_{BZ}) = 1$ such that σ_{BZ} is classical with respect to $\{|z\rangle\}_{z\in\mathcal{Z}}$ and $H_{\min}(\bar{\rho}_{ABZ}|\sigma_{BZ}) \geq H_{\min}(\bar{\rho}'_{ABZ}|\sigma'_{BZ})$.

Let thus $\bar{\rho}'_{ABZ} \in \mathcal{B}^\varepsilon(\rho_{ABZ})$ and $\sigma'_{BZ} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_Z)$ be fixed. Define $\bar{\rho}_{ABZ} := (\mathrm{id}_{AB} \otimes \mathcal{E}_Z)(\bar{\rho}'_{ABZ})$ and $\sigma_{BZ} := (\mathrm{id}_B \otimes \mathcal{E}_Z)(\sigma'_{BZ})$ where $\mathcal{E}_Z$ is the projective measurement operation on $\mathcal{H}_Z$, i.e.,

$$\mathcal{E}_Z(\rho) := \sum_{z\in\mathcal{Z}} |z\rangle\langle z|\rho|z\rangle\langle z| .$$

Note that σ_{BZ} is classical with respect to $\{|z\rangle\}_{z\in\mathcal{Z}}$ and, because $\mathcal{E}_Z$ is trace-preserving, $\mathrm{tr}(\sigma_{BZ}) = \mathrm{tr}(\sigma'_{BZ}) = 1$. Similarly, $\mathrm{tr}(\bar{\rho}_{ABZ}) = \mathrm{tr}(\bar{\rho}'_{ABZ})$. Moreover, because $(\mathrm{id}_{AB} \otimes \mathcal{E}_Z)(\rho_{ABZ}) = \rho_{ABZ}$ and because the distance can only decrease when applying $\mathrm{id}_{AB} \otimes \mathcal{E}_Z$ (cf. Lemma A.2.1), we have

$$\|\bar{\rho}_{ABZ} - \rho_{ABZ}\|_1 \leq \|\bar{\rho}'_{ABZ} - \rho_{ABZ}\|_1$$

which implies $\bar{\rho}_{ABZ} \in \mathcal{B}^\varepsilon(\rho_{ABZ})$. Finally, using Lemma 3.1.12, we find $H_{\min}(\bar{\rho}_{ABZ}|\sigma_{BZ}) \geq H_{\min}(\bar{\rho}'_{ABZ}|\sigma'_{BZ})$. □

3.2.2 Basic properties of smooth min-entropy

Superadditivity

The following is a generalization of (one direction of) Lemma 3.1.6 to smooth min-entropy.

Lemma 3.2.6. *Let* $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$*,* $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ *and, similarly,* $\rho_{A'B'} \in \mathcal{P}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$*,* $\sigma_{B'} \in \mathcal{P}(\mathcal{H}_{B'})$*, and let* $\varepsilon, \varepsilon' \geq 0$*. Then*

$$H_{\min}^{\varepsilon+\varepsilon'}(\rho_{AB} \otimes \rho_{A'B'}|\sigma_B \otimes \sigma_{B'}) \geq H_{\min}^{\varepsilon}(\rho_{AB}|\sigma_B) + H_{\min}^{\varepsilon'}(\rho_{A'B'}|\sigma_{B'}) \ .$$

Proof. For any $\nu > 0$, there exist $\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})$ and $\bar{\rho}_{A'B'} \in \mathcal{B}^{\varepsilon'}(\rho_{A'B'})$ such that

$$\begin{aligned} H_{\min}(\bar{\rho}_{AB}|\sigma_B) &> H_{\min}^{\varepsilon}(\rho_{AB}|\sigma_B) - \nu \\ H_{\min}(\bar{\rho}_{A'B'}|\sigma_{B'}) &> H_{\min}^{\varepsilon'}(\rho_{A'B'}|\sigma_{B'}) - \nu \ . \end{aligned}$$

Hence, by Lemma 3.1.6,

$$H_{\min}(\bar{\rho}_{AB} \otimes \bar{\rho}_{A'B'}|\sigma_B \otimes \sigma_{B'}) > H_{\min}^{\varepsilon}(\rho_{AB}|\sigma_B) + H_{\min}^{\varepsilon'}(\rho_{A'B'}|\sigma_{B'}) - 2\nu \ .$$

Because this holds for any $\nu > 0$, it remains to verify that $\bar{\rho}_{AB} \otimes \bar{\rho}_{A'B'} \in \mathcal{B}^{\varepsilon+\varepsilon'}(\rho_{AB} \otimes \rho_{A'B'})$. This is however a direct consequence of the triangle inequality, i.e.,

$$\begin{aligned} &\big\|\bar{\rho}_{AB} \otimes \bar{\rho}_{A'B'} - \rho_{AB} \otimes \rho_{A'B'}\big\|_1 \\ &\qquad \leq \operatorname{tr}(\bar{\rho}_{A'B'}) \cdot \big\|\bar{\rho}_{AB} - \rho_{AB}\big\|_1 + \operatorname{tr}(\rho_{AB}) \cdot \big\|\bar{\rho}_{A'B'} - \rho_{A'B'}\big\|_1 \\ &\qquad \leq \operatorname{tr}(\rho_{AB} \otimes \rho_{A'B'})(\varepsilon + \varepsilon') \ . \end{aligned}$$

□

Strong subadditivity

The following statement is a generalization of Lemma 3.1.7 to smooth min-entropy.

Lemma 3.2.7. *Let* $\rho_{ABC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$*,* $\sigma_{BC} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_C)$*, and let* $\varepsilon \geq 0$*. Then*

$$H_{\min}^{\varepsilon}(\rho_{ABC}|\sigma_{BC}) \leq H_{\min}^{\varepsilon}(\rho_{AB}|\sigma_B) \ .$$

Proof. For any $\nu > 0$, there exists $\bar{\rho}_{ABC} \in \mathcal{B}^{\varepsilon}(\rho_{ABC})$ such that

$$H_{\min}(\bar{\rho}_{ABC}|\sigma_{BC}) \geq H_{\min}^{\varepsilon}(\rho_{ABC}|\sigma_{BC}) - \nu \ .$$

Hence, by Lemma 3.1.7, applied to the operator $\bar{\rho}_{ABC}$,

$$H_{\min}(\bar{\rho}_{AB}|\sigma_B) \geq H_{\min}^{\varepsilon}(\rho_{ABC}|\sigma_{BC}) - \nu \ .$$

Because this holds for any $\nu > 0$, it remains to show that $\bar{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})$. This is however a direct consequence of the fact that the L_1-distance cannot increase when taking the partial trace (cf. Lemma A.2.1), i.e.,

$$\|\bar{\rho}_{AB} - \rho_{AB}\|_1 \leq \|\bar{\rho}_{ABC} - \rho_{ABC}\|_1 \leq \operatorname{tr}(\rho_{ABC}) \cdot \varepsilon \ .$$

□

Conditioning on classical information

The following lemma generalizes (one direction of) Lemma 3.1.8 to smooth min-entropy.

Lemma 3.2.8. *Let $\rho_{ABZ} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_Z)$ and $\sigma_{BZ} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_Z)$ be classical with respect to an orthonormal basis $\{|z\rangle\}_{z \in \mathcal{Z}}$ of $\mathcal{H}_Z$, let ρ_{AB}^z and σ_B^z be the corresponding (non-normalized) conditional operators, and let $\varepsilon \geq 0$. Then*

$$H_{\min}^{\varepsilon}(\rho_{ABZ}|\sigma_{BZ}) \geq \inf_{z \in \mathcal{Z}} H_{\min}^{\varepsilon}(\rho_{AB}^z|\sigma_B^z) \ .$$

Proof. For any $\nu > 0$ and $z \in \mathcal{Z}$, there exists $\bar{\rho}_{AB}^z \in \mathcal{B}^{\varepsilon}(\rho_{AB}^z)$ such that

$$H_{\min}(\bar{\rho}_{AB}^z|\sigma_B^z) \geq H_{\min}^{\varepsilon}(\rho_{AB}^z|\sigma_B^z) - \nu \ .$$

Let

$$\bar{\rho}_{ABZ} := \sum_{z \in \mathcal{Z}} \bar{\rho}_{AB}^z \otimes |z\rangle\langle z| \ .$$

Using Lemma 3.1.8, we find

$$H_{\min}(\bar{\rho}_{ABZ}|\sigma_{BZ}) = \inf_{z \in \mathcal{Z}} H_{\min}(\bar{\rho}_{AB}^z|\sigma_B^z) \geq \inf_{z \in \mathcal{Z}} H_{\min}^{\varepsilon}(\rho_{AB}^z|\sigma_B^z) - \nu \ . \quad (3.8)$$

Because this holds for any value of $\nu > 0$, it suffices to verify that $\bar{\rho}_{ABZ} \in \mathcal{B}^{\varepsilon}(\rho_{ABZ})$. This is however a direct consequence of

$$\left\|\bar{\rho}_{ABZ} - \rho_{ABZ}\right\|_1 = \sum_{z \in \mathcal{Z}} \left\|\bar{\rho}_{AB}^z - \rho_{AB}^z\right\|_1 \leq \sum_{z \in \mathcal{Z}} \mathrm{tr}(\rho_{AB}^z) \cdot \varepsilon = \mathrm{tr}(\rho_{ABZ}) \cdot \varepsilon \ ,$$

where the first equality follows from Lemma A.2.2. □

3.2.3 Chain rules for smooth min-entropy

The following lemma generalizes (one direction of) Lemma 3.1.10 to smooth min-entropy.

Lemma 3.2.9. *Let $\rho_{ABC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, $\sigma_C \in \mathcal{P}(\mathcal{H}_C)$, let $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ be the fully mixed state on the support of ρ_B, and let $\varepsilon \geq 0$. Then*

$$H_{\min}^{\varepsilon}(\rho_{ABC}|\sigma_C) \leq H_{\min}^{\varepsilon}(\rho_{ABC}|\sigma_B \otimes \sigma_C) + H_{\max}(\rho_B) \ .$$

Proof. According to Remark 3.2.4, for any $\nu > 0$, there exists $\bar{\rho}_{ABC} \in \mathcal{B}^{\varepsilon}(\rho_{ABC})$ such that

$$H_{\min}(\bar{\rho}_{ABC}|\sigma_C) \geq H_{\min}^{\varepsilon}(\rho_{ABC}|\sigma_C) - \nu \quad (3.9)$$

and $\mathrm{supp}(\bar{\rho}_{ABC}) \subseteq \mathrm{supp}(\rho_{AB}) \otimes \mathcal{H}_C = \mathrm{supp}(\rho_{AB} \otimes \mathrm{id}_C)$. Hence, from Lemma B.4.2, $\mathrm{supp}(\bar{\rho}_B) \subseteq \mathrm{supp}(\rho_B)$. Consequently, the operator $\bar{\rho}_B$ is arbitrarily close to an operator whose support is equal to the support of ρ_B. By continuity, we can thus assume without loss of generality that $\mathrm{supp}(\bar{\rho}_B) = \mathrm{supp}(\rho_B)$, that is,

$$H_{\max}(\bar{\rho}_B) = H_{\max}(\rho_B) \ . \tag{3.10}$$

Moreover, since $\bar{\rho}_{ABC} \in \mathcal{B}^{\varepsilon}(\rho_{ABC})$, we have

$$H_{\min}^{\varepsilon}(\rho_{ABC}|\sigma_B \otimes \sigma_C) \geq H_{\min}(\bar{\rho}_{ABC}|\sigma_B \otimes \sigma_C) \ . \tag{3.11}$$

Finally, because σ_B is the fully mixed state on $\mathrm{supp}(\rho_B) = \mathrm{supp}(\bar{\rho}_B)$, Lemma 3.1.10, applied to the state $\bar{\rho}_{ABC}$, gives

$$H_{\min}(\bar{\rho}_{ABC}|\sigma_C) = H_{\min}(\bar{\rho}_{ABC}|\sigma_B \otimes \sigma_C) + H_{\max}(\bar{\rho}_B) \ .$$

Combining this with (3.9), (3.10), and (3.11) concludes the proof. □

Data processing

The following lemma is a generalization of Lemma 3.1.11 to smooth min-entropy.

Lemma 3.2.10. *Let $\rho_{AYC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_Y \otimes \mathcal{H}_C)$ be classical with respect to an orthonormal basis $\{|y\rangle\}_{y \in \mathcal{Y}}$ of $\mathcal{H}_Y$ such that the corresponding conditional operators ρ_{AC}^y, for any $y \in \mathcal{Y}$, have product form, let $\sigma_C \in \mathcal{P}(\mathcal{H}_C)$, and let $\varepsilon \geq 0$. Then*

$$H_{\min}^{\varepsilon}(\rho_{AYC}|\sigma_C) \geq H_{\min}^{\varepsilon}(\rho_{YC}|\sigma_C) + H_{\min}(\rho_{AY}|\rho_Y) \ .$$

Proof. For any $y \in \mathcal{Y}$, let $p_y := \mathrm{tr}(\rho_{AC}^y)$ and define $\tilde{\rho}_A^y := \frac{1}{p_y}\rho_A^y$. Because ρ_{AC}^y has product form, we have

$$\rho_{AYC} = \sum_{y \in \mathcal{Y}} \tilde{\rho}_A^y \otimes |y\rangle\langle y| \otimes \rho_C^y \ .$$

According to Remark 3.2.4, for any $\nu > 0$, there exists a nonnegative operator $\bar{\rho}_{YC} \in \mathcal{B}^{\varepsilon}(\rho_{YC})$ such that

$$H_{\min}(\bar{\rho}_{YC}|\sigma_C) \geq H_{\min}^{\varepsilon}(\rho_{YC}|\sigma_C) - \nu \tag{3.12}$$

where $\bar{\rho}_{YC}$ is classical with respect to $\{|y\rangle\}_{y \in \mathcal{Y}}$, that is, $\bar{\rho}_{YC} = \sum_{y \in \mathcal{Y}} |y\rangle\langle y| \otimes \bar{\rho}_C^y$, for some family $\{\bar{\rho}_C^y\}_{y \in \mathcal{Y}}$ of conditional operators on $\mathcal{H}_C$. Let $\bar{\rho}_{AYC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_Y \otimes \mathcal{H}_C)$ be defined by

$$\bar{\rho}_{AYC} := \sum_{y \in \mathcal{Y}} \tilde{\rho}_A^y \otimes |y\rangle\langle y| \otimes \bar{\rho}_C^y \ .$$

Because the operators $\tilde{\rho}_A^y$ are normalized, we have

$$\begin{aligned}\|\bar{\rho}_{AYC} - \rho_{AYC}\|_1 &= \sum_y \|\tilde{\rho}_A^y \otimes \bar{\rho}_C^y - \tilde{\rho}_A^y \otimes \rho_C^y\|_1 \\ &= \sum_y \|\bar{\rho}_C^y - \rho_C^y\|_1 \\ &= \|\bar{\rho}_{YC} - \rho_{YC}\|_1 ,\end{aligned}$$

where the first and the last equality follow from Lemma A.2.2. Because $\bar{\rho}_{YC} \in \mathcal{B}^\varepsilon(\rho_{YC})$, this implies $\bar{\rho}_{AYC} \in \mathcal{B}^\varepsilon(\rho_{AYC})$ and thus

$$H_{\min}^\varepsilon(\rho_{AYC}|\sigma_C) \geq H_{\min}(\bar{\rho}_{AYC}|\sigma_C) . \tag{3.13}$$

Moreover, using Lemma 3.1.8 and the fact that, for any $y \in \mathcal{Y}$, the operators $\tilde{\rho}_A^y$ and ρ_A^y only differ by a factor p_y, we have

$$\begin{aligned}H_{\min}(\bar{\rho}_{AY}|\bar{\rho}_Y) &= \inf_{y\in\mathcal{Y}} H_{\min}(\tilde{\rho}_A^y|\mathrm{tr}(\tilde{\rho}_A^y)) \\ &= \inf_{y\in\mathcal{Y}} H_{\min}(\rho_A^y|\mathrm{tr}(\rho_A^y)) \\ &= H_{\min}(\rho_{AY}|\rho_Y) .\end{aligned} \tag{3.14}$$

Finally, applying Lemma 3.1.11 to the state $\bar{\rho}_{AYC}$ gives

$$H_{\min}(\bar{\rho}_{AYC}|\sigma_C) \geq H_{\min}(\bar{\rho}_{YC}|\sigma_C) + H_{\min}(\bar{\rho}_{AY}|\bar{\rho}_Y) .$$

Combining this with (3.12), (3.13), and (3.14) concludes the proof. □

3.2.4 Smooth min-entropy of superpositions

The following statement generalizes Lemma 3.1.14.

Lemma 3.2.11. *Let ρ_{ABE}, $\tilde{\rho}_{ABEX}$ be defined by* (3.4) *and* (3.5)*, respectively, for mutually orthogonal vectors $|\psi^x\rangle$, let $\sigma_{BX} \in \mathcal{P}(\mathcal{H}_B \otimes \mathcal{H}_\mathcal{X})$, and let $\varepsilon \geq 0$. Then*

$$H_{\min}^\varepsilon(\rho_{AB}|\sigma_B) \geq H_{\min}^{\tilde{\varepsilon}}(\tilde{\rho}_{ABX}|\sigma_{BX}) - H_{\max}(\tilde{\rho}_X) ,$$

where $\tilde{\varepsilon} = \frac{\varepsilon^2}{6|\mathcal{X}|}$.

Proof. By Remark 3.2.4, for any $\nu > 0$, there exists an operator $\bar{\tilde{\rho}}_{ABX} \in \mathcal{B}^{\tilde{\varepsilon}}(\tilde{\rho}_{ABX})$ which is classical with respect to the basis $\{|x\rangle\}_{x\in\mathcal{X}}$ such that

$$H_{\min}(\bar{\tilde{\rho}}_{ABX}|\sigma_{BX}) \geq H_{\min}^{\tilde{\varepsilon}}(\tilde{\rho}_{ABX}|\sigma_{BX}) - \nu . \tag{3.15}$$

Let $\{\bar{\tilde{\rho}}_{AB}^x\}_{x\in\mathcal{X}}$ be the family of conditional operators defined by $\bar{\tilde{\rho}}_{ABX}$ and $\{|x\rangle\}_{x\in\mathcal{X}}$, i.e., $\bar{\tilde{\rho}}_{ABX} = \sum_{x\in\mathcal{X}} \bar{\tilde{\rho}}_{AB}^x \otimes |x\rangle\langle x|$. According to Lemma A.2.7, for any $x \in \mathcal{X}$, there exists a purification $|\bar{\psi}^x\rangle\langle\bar{\psi}^x|$ of $\bar{\tilde{\rho}}_{AB}^x$ such that

$$\||\psi^x\rangle - |\bar{\psi}^x\rangle\| \leq \sqrt{\|\tilde{\rho}_{AB}^x - \bar{\tilde{\rho}}_{AB}^x\|_1} .$$

Let $|\bar{\psi}\rangle := \sum_{x \in \mathcal{X}} |\bar{\psi}^x\rangle$ and define $\bar{\rho}_{ABE} := |\bar{\psi}\rangle\langle\bar{\psi}|$. By the triangle inequality, we find

$$\big\||\psi\rangle - |\bar{\psi}\rangle\big\| \leq \sum_{x \in \mathcal{X}} \big\||\psi^x\rangle - |\bar{\psi}^x\rangle\big\| \leq \sum_{x \in \mathcal{X}} \sqrt{\big\|\tilde{\rho}^x_{AB} - \bar{\tilde{\rho}}^x_{AB}\big\|_1} \ .$$

Hence, with Jensen's inequality,

$$\begin{aligned}\big\||\psi\rangle - |\bar{\psi}\rangle\big\| &\leq \sqrt{|\mathcal{X}| \sum_{x \in \mathcal{X}} \big\|\tilde{\rho}^x_{AB} - \bar{\tilde{\rho}}^x_{AB}\big\|_1} \\ &= \sqrt{|\mathcal{X}| \cdot \big\|\tilde{\rho}_{ABX} - \bar{\tilde{\rho}}_{ABX}\big\|_1} \ ,\end{aligned}$$

where the equality follows from Lemma A.2.2. Because the vectors $|\psi^x\rangle$ are orthogonal, we have $\mathrm{tr}(\tilde{\rho}_{ABX}) = \mathrm{tr}(\rho_{AB})$. Consequently, since $\bar{\tilde{\rho}}_{ABX} \in \mathcal{B}^{\tilde{\varepsilon}}(\tilde{\rho}_{ABX})$, we obtain

$$\big\||\psi\rangle - |\bar{\psi}\rangle\big\| \leq \sqrt{|\mathcal{X}| \cdot \tilde{\varepsilon} \cdot \mathrm{tr}(\tilde{\rho}_{ABX})} = \sqrt{|\mathcal{X}| \cdot \tilde{\varepsilon} \cdot \mathrm{tr}(\rho_{AB})} \ . \tag{3.16}$$

Assume without loss of generality that $|\mathcal{X}| \cdot \tilde{\varepsilon} \leq \frac{1}{6}$ (otherwise, the assertion is trivial). Then, because $\sqrt{\mathrm{tr}(\rho_{AB})} = \||\psi\rangle\|$, we have

$$\begin{aligned}\||\psi\rangle\| + \||\bar{\psi}\rangle\| &\leq 2\||\psi\rangle\| + \||\psi\rangle - |\bar{\psi}\rangle\| \\ &\leq 2\sqrt{\mathrm{tr}(\rho_{AB})} + \sqrt{\frac{1}{6}\mathrm{tr}(\rho_{AB})} < \sqrt{6\,\mathrm{tr}(\rho_{AB})} \ .\end{aligned}$$

and thus, by Lemma A.2.5,

$$\big\|\rho_{AB} - \bar{\rho}_{AB}\big\|_1 \leq \sqrt{6\,\mathrm{tr}(\rho_{AB})} \cdot \big\||\psi\rangle - |\bar{\psi}\rangle\big\| \leq \mathrm{tr}(\rho_{AB}) \cdot \varepsilon \ ,$$

where the last inequality follows from (3.16). This implies

$$H^{\varepsilon}_{\min}(\rho_{AB}|\sigma_B) \geq H_{\min}(\bar{\rho}_{AB}|\sigma_B) \ . \tag{3.17}$$

Note that $\bar{\tilde{\rho}}_{ABX}$ can be seen as the operator obtained by taking the partial trace of

$$\bar{\tilde{\rho}}_{ABEX} := \sum_{x \in \mathcal{X}} |\bar{\psi}^x\rangle\langle\bar{\psi}^x| \otimes |x\rangle\langle x| \ .$$

We can thus apply Lemma 3.1.14 to the operators $\bar{\rho}_{ABE}$ and $\bar{\tilde{\rho}}_{ABEX}$, which gives

$$H_{\min}(\bar{\rho}_{AB}|\sigma_B) \geq H_{\min}(\bar{\tilde{\rho}}_{ABX}|\sigma_{BX}) - H_{\max}(\bar{\tilde{\rho}}_X) \ .$$

Finally, because the support of $\bar{\tilde{\rho}}_X$ is contained in the support of $\tilde{\rho}_X$, we have $H_{\max}(\bar{\tilde{\rho}}_X) \leq H_{\max}(\tilde{\rho}_X)$ and thus

$$H_{\min}(\bar{\rho}_{AB}|\sigma_B) \geq H_{\min}(\bar{\tilde{\rho}}_{ABX}|\sigma_{BX}) - H_{\max}(\tilde{\rho}_X) \ .$$

Combining this with (3.17) and (3.15) concludes the proof. □

3.2.5 Smooth min-entropy calculus

The properties proven so far are formulated in terms of the smooth min-entropy $H(\rho_{AB}|\sigma_B)$ relative to an operator σ_B (Definition 3.2.1). The following theorem translates these statements to conditional smooth min-entropy $H(\rho_{AB}|B)$ (Definition 3.2.2).

Theorem 3.2.12. *Let $\varepsilon, \varepsilon' \geq 0$. Then the following inequalities hold:*

- *(Super-)additivity:*

$$H_{\min}^{\varepsilon+\varepsilon'}(\rho_{AB} \otimes \rho_{A'B'}|BB') \geq H_{\min}^{\varepsilon}(\rho_{AB}|B) + H_{\min}^{\varepsilon'}(\rho_{A'B'}|B') \ , \tag{3.18}$$

  for $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $\rho_{A'B'} \in \mathcal{P}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$.

- *Strong subadditivity:*

$$H_{\min}^{\varepsilon}(\rho_{ABC}|BC) \leq H_{\min}^{\varepsilon}(\rho_{AB}|B) \ , \tag{3.19}$$

  for $\rho_{ABC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$.

- *Conditioning on classical information:*

$$H_{\min}^{\varepsilon}(\rho_{ABZ}|BZ) \geq \inf_{z \in \mathcal{Z}} H_{\min}^{\varepsilon}(\bar{\rho}_{AB}^{z}|B) \ , \tag{3.20}$$

  for $\rho_{ABZ} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_Z)$ normalized and classical on $\mathcal{H}_Z$, and for normalized conditional operators $\bar{\rho}_{AB}^{z}$.

- *Chain rule:*

$$H_{\min}^{\varepsilon}(\rho_{ABC}|C) \leq H_{\min}^{\varepsilon}(\rho_{ABC}|BC) + H_{\max}(\rho_B) \ , \tag{3.21}$$

  for $\rho_{ABC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$.

- *Data processing:*

$$H_{\min}^{\varepsilon}(\rho_{AYC}|C) \geq H_{\min}^{\varepsilon}(\rho_{YC}|C) + H_{\min}(\rho_{AY}|\rho_Y) \ , \tag{3.22}$$

  for $\rho_{AYC} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_Y \otimes \mathcal{H}_C)$ classical on $\mathcal{H}_Y$ such that the conditional operators ρ_{AC}^{y} have product form.

Proof. The statements follow immediately from Lemmata 3.2.6, 3.2.7, 3.2.8, 3.2.9, and 3.2.10. □

3.3 Smooth min- and max-entropy of products

In this section, we show that the smooth min- and max-entropies of product states are asymptotically equal to the von Neumann entropy. In a first step, we consider a purely classical situation, i.e., we prove that the smooth min- and max-entropies of a sequence of independent and identically distributed random variables can be expressed in terms of Shannon entropy (which is the classical analogue of the von Neumann entropy). Then, in a second step, we generalize this statement to quantum states (Section 3.3.2).

3.3.1 The classical case

The proof of the main result of this section (Theorem 3.3.4) is based on a Chernoff style bound (Theorem 3.3.3) which is actually a variant of the asymptotic equipartition property (AEP) known from information theory (see, e.g., [CT91]). It states that, with high probability, the negative logarithm of the probability of an n-tuple of values chosen according to a product distribution P^n is close to the Shannon entropy of P^n.

Typical sequences and their probabilities

Lemma 3.3.1. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a probability distribution. Then, for any $t \in \mathbb{R}$ with $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$,*

$$\log \mathop{\mathbb{E}}_{x,y}\left[P_{X|Y}(x,y)^{-t}\right] \leq tH(X|Y) + \frac{1}{2}t^2 \log(|\mathcal{X}|+3)^2 \ ,$$

where the expectation is taken over pairs (x,y) chosen according to P_{XY}.

Proof. For any $t \in \mathbb{R}$, let r_t be the function on the open interval $(0,\infty)$ defined by

$$r_t(z) := z^t - t \ln z - 1 \ . \tag{3.23}$$

We will use several properties of this function proven in Appendix B.6.

For any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, let $p_{x,y} := P_{X|Y}(x,y)$. If $p_{x,y} > 0$ then

$$p_{x,y}^{-t} = r_t\big(\frac{1}{p_{x,y}}\big) + t \ln \frac{1}{p_{x,y}} + 1 \leq r_t\big(\frac{1}{p_{x,y}} + 3\big) + t \ln \frac{1}{p_{x,y}} + 1 \ ,$$

where the inequality holds because r_t is monotonically increasing on the interval $[1,\infty)$ (Lemma B.6.1) and $\frac{1}{p_{x,y}} = \frac{P_Y(y)}{P_{XY}(x,y)} \geq 1$. Because $\frac{1}{p_{x,y}} + 3 \in [4,\infty)$ and because r_t is concave on this interval (Lemma B.6.3 which can be applied because $t \in [-\frac{1}{2}, \frac{1}{2}]$), Jensen's inequality leads to

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y}\left[p_{x,y}^{-t}\right] &\leq \mathop{\mathbb{E}}_{x,y}\Big[r_t\big(\frac{1}{p_{x,y}} + 3\big)\Big] + t \mathop{\mathbb{E}}_{x,y}\Big[\ln \frac{1}{p_{x,y}}\Big] + 1 \\ &\leq r_t\Big(\mathop{\mathbb{E}}_{x,y}\big[\frac{1}{p_{x,y}} + 3\big]\Big) + t(\ln 2) \mathop{\mathbb{E}}_{x,y}\Big[\log \frac{1}{p_{x,y}}\Big] + 1 \ ,\end{aligned}$$

where $\mathbb{E}_{x,y}[\cdot]$ denotes the expectation with respect to (x,y) chosen according to the distribution P_{XY}. Because $\mathbb{E}_{x,y}[\frac{1}{p_{x,y}}] = \sum_{x,y} P_{XY}(x,y) \frac{P_Y(y)}{P_{XY}(x,y)} = |\mathcal{X}|$ and $\mathbb{E}_{x,y}[\log \frac{1}{p_{x,y}}] = H(X|Y)$, we obtain

$$\mathop{\mathbb{E}}_{x,y}\left[p_{x,y}^{-t}\right] \leq r_t(|\mathcal{X}|+3) + t(\ln 2)H(X|Y) + 1 \ .$$

Furthermore, because $\log a \leq \frac{1}{\ln 2}(a-1)$,

$$\log \mathop{\mathbb{E}}_{x,y}\left[p_{x,y}^{-t}\right] \leq \frac{1}{\ln 2} r_t\big(|\mathcal{X}|+3\big) + tH(X|Y) \ .$$

Finally, together with Lemma B.6.4, since $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$, we conclude

$$\log \mathop{\mathbb{E}}_{x,y}\bigl[p_{x,y}^{-t}\bigr] \leq \bigl(\frac{1}{\ln 2} - 1\bigr) t^2 \log(|\mathcal{X}|+3)^2 + tH(X|Y) \ .$$

The assertion follows because $\frac{1}{\ln 2} - 1 \leq \frac{1}{2}$. □

Lemma 3.3.2. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a probability distribution and let γ be the function on $\mathcal{X} \times \mathcal{Y}$ defined by*

$$\gamma(x,y) := -\log P_{X|Y}(x,y) - H(X|Y) \ .$$

Then, for any $t \in \mathbb{R}$ with $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$,

$$\mathop{\mathbb{E}}_{x,y}\bigl[2^{t\gamma(x,y)}\bigr] \leq 2^{\frac{1}{2}t^2 \log(|\mathcal{X}|+3)^2} \ .$$

Proof. The assertion follows directly from Lemma 3.3.1, that is,

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y}\bigl[2^{t\gamma(x,y)}\bigr] &= 2^{-tH(X|Y)} \mathop{\mathbb{E}}_{x,y}\bigl[P_{X|Y}(x,y)^{-t}\bigr] \\ &\leq 2^{-tH(X|Y)} \cdot 2^{tH(X|Y)+\frac{1}{2}t^2 \log(|\mathcal{X}|+3)^2} \ .\end{aligned}$$ □

Theorem 3.3.3. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a probability distribution and let $n \in \mathbb{N}$. Then, for any $\delta \in [0, \log|\mathcal{X}|]$ and $(\mathbf{x},\mathbf{y})$ chosen according to $P_{X^nY^n} := (P_{XY})^n$,*

$$\Pr_{\mathbf{x},\mathbf{y}}\bigl[-\log P_{X^n|Y^n}(\mathbf{x},\mathbf{y}) \geq n\bigl(H(X|Y)+\delta\bigr)\bigr] \leq 2^{-\frac{n\delta^2}{2\log(|\mathcal{X}|+3)^2}} \ ,$$

and, similarly,

$$\Pr_{\mathbf{x},\mathbf{y}}\bigl[-\log P_{X^n|Y^n}(\mathbf{x},\mathbf{y}) \leq n\bigl(H(X|Y)-\delta\bigr)\bigr] \leq 2^{-\frac{n\delta^2}{2\log(|\mathcal{X}|+3)^2}} \ .$$

Proof. Let $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, and let γ be the function defined in Lemma 3.3.2 for the probability distribution P_{XY}. Then

$$\sum_{i=1}^n \gamma(x_i, y_i) = -\log P_{X^n|Y^n}(\mathbf{x},\mathbf{y}) - nH(X|Y) \ . \tag{3.24}$$

Using Markov's inequality, for any $t > 0$,

$$\begin{aligned}\Pr_{\mathbf{x},\mathbf{y}}\bigl[\sum_{i=1}^n \gamma(x_i,y_i) \geq n\delta\bigr] &= \Pr_{\mathbf{x},\mathbf{y}}\bigl[2^{t\sum_{i=1}^n \gamma(x_i,y_i)} \geq 2^{tn\delta}\bigr] \\ &\leq \frac{\mathbb{E}_{\mathbf{x},\mathbf{y}}\bigl[2^{t\sum_{i=1}^n \gamma(x_i,y_i)}\bigr]}{2^{tn\delta}} \ .\end{aligned} \tag{3.25}$$

Moreover, because the pairs (x_i, y_i) are chosen independently,

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}}\big[2^{t\sum_{i=1}^n \gamma(x_i,y_i)}\big] &= \mathop{\mathbb{E}}_{\mathbf{x},\mathbf{y}}\big[\prod_{i=1}^n 2^{t\gamma(x_i,y_i)}\big] \\
&= \prod_{i=1}^n \mathop{\mathbb{E}}_{x_i,y_i}\big[2^{t\gamma(x_i,y_i)}\big] \\
&\leq \big(2^{\frac{1}{2}t^2 \log(|\mathcal{X}|+3)^2}\big)^n ,
\end{aligned}
$$

where the inequality follows from Lemma 3.3.2, for any $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$. Combining this with (3.25) gives

$$
\Pr_{\mathbf{x},\mathbf{y}}\big[\sum_{i=1}^n \gamma(x_i,y_i) \geq n\delta\big] \leq 2^{\frac{1}{2}nt^2 \log(|\mathcal{X}|+3)^2 - tn\delta} .
$$

With $t := \frac{\delta}{\log(|\mathcal{X}|+3)^2}$ (note that $t \leq \frac{1}{\log(|\mathcal{X}|+3)}$ because $\delta \leq \log|\mathcal{X}|$), we conclude

$$
\Pr_{\mathbf{x},\mathbf{y}}\big[\sum_{i=1}^n \gamma(x_i,y_i) \geq n\delta\big] \leq 2^{-\frac{n\delta^2}{2\log(|\mathcal{X}|+3)^2}} .
$$

The first inequality of the lemma then follows from (3.24).

Similarly, if $t < 0$,

$$
\begin{aligned}
\Pr_{\mathbf{x},\mathbf{y}}\big[\sum_{i=1}^n \gamma(x_i,y_i) \leq -n\delta\big] &= \Pr_{\mathbf{x},\mathbf{y}}\big[2^{t\sum_{i=1}^n \gamma(x_i,y_i)} \geq 2^{-tn\delta}\big] \\
&\leq \frac{\mathbb{E}_{\mathbf{x},\mathbf{y}}\big[2^{t\sum_{i=1}^n \gamma(x_i,y_i)}\big]}{2^{-tn\delta}} ,
\end{aligned}
$$

and thus

$$
\Pr_{\mathbf{x},\mathbf{y}}\big[\sum_{i=1}^n \gamma(x_i,y_i) \leq -n\delta\big] \leq 2^{\frac{1}{2}nt^2 \log(|\mathcal{X}|+3)^2 + tn\delta} .
$$

The second inequality follows with $t := -\frac{\delta}{\log(|\mathcal{X}|+3)^2}$. □

Asymptotic equality of smooth entropy and Shannon entropy

Theorem 3.3.4. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a probability distribution and let $n \in \mathbb{N}$. Then, for any $\varepsilon \geq 0$ and $P_{X^nY^n} := (P_{XY})^n$,*

$$
\begin{aligned}
\frac{1}{n} H_{\max}^{\varepsilon}(P_{X^nY^n}|P_{Y^n}) &\leq H(X|Y) + \delta \\
\frac{1}{n} H_{\min}^{\varepsilon}(P_{X^nY^n}|P_{Y^n}) &\geq H(X|Y) - \delta ,
\end{aligned}
$$

where $\delta := \log(|\mathcal{X}| + 3)\sqrt{\frac{2\log(1/\varepsilon)}{n}}$.

Proof. We first prove the bound on the (classical) smooth max-entropy $H^\varepsilon_{\max}(P_{X^nY^n}|P_{Y^n})$. For any $\mathbf{y} \in \mathcal{Y}^n$ with $P_{Y^n}(\mathbf{y}) > 0$, let $\bar{\mathcal{X}}_\mathbf{y}$ be the set of all n-tuples $\mathbf{x} \in \mathcal{X}^n$ such that

$$-\log P_{X^n|Y^n}(\mathbf{x},\mathbf{y}) \leq n\bigl(H(X|Y)+\delta\bigr) \ .$$

Furthermore, let $P_{\bar{X}^n\bar{Y}^n}$ be the nonnegative function on $\mathcal{X}^n \times \mathcal{Y}^n$ defined by

$$P_{\bar{X}^n\bar{Y}^n}(\mathbf{x},\mathbf{y}) = \begin{cases} P_{X^nY^n}(\mathbf{x},\mathbf{y}) & \text{if } \mathbf{x} \in \bar{\mathcal{X}}_\mathbf{y} \\ 0 & \text{otherwise.} \end{cases} \tag{3.26}$$

We can assume without loss of generality that $\delta \leq \log|\mathcal{X}|$ (otherwise, the statement is trivial). Hence, by the first inequality of Theorem 3.3.3, $\Pr_{\mathbf{x},\mathbf{y}}[\mathbf{x} \notin \bar{\mathcal{X}}_\mathbf{y}] \leq \varepsilon$. This implies $\|P_{X^nY^n} - P_{\bar{X}^n\bar{Y}^n}\|_1 \leq \varepsilon$ and thus

$$H^\varepsilon_{\max}(P_{X^nY^n}|P_{Y^n}) \leq H_{\max}(P_{\bar{X}^n\bar{Y}^n}|P_{Y^n}) \ . \tag{3.27}$$

For any fixed $\mathbf{y} := (y_1, \ldots, y_n) \in \mathcal{Y}^n$ with $P_{Y^n}(\mathbf{y}) > 0$,

$$1 \geq \sum_{\mathbf{x}\in\bar{\mathcal{X}}_\mathbf{y}} \prod_{i=1}^n P_{X|Y}(x_i, y_i) \geq |\bar{\mathcal{X}}_\mathbf{y}| 2^{-n(H(X|Y)+\delta)} \ ,$$

where the second inequality follows from the definition of the set $\bar{\mathcal{X}}_\mathbf{y}$. Consequently, we have $|\bar{\mathcal{X}}_\mathbf{y}| \leq 2^{n(H(X|Y)+\delta)}$. Moreover, by the definition of $P_{\bar{X}^n\bar{Y}^n}$, the support of the function $\mathbf{x} \mapsto P_{\bar{X}^n\bar{Y}^n}(\mathbf{x},\mathbf{y})$ is contained in $\bar{\mathcal{X}}_\mathbf{y}$. Hence, using Remark 3.1.4,

$$H_{\max}(P_{\bar{X}^n\bar{Y}^n}|P_{Y^n}) \leq \log\Bigl(\sum_{\mathbf{y}\in\mathcal{Y}^n} P_{Y^n}(\mathbf{y}) \cdot |\bar{\mathcal{X}}_\mathbf{y}|\Bigr) \leq n\bigl(H(X|Y)+\delta\bigr) \ .$$

Combining this with (3.27) proves the first inequality of the lemma.

To prove the bound on the min-entropy $H^\varepsilon_{\min}(P_{X^nY^n}|P_{Y^n})$, let $\bar{\mathcal{X}}_\mathbf{y}$, for any $\mathbf{y} \in \mathcal{Y}^n$ with $P_{Y^n}(\mathbf{y}) > 0$, be the set of n-tuples $\mathbf{x} \in \mathcal{X}^n$ such that

$$-\log P_{X^n|Y^n}(\mathbf{x},\mathbf{y}) \geq n\bigl(H(X|Y)-\delta\bigr) \ ,$$

and let again $P_{\bar{X}^n\bar{Y}^n}$ be defined by (3.26). By the second inequality of Theorem 3.3.3, $\Pr_{\mathbf{x},\mathbf{y}}[\mathbf{x} \notin \bar{\mathcal{X}}_\mathbf{y}] \leq \varepsilon$, which, similarly to the previous argument, implies

$$H^\varepsilon_{\min}(P_{X^nY^n}|P_{Y^n}) \geq H_{\min}(P_{\bar{X}^n\bar{Y}^n}|P_{Y^n}) \ . \tag{3.28}$$

Moreover, using Remark 3.1.4

$$\begin{aligned} H_{\min}(P_{\bar{X}^n\bar{Y}^n}|P_{Y^n}) &= -\log \max_{\mathbf{y}\in\mathrm{supp}(P_{Y^n})} \max_{\mathbf{x}\in\mathcal{X}^n} \frac{P_{\bar{X}^n\bar{Y}^n}(\mathbf{x},\mathbf{y})}{P_{Y^n}(\mathbf{y})} \\ &= -\log \max_{\mathbf{y}\in\mathrm{supp}(P_{Y^n})} \max_{\mathbf{x}\in\bar{\mathcal{X}}_\mathbf{y}} \frac{P_{X^nY^n}(\mathbf{x},\mathbf{y})}{P_{Y^n}(\mathbf{y})} \\ &\geq n\bigl(H(X|Y)-\delta\bigr) \ , \end{aligned}$$

where the inequality follows from the definition of the set $\bar{\mathcal{X}}_\mathbf{y}$. Combining this with (3.28) proves the second inequality of the lemma. □

Because the min-entropy $H_{\min}(P_{X^nY^n}|P_{Y^n})$ cannot be larger than the max-entropy $H_{\max}(P_{X^nY^n}|P_{Y^n})$ (cf. Lemma 3.1.5), Theorem 3.3.4 implies that

$$\frac{1}{n}H^\varepsilon_{\min}(P_{X^nY^n}|P_{Y^n}) \approx \frac{1}{n}H^\varepsilon_{\max}(P_{X^nY^n}|P_{Y^n}) \approx \frac{1}{n}H(X^n|Y^n) , \qquad (3.29)$$

where asymptotically, for increasing n, the approximation becomes an equality.

Remark 3.3.5. It is easy to see that Theorem 3.3.4 can be generalized to probability distributions $P_{X^nY^n}$ which are the product of not necessarily identical distributions $P_{X_iY_i}$. That is, for any distribution of the form $P_{X^nY^n} = \prod_{i=1}^n P_{X_iY_i}$, the approximation (3.29) still holds.

3.3.2 The quantum case

The following theorem and its corollary can be seen as a quantum version of Theorem 3.3.4 for smooth min-entropy (where the Shannon entropy is replaced by the von Neumann entropy). The proof essentially follows the same line as the classical argument described above.[7] A similar argument shows that the statement also holds for smooth max-entropy.

Theorem 3.3.6. *Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$ be density operators, and let $n \in \mathbb{N}$. Then, for any $\varepsilon \geq 0$,*

$$\frac{1}{n}H^\varepsilon_{\min}(\rho_{AB}^{\otimes n}|\sigma_B^{\otimes n}) \geq H(\rho_{AB}) - H(\rho_B) - D(\rho_B\|\sigma_B) - \delta ,$$

where $\delta := 2\log\bigl(\mathrm{rank}(\rho_A) + \mathrm{tr}(\rho_{AB}^2(\mathrm{id}_A \otimes \sigma_B^{-1})) + 2\bigr)\sqrt{\frac{\log(1/\varepsilon)}{n}} + 1$.

Proof. Define $H(\rho_{AB}|\sigma_B) := H(\rho_{AB}) - H(\rho_B) - D(\rho_B\|\sigma_B)$. We show that there exists a density operator $\bar{\rho}_{A^nB^n} \in \mathcal{B}^\varepsilon(\rho_{AB}^{\otimes n})$ such that

$$H_{\min}(\bar{\rho}_{A^nB^n}|\sigma_B^{\otimes n}) \geq nH(\rho_{AB}|\sigma_B) - n\delta . \qquad (3.30)$$

According to the definition of min-entropy, this is equivalent to saying that the operator $\lambda \cdot (\mathrm{id}_A \otimes \sigma_B)^{\otimes n} - \bar{\rho}_{A^nB^n}$ is nonnegative, for $\lambda \geq 0$ such that $-\log\lambda = nH(\rho_{AB}|\sigma_B) - n\delta$.

Let

$$(\mathrm{id}_A \otimes \sigma_B)^{\otimes n} = \sum_{\mathbf{z}\in\mathcal{Z}^n} q_{\mathbf{z}}|\mathbf{z}\rangle\langle\mathbf{z}|$$

[7] An alternative method to prove the statement $\frac{1}{n}H^\varepsilon_{\min}(\rho_{AB}^{\otimes n}|\rho_B^{\otimes n}) \gtrsim H(\rho_{AB}) - H(\rho_B)$ is to use a chain rule of the form $H^\varepsilon_{\min}(\rho_{AB}^{\otimes n}|\rho_B^{\otimes n}) \gtrsim H^\varepsilon_{\min}(\rho_{AB}^{\otimes n}) - H^\varepsilon_{\max}(\rho_B^{\otimes n})$. The entropies on the right hand side of this inequality can be rewritten as the entropies of the classical probability distributions defined by the eigenvalues of $\rho_{AB}^{\otimes n}$ and $\rho_B^{\otimes n}$, respectively. The desired bound then follows from the classical Theorem 3.3.4. However, the results obtained with such an alternative method are less tight and less general than Theorem 3.3.6.

be a spectral decomposition of $(\mathrm{id}_A \otimes \sigma_B)^{\otimes n}$. We can assume without loss of generality that there exists an order relation on the values $\mathcal{Z}^n$ such that $q_{\mathbf{z}} \geq q_{\mathbf{z}'}$, for any $\mathbf{z} \geq \mathbf{z}'$. For any $\mathbf{z} \in \mathcal{Z}$, let $B_{\mathbf{z}}$ be the projector defined by

$$B_{\mathbf{z}} := \sum_{\mathbf{z}' :\, \mathbf{z}' \geq \mathbf{z}} |\mathbf{z}'\rangle\langle\mathbf{z}'| \ .$$

Moreover, let $\beta_{\mathbf{z}}$, for $\mathbf{z} \in \mathcal{Z}^n$, be nonnegative coefficients such that, for any $\mathbf{z}' \in \mathcal{Z}^n$,

$$\sum_{\mathbf{z} :\, \mathbf{z} \leq \mathbf{z}'} \beta_{\mathbf{z}} = q_{\mathbf{z}'} \ .$$

Note that the spectral decomposition above can then be rewritten as

$$(\mathrm{id}_A \otimes \sigma_B)^{\otimes n} = \sum_{\mathbf{z} \in \mathcal{Z}^n} \beta_{\mathbf{z}} B_{\mathbf{z}} \ . \tag{3.31}$$

Let

$$\rho_{AB}^{\otimes n} = \sum_{\mathbf{x} \in \mathcal{X}^n} p_{\mathbf{x}} |\mathbf{x}\rangle\langle\mathbf{x}|$$

be a spectral decomposition of $\rho_{AB}^{\otimes n}$. In the following, we denote by ∞ an element which is larger than any element of $\mathcal{Z}^n$. Moreover, let $p_{\mathbf{x},\mathbf{z}}$, for $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{z} \in \mathcal{Z}^n \cup \{\infty\}$, be nonnegative coefficients such that, for any $\mathbf{z}' \in \mathcal{Z}^n$,

$$\sum_{\mathbf{z} :\, \mathbf{z} \leq \mathbf{z}'} p_{\mathbf{x},\mathbf{z}} = \min(p_{\mathbf{x}}, \lambda q_{\mathbf{z}'})$$

$$\sum_{\mathbf{z} \in \mathcal{Z}^n \cup \{\infty\}} p_{\mathbf{x},\mathbf{z}} = p_{\mathbf{x}} \ .$$

We show that inequality (3.30) holds for the operator

$$\bar{\rho}_{A^n B^n} := \sum_{\mathbf{x} \in \mathcal{X}^n} \sum_{\mathbf{z} \in \mathcal{Z}^n} p_{\mathbf{x},\mathbf{z}} B_{\mathbf{z}} |\mathbf{x}\rangle\langle\mathbf{x}| B_{\mathbf{z}} \ .$$

Note first that, by the definition of $p_{\mathbf{x},\mathbf{z}}$ and $\beta_{\mathbf{z}}$, we have $p_{\mathbf{x},\mathbf{z}} \leq \lambda \beta_{\mathbf{z}}$, for any $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{z} \in \mathcal{Z}^n$, that is, the operator

$$\sum_{\mathbf{z} \in \mathcal{Z}^n} \lambda \beta_{\mathbf{z}} B_{\mathbf{z}} B_{\mathbf{z}} - \bar{\rho}_{A^n B^n} = \sum_{\mathbf{z} \in \mathcal{Z}^n} \sum_{\mathbf{x} \in \mathcal{X}^n} (\lambda \beta_{\mathbf{z}} - p_{\mathbf{x},\mathbf{z}}) B_{\mathbf{z}} |\mathbf{x}\rangle\langle\mathbf{x}| B_{\mathbf{z}}$$

is nonnegative. Using (3.31) and the fact that the operators $B_{\mathbf{z}}$ are projectors, we conclude that the operator

$$\lambda \cdot (\mathrm{id}_A \otimes \sigma_B)^{\otimes n} - \bar{\rho}_{A^n B^n} = \sum_{\mathbf{z} \in \mathcal{Z}^n} \lambda \beta_{\mathbf{z}} B_{\mathbf{z}} - \bar{\rho}_{A^n B^n}$$

is nonnegative, which implies (3.30). It thus remains to be proven that $\bar{\rho}_{A^n B^n} \in \mathcal{B}^{\varepsilon}(\rho_{AB}^{\otimes n})$.

Using the above definitions and the convention that B_∞ is the zero matrix, we have

$$\begin{aligned}\left\|\rho_{AB}^{\otimes n} - \bar{\rho}_{A^n B^n}\right\|_1 &= \Big\| \sum_{\mathbf{x}\in\mathcal{X}^n} \sum_{\mathbf{z}\in\mathcal{Z}^n\cup\{\infty\}} p_{\mathbf{x},\mathbf{z}}\big(|\mathbf{x}\rangle\langle\mathbf{x}| - B_{\mathbf{z}}|\mathbf{x}\rangle\langle\mathbf{x}|B_{\mathbf{z}}\big)\Big\|_1 \\ &\leq \sum_{\mathbf{x}\in\mathcal{X}^n} \sum_{\mathbf{z}\in\mathcal{Z}^n\cup\{\infty\}} p_{\mathbf{x},\mathbf{z}}\left\||\mathbf{x}\rangle\langle\mathbf{x}| - B_{\mathbf{z}}|\mathbf{x}\rangle\langle\mathbf{x}|B_{\mathbf{z}}\right\|_1 .\end{aligned}$$

We can use Lemma A.2.8 to bound the trace distance on the right hand side of this inequality, that is,

$$\left\||\mathbf{x}\rangle\langle\mathbf{x}| - B_{\mathbf{z}}|\mathbf{x}\rangle\langle\mathbf{x}|B_{\mathbf{z}}\right\|_1 \leq 2\sqrt{1 - \mathrm{tr}(B_{\mathbf{z}}|\mathbf{x}\rangle\langle\mathbf{x}|B_{\mathbf{z}})} .$$

Because ρ_{AB} is a density operator, the nonnegative coefficients $p_{\mathbf{x},\mathbf{z}}$ sum up to one. We can thus apply Jensen's inequality which gives

$$\begin{aligned}\left\|\rho_{AB}^{\otimes n} - \bar{\rho}_{A^n B^n}\right\|_1 &\leq 2 \sum_{\mathbf{x}\in\mathcal{X}^n} \sum_{\mathbf{z}\in\mathcal{Z}^n\cup\{\infty\}} p_{\mathbf{x},\mathbf{z}}\sqrt{1 - \mathrm{tr}(B_{\mathbf{z}}|\mathbf{x}\rangle\langle\mathbf{x}|B_{\mathbf{z}})} \\ &\leq 2\sqrt{\sum_{\mathbf{x}\in\mathcal{X}^n} \sum_{\mathbf{z}\in\mathcal{Z}^n\cup\{\infty\}} p_{\mathbf{x},\mathbf{z}}(1 - \mathrm{tr}(B_{\mathbf{z}}|\mathbf{x}\rangle\langle\mathbf{x}|B_{\mathbf{z}}))} \qquad (3.32) \\ &= 2\sqrt{1 - \mathrm{tr}(\bar{\rho}_{A^n B^n})} .\end{aligned}$$

The trace in the square root can be rewritten as

$$\begin{aligned}\mathrm{tr}(\bar{\rho}_{A^n B^n}) &= \sum_{\mathbf{z}'\in\mathcal{Z}^n} \langle\mathbf{z}'|\Big(\sum_{\mathbf{x}\in\mathcal{X}^n}\sum_{\mathbf{z}\in\mathcal{Z}^n} p_{\mathbf{x},\mathbf{z}} B_{\mathbf{z}}|\mathbf{x}\rangle\langle\mathbf{x}|B_{\mathbf{z}}\Big)|\mathbf{z}'\rangle \\ &= \sum_{\mathbf{z}'\in\mathcal{Z}^n}\sum_{\mathbf{z}:\,\mathbf{z}\leq\mathbf{z}'}\sum_{\mathbf{x}\in\mathcal{X}^n} p_{\mathbf{x},\mathbf{z}}|\langle\mathbf{z}'|\mathbf{x}\rangle|^2 .\end{aligned}$$

Because the terms in the sum are all nonnegative, the sum can only become smaller if we restrict the set of values $\mathbf{x}$ over which the sum is taken. Consequently,

$$\mathrm{tr}(\bar{\rho}_{A^n B^n}) \geq \sum_{\mathbf{z}'\in\mathcal{Z}^n}\sum_{\mathbf{x}:\,p_{\mathbf{x}}\leq\lambda q_{\mathbf{z}'}} |\langle\mathbf{z}'|\mathbf{x}\rangle|^2 \sum_{\mathbf{z}:\,\mathbf{z}\leq\mathbf{z}'} p_{\mathbf{x},\mathbf{z}} .$$

By the definition of $p_{\mathbf{x},\mathbf{z}}$, we have $\sum_{\mathbf{z}:\,\mathbf{z}\leq\mathbf{z}'} p_{\mathbf{x},\mathbf{z}} = p_{\mathbf{x}}$, for any $(\mathbf{x},\mathbf{z}')$ such that $p_{\mathbf{x}} \leq \lambda q_{\mathbf{z}'}$, and hence

$$\mathrm{tr}(\bar{\rho}_{A^n B^n}) \geq \sum_{(\mathbf{x},\mathbf{z}'):\,p_{\mathbf{x}}\leq\lambda q_{\mathbf{z}'}} p_{\mathbf{x}}|\langle\mathbf{z}'|\mathbf{x}\rangle|^2 .$$

Because $\sum_{\mathbf{z},\mathbf{x}} p_{\mathbf{x}}|\langle\mathbf{z}|\mathbf{x}\rangle|^2 = 1$, this inequality can be rewritten as

$$1 - \mathrm{tr}(\bar{\rho}_{A^n B^n}) \leq \sum_{(\mathbf{x},\mathbf{z}):\,p_{\mathbf{x}}>\lambda q_{\mathbf{z}}} p_{\mathbf{x}}|\langle\mathbf{z}|\mathbf{x}\rangle|^2$$

Recall that we need to prove that $\bar{\rho}_{A^n B^n} \in \mathcal{B}^{\varepsilon}(\rho_{AB}^{\otimes n})$. Hence, combining (3.32) with the above bound on $\mathrm{tr}(\bar{\rho}_{A^n B^n})$, it remains to be shown that

$$\sum_{(\mathbf{x},\mathbf{z}):\, p_{\mathbf{x}} > \lambda q_{\mathbf{z}}} p_{\mathbf{x}} |\langle \mathbf{z} | \mathbf{x} \rangle|^2 \leq \left(\frac{\varepsilon}{2}\right)^2 . \tag{3.33}$$

Let

$$\mathrm{id}_A \otimes \sigma_B = \sum_{\bar{z} \in \bar{\mathcal{Z}}} \bar{q}_{\bar{z}} |\bar{z}\rangle\langle\bar{z}|$$

and

$$\rho_{AB} = \sum_{\bar{x} \in \bar{\mathcal{X}}} \bar{p}_{\bar{x}} |\bar{x}\rangle\langle\bar{x}|$$

be spectral decompositions of $\mathrm{id}_A \otimes \sigma_B$ and ρ_{AB}, respectively. Moreover, let $P_{\bar{X}\bar{Z}}$ be the probability distribution defined by

$$P_{\bar{X}\bar{Z}}(\bar{x}, \bar{z}) := \bar{p}_{\bar{x}} |\langle \bar{z} | \bar{x} \rangle|^2 .$$

Note that $|\mathbf{x}\rangle$ and $p_{\mathbf{x}}$, as used above, can be defined as $|\mathbf{x}\rangle := \bigotimes_{i=1}^n |x_i\rangle$ and $p_{\mathbf{x}} = p_{(x_1,\ldots,x_n)} := \prod_{i=1}^n \bar{p}_{x_i}$. Similarly, we can set $|\mathbf{z}\rangle := \bigotimes_{i=1}^n |z_i\rangle$ and $q_{\mathbf{z}} = q_{(z_1,\ldots,z_n)} := \prod_{i=1}^n \bar{q}_{z_i}$. Then, the left hand side of (3.33) can be rewritten as

$$\begin{aligned}
\sum_{(\mathbf{x},\mathbf{z}):\, p_{\mathbf{x}} > \lambda q_{\mathbf{z}}} p_{\mathbf{x}} |\langle \mathbf{z} | \mathbf{x} \rangle|^2 &= \Pr_{\mathbf{x},\mathbf{z}}[p_{\mathbf{x}} > \lambda q_{\mathbf{z}}] \\
&= \Pr_{\mathbf{x},\mathbf{z}}[-\log p_{\mathbf{x}} + \log q_{\mathbf{z}} < -\log \lambda] \\
&= \Pr_{\mathbf{x},\mathbf{z}}\Bigl[\sum_{i=1}^n -\log \bar{p}_{x_i} + \log \bar{q}_{z_i} < -\log \lambda\Bigr]
\end{aligned} \tag{3.34}$$

for $(\mathbf{x}, \mathbf{z})$ chosen according to the probability distribution $(P_{\bar{X}\bar{Z}})^n$.

By the definition of $H(\rho_{AB}|\sigma_B)$, we have

$$\begin{aligned}
H(\rho_{AB}|\sigma_B) &= -\mathrm{tr}(\rho_{AB} \log \rho_{AB}) + \mathrm{tr}(\rho_{AB} \log \mathrm{id}_A \otimes \sigma_B) \\
&= \sum_{\bar{x},\bar{z}} \bar{p}_{\bar{x}} |\langle \bar{z} | \bar{x} \rangle|^2 \bigl(\log \frac{1}{\bar{p}_{\bar{x}}} - \log \frac{1}{\bar{q}_{\bar{z}}}\bigr) \\
&= \mathop{\mathbb{E}}_{\bar{x},\bar{z}}[-\log \bar{p}_{\bar{x}} + \log \bar{q}_{\bar{z}}] ,
\end{aligned}$$

for $(\bar{x}, \bar{y})$ chosen according to $P_{\bar{X}\bar{Y}}$. According to Birkhoff's theorem (cf. Theorem B.2.2) there exist nonnegative coefficients μ_π parameterized by the bijections π from $\mathcal{X}$ to $\mathcal{Z}$ such that $\sum_\pi \mu_\pi = 1$ and $|\langle \bar{z} | \bar{x} \rangle|^2 = \sum_\pi \mu_\pi \delta_{\bar{z}, \pi(\bar{x})}$. The identity above can thus be rewritten as

$$H(\rho_{AB}|\sigma_B) = \sum_{\bar{x},\bar{z}} \bar{p}_{\bar{x}} |\langle \bar{z} | \bar{x} \rangle|^2 \log \frac{\bar{q}_{\bar{z}}}{\bar{p}_{\bar{x}}} = \sum_\pi \mu_\pi \sum_{\bar{x}} \bar{p}_{\bar{x}} \log \frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}} . \tag{3.35}$$

For $(\bar{x}, \bar{z})$ chosen according to $P_{\bar{X}\bar{Y}}$,

$$\mathop{\mathbb{E}}_{\bar{x},\bar{z}}\bigl[2^{-t(\log \bar{p}_{\bar{x}} - \log \bar{q}_{\bar{z}})}\bigr] = \sum_{\bar{x},\bar{z}} \bar{p}_{\bar{x}} |\langle \bar{z}|\bar{x}\rangle|^2 \bigl(\frac{\bar{p}_{\bar{x}}}{\bar{q}_{\bar{z}}}\bigr)^{-t} = \sum_{\pi} \mu_\pi \sum_{\bar{x}} \bar{p}_{\bar{x}} \Bigl(\frac{\bar{p}_{\bar{x}}}{\bar{q}_{\pi(\bar{x})}}\Bigr)^{-t} .$$

For any $t \in \mathbb{R}$, let r_t be the function defined by (3.23). The last term in the sum above can then be bounded by

$$\begin{aligned}\Bigl(\frac{\bar{p}_{\bar{x}}}{\bar{q}_{\pi(\bar{x})}}\Bigr)^{-t} &= r_t\Bigl(\frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}}\Bigr) + t \ln \frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}} + 1 \\ &\leq r_{|t|}\Bigl(\frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}} + \frac{\bar{p}_{\bar{x}}}{\bar{q}_{\pi(\bar{x})}} + 2\Bigr) + t \ln \frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}} + 1\end{aligned}$$

where the inequality follows from the fact that, for all $z > 0$, $r_t(z) \leq r_{|t|}(z + \frac{1}{z})$ (Lemma B.6.2) and the fact that r_t is monotonically increasing (Lemma B.6.1) on the interval $[1, \infty)$. Because $\frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}} + \frac{\bar{p}_{\bar{x}}}{\bar{q}_{\pi(\bar{x})}} + 2 \in [4, \infty)$ and because r_t is concave on this interval (Lemma B.6.3) we can apply Jensen's inequality, which gives

$$\begin{aligned}\mathop{\mathbb{E}}_{\bar{x},\bar{z}}[2^{-t(\log \bar{p}_{\bar{x}} - \log \bar{q}_{\bar{z}})}] \leq r_{|t|}\Bigl(\sum_{\pi} \mu_\pi \sum_{\bar{x}} \bar{p}_{\bar{x}} \bigl(\frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}} + \frac{\bar{p}_{\bar{x}}}{\bar{q}_{\pi(\bar{x})}} + 2\bigr)\Bigr) \\ + t(\ln 2) \sum_{\pi} \mu_\pi \sum_{\bar{x}} \bar{p}_{\bar{x}} \log \frac{\bar{q}_{\pi(\bar{x})}}{\bar{p}_{\bar{x}}} + 1 . \qquad (3.36)\end{aligned}$$

Note that $\sum_{\bar{z}} \bar{q}_{\bar{z}} = \mathrm{tr}(\mathrm{id}_A \otimes \sigma_B) = \dim(\mathcal{H}_A)$. As we can assume without loss of generality that $\mathcal{H}_A$ is restricted to the support of ρ_A, we have

$$\sum_{\pi} \mu_\pi \sum_{\bar{x}} \bar{q}_{\pi(\bar{x})} = \mathrm{rank}(\rho_A) .$$

Moreover,

$$\sum_{\pi} \mu_\pi \sum_{\bar{x}} \frac{\bar{p}_{\bar{x}}^2}{\bar{q}_{\pi(\bar{x})}} = \sum_{\bar{x},\bar{z}} |\langle \bar{x}|\bar{z}\rangle|^2 \bar{p}_{\bar{x}}^2 \bar{q}_{\bar{z}}^{-1} = \mathrm{tr}\bigl(\rho_{AB}^2 (\mathrm{id}_A \otimes \sigma_B^{-1})\bigr) .$$

Hence, together with (3.35), the bound (3.36) can be rewritten as

$$\mathop{\mathbb{E}}_{\bar{x},\bar{z}}[2^{-t(\log \bar{p}_{\bar{x}} - \log \bar{q}_{\bar{z}})}] \leq r_{|t|}(\gamma + 2) + t(\ln 2) H(\rho_{AB}|\sigma_B) + 1 ,$$

where $\gamma := \mathrm{rank}(\rho_A) + \mathrm{tr}\bigl(\rho_{AB}^2(\mathrm{id}_A \otimes \sigma_B^{-1})\bigr)$. Furthermore, using the fact that $\log a \leq \frac{1}{\ln 2}(a - 1)$ we find

$$\begin{aligned}\mathop{\mathbb{E}}_{\bar{x},\bar{z}}[2^{-t(\log \bar{p}_{\bar{x}} - \log \bar{q}_{\bar{z}})}] &\leq 2^{\log\bigl(r_{|t|}(\gamma+2) + t(\ln 2) H(\rho_{AB}|\sigma_B) + 1\bigr)} \\ &\leq 2^{\frac{1}{\ln 2} r_{|t|}(\gamma+2) + t H(\rho_{AB}|\sigma_B)} .\end{aligned}$$

With Lemma B.6.4, we conclude

$$\mathbb{E}_{\bar{x},\bar{z}}[2^{t(-\log \bar{p}_{\bar{x}}+\log \bar{q}_{\bar{z}}-H(\rho_{AB}|\sigma_B))}] \leq 2^{(\frac{1}{\ln 2}-1)t^2 \log(\gamma+2)^2} \leq 2^{\frac{1}{2}t^2 \log(\gamma+2)^2} \ . \quad (3.37)$$

Let now $w(\mathbf{x},\mathbf{z}) := \sum_{i=1}^n(-\log \bar{p}_{x_i} + \log \bar{q}_{z_i} - H(\rho_{AB}|\sigma_B))$. Because the expectation of the product of independent values is equal to the product of the expectation of these values, we have, for $(\mathbf{x},\mathbf{z})$ chosen according to $(P_{\bar{X}\bar{Z}})^n$,

$$\mathbb{E}_{\mathbf{x},\mathbf{z}}[2^{tw(\mathbf{x},\mathbf{z})}] = \mathbb{E}_{\bar{x},\bar{z}}[2^{t(-\log \bar{p}_{\bar{x}}+\log \bar{q}_{\bar{z}}-H(\rho_{AB}|\sigma_B))}]^n \ .$$

Hence, by Markov's inequality, for any $t \leq 0$,

$$\begin{aligned}\Pr_{\mathbf{x},\mathbf{z}}[w(\mathbf{x},\mathbf{z}) \leq -n\delta] &= \Pr_{\mathbf{x},\mathbf{z}}[2^{tw(\mathbf{x},\mathbf{z})} \geq 2^{-tn\delta}] \\ &\leq \frac{\mathbb{E}_{\mathbf{x},\mathbf{z}}[2^{tw(\mathbf{x},\mathbf{z})}]}{2^{-tn\delta}} \\ &= \frac{\mathbb{E}_{\bar{x},\bar{z}}[2^{t(-\log \bar{p}_{\bar{x}}+\log \bar{q}_{\bar{z}}-H(\rho_{AB}|\sigma_B))}]^n}{2^{-tn\delta}}\end{aligned}$$

and thus, using (3.37),

$$\Pr_{\mathbf{x},\mathbf{z}}[w(\mathbf{x},\mathbf{z}) \leq -n\delta] \leq 2^{\frac{1}{2}t^2 n \log(\gamma+2)^2 + tn\delta} \ .$$

Consequently, with $t := -\frac{\delta}{\log(\gamma+2)^2}$,

$$\begin{aligned}&\Pr_{\mathbf{x},\mathbf{z}}[\sum_{i=1}^n -\log \bar{p}_{x_i} + \log \bar{q}_{z_i} < nH(\rho_{AB}|\sigma_B) - n\delta] \\ &\qquad \leq \Pr_{\mathbf{x},\mathbf{z}}\big[w(\mathbf{x},\mathbf{z}) \leq -n\delta\big] \leq 2^{-\frac{n\delta^2}{2\log(\gamma+2)^2}} \leq \Big(\frac{\varepsilon}{2}\Big)^2 \ .\end{aligned}$$

Combining this with (3.34) implies (3.33) and thus concludes the proof. □

The following corollary specializes Theorem 3.3.6 to the case where the first part of the state $\rho_{AB} = \rho_{XB}$ is classical and where $\sigma_B = \rho_B$.

Corollary 3.3.7. *Let $\rho_{XB} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_B)$ be a density operator which is classical on $\mathcal{H}_X$. Then, for any $\varepsilon \geq 0$,*

$$\frac{1}{n} H_{\min}^{\varepsilon}(\rho_{XB}^{\otimes n}|\rho_B^{\otimes n}) \geq H(\rho_{XB}) - H(\rho_B) - \delta \ ,$$

where $\delta := \big(2H_{\max}(\rho_X) + 3\big)\sqrt{\frac{\log(1/\varepsilon)}{n} + 1}$.

Proof. Assume without loss of generality that ρ_B is invertible (the general statement then follows by continuity). Because the operator

$$\mathrm{id}_X \otimes \rho_B - \rho_{XB} = \sum_{x \in \mathcal{X}} \mathrm{id}_X \otimes \rho_B^x - |x\rangle\langle x| \otimes \rho_B^x$$

is nonnegative, we can apply Lemma B.5.4 which gives

$$\lambda_{\max}\bigl(\rho_{XB}^{1/2}(\mathrm{id}_X \otimes \rho_B^{-1})\rho_{XB}^{1/2}\bigr) \leq 1 \ .$$

Hence, since ρ_{XB} is normalized,

$$\mathrm{tr}\bigl(\rho_{XB}^2(\mathrm{id}_X \otimes \rho_B^{-1})\bigr) = \mathrm{tr}\bigl(\rho_{XB}\rho_{XB}^{1/2}(\mathrm{id}_X \otimes \rho_B^{-1})\rho_{XB}^{1/2}\bigr) \leq 1 \ .$$

Using the fact that, for any $a \geq 2$, $\log(a+3) \leq \log a + \frac{3}{2}$, we thus have

$$\begin{aligned}
\log\Bigl(\mathrm{rank}(\rho_X) + \mathrm{tr}\bigl(\rho_{XB}^2(\mathrm{id}_X \otimes \sigma_B^{-1})\bigr) + 2\Bigr) &\leq \log\bigl(\mathrm{rank}(\rho_X) + 3\bigr) \\
&\leq \log \mathrm{rank}(\rho_X) + \frac{3}{2} \\
&= H_{\max}(\rho_X) + \frac{3}{2} \ .
\end{aligned}$$

The assertion then follows directly from Theorem 3.3.6 with $\rho_{AB} := \rho_{XB}$ and $\sigma_B := \rho_B$. □

Chapter 4

Symmetric States

The state of an n-partite quantum system is said to be *symmetric* or *permutation-invariant* if it is unchanged under reordering of the subsystems. Such states have nice properties which are actually very similar to those of product states.

The chapter is organized as follows: We first review some basic properties of symmetric subspaces of product spaces (Section 4.1) and show that any permutation-invariant density operator has a purification in such a space (Section 4.2). Next, we state our main result on the structure of symmetric states, which generalizes the so-called *de Finetti representation theorem* (Section 4.3). Based on this result, we derive expressions for the smooth min-entropy (Section 4.4) and the measurement statistics (Section 4.5) of symmetric states.

4.1 Definition and basic properties

4.1.1 Symmetric subspace of $\mathcal{H}^{\otimes n}$

Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{S}_n$ be the set of permutations on $\{1, \ldots, n\}$. For any $\pi \in \mathcal{S}_n$, we denote by the same letter π the unitary operation on $\mathcal{H}^{\otimes n}$ which permutes the n subsystems, that is,

$$\pi(|\theta_1\rangle \otimes \cdots \otimes |\theta_n\rangle) := |\theta_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |\theta_{\pi^{-1}(n)}\rangle \ ,$$

for any $|\theta_1\rangle, \ldots, |\theta_n\rangle \in \mathcal{H}$.

Definition 4.1.1. Let $\mathcal{H}$ be a Hilbert space and let $n \geq 0$. The *symmetric subspace* $\mathrm{Sym}(\mathcal{H}^{\otimes n})$ *of* $\mathcal{H}^{\otimes n}$ is the subspace of $\mathcal{H}^{\otimes n}$ spanned by all vectors which are invariant under permutations of the subsystems, that is,

$$\mathrm{Sym}(\mathcal{H}^{\otimes n}) := \bigl\{|\Psi\rangle \in \mathcal{H}^{\otimes n} : \pi|\Psi\rangle = |\Psi\rangle\bigr\} \ .$$

Remark 4.1.2. For any $n', n'' \geq 0$,

$$\mathrm{Sym}(\mathcal{H}^{\otimes n'+n''}) \subseteq \mathrm{Sym}(\mathcal{H}^{\otimes n'}) \otimes \mathrm{Sym}(\mathcal{H}^{\otimes n''}) \ .$$

Lemma 4.1.3 below provides an alternative characterization of the symmetric subspace $\mathrm{Sym}(\mathcal{H}^{\otimes n})$.

Lemma 4.1.3. *Let $\mathcal{H}$ be a Hilbert space and let $n \geq 0$. Then*

$$\mathrm{Sym}(\mathcal{H}^{\otimes n}) = \mathrm{span}\{|\theta\rangle^{\otimes n} : |\theta\rangle \in \mathcal{H}\} .$$

Proof. For a proof of this statement, we refer to the standard literature on symmetric functions or representation theory (see, e.g., [WG00]). □

A basis of the symmetric subspace

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be an n-tuple of elements from $\mathcal{X}$. The *frequency distribution* $\lambda_{\mathbf{x}}$ of $\mathbf{x}$ is the probability distribution on $\mathcal{X}$ defined by the relative number of occurrences of each symbol, that is,

$$\lambda_{\mathbf{x}}(x) := \frac{1}{n}\big|\{i : x_i = x\}\big| ,$$

for any $x \in \mathcal{X}$. In the following, we denote by $\mathcal{Q}_n^{\mathcal{X}}$ the set of frequency distributions of n-tuples on $\mathcal{X}$, also called *types with denominator n on $\mathcal{X}$*. Moreover, for any type $Q \in \mathcal{Q}_n^{\mathcal{X}}$, we denote by Λ_n^Q the corresponding *type class*, i.e., the set of all n-tuples $\mathbf{x} = (x_1, \ldots, x_n)$ with frequency distribution $\lambda_{\mathbf{x}} = Q$.

Let $\{|x\rangle\}_{x\in\mathcal{X}}$ be an orthonormal basis of $\mathcal{H}$. For any $Q \in \mathcal{Q}_n^{\mathcal{X}}$, we define the vector $|\Theta^Q\rangle$ on $\mathrm{Sym}(\mathcal{H}^{\otimes n})$ by

$$|\Theta^Q\rangle := \frac{1}{\sqrt{|\Lambda_n^Q|}} \sum_{(x_1,\ldots,x_n)\in\Lambda_n^Q} |x_1\rangle \otimes \cdots \otimes |x_n\rangle , \tag{4.1}$$

where, according to Lemma (B.1.2), $|\Lambda_n^Q| = \frac{n!}{\prod_x (nQ(x))!}$.

The vectors $|\Theta^Q\rangle$, for $Q \in \mathcal{Q}_n^{\mathcal{X}}$, are mutually orthogonal and normalized. We will see below (cf. Lemma 4.1.5) that the family $\{|\Theta^Q\rangle\}_{Q\in\mathcal{Q}_n^{\mathcal{X}}}$ is a basis of $\mathrm{Sym}(\mathcal{H}^{\otimes n})$. In particular, if $\mathcal{H}$ has dimension d, then $\dim(\mathrm{Sym}(\mathcal{H}^{\otimes n})) = |\mathcal{Q}_n^{\mathcal{X}}| = \binom{n+d-1}{n}$ (cf. Lemma B.1.1).

4.1.2 Symmetric subspace along product states

Let $\mathcal{H}$ be a Hilbert space, let $|\theta\rangle \in \mathcal{H}$ be fixed, and let $0 \leq m \leq n$. We denote by $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ the set of vectors $|\Psi\rangle \in \mathcal{H}^{\otimes n}$ which, after some reordering of the subsystems, are of the form $|\theta\rangle^{\otimes m} \otimes |\tilde{\Psi}\rangle$, that is,

$$\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m}) := \{\pi(|\theta\rangle^{\otimes m} \otimes |\tilde{\Psi}\rangle) : \pi \in \mathcal{S}_n, |\tilde{\Psi}\rangle \in \mathcal{H}^{\otimes n-m}\} . \tag{4.2}$$

We will be interested in the subspace of $\mathrm{Sym}(\mathcal{H}^{\otimes n})$ which only consists of linear combinations of vectors from $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$.

Definition 4.1.4. Let $\mathcal{H}$ be a Hilbert space, let $|\theta\rangle \in \mathcal{H}$, and let $0 \leq m \leq n$. The *symmetric subspace* $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ *of* $\mathcal{H}^{\otimes n}$ *along* $|\theta\rangle^{\otimes m}$ is

$$\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m}) := \mathrm{Sym}(\mathcal{H}^{\otimes n}) \cap \operatorname{span} \mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m}) ,$$

where $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ denotes the subset of $\mathcal{H}^{\otimes n}$ defined by (4.2).

Note that $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m}) \subseteq \mathrm{Sym}(\mathcal{H}^{\otimes n})$, where equality holds if $m = 0$. In Section 4.4 and 4.5, we shall see that, if $r := n - m$ is small compared to n, then the states in $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ have similar properties as product states $|\theta\rangle^{\otimes n}$.

Lemma 4.1.5. *Let $\mathcal{H}$ be a Hilbert space with orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$, let $|\theta\rangle := |\bar{x}\rangle$ for some $\bar{x} \in \mathcal{X}$, and let $0 \leq m \leq n$. Then the family*

$$\mathcal{B} := \{|\Theta^Q\rangle\}_{Q\in\mathcal{Q}_n^{\mathcal{X}} : Q(\bar{x}) \geq \frac{m}{n}}$$

of vectors $|\Theta^Q\rangle$ defined by (4.1) *is an orthonormal basis of* $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$.

Note that, for $m = 0$, Lemma 4.1.5 implies that the family $\{|\Theta^Q\rangle\}_{Q\in\mathcal{Q}_n^{\mathcal{X}}}$ is an orthonormal basis of $\mathrm{Sym}(\mathcal{H}^{\otimes n})$.

Proof. For any $Q \in \mathcal{Q}_n^{\mathcal{X}}$, the vector $|\Theta^Q\rangle$ is invariant under permutations of the subsystems, that is, $|\Theta^Q\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n})$. Moreover, if $Q(\bar{x}) \geq \frac{m}{n}$ then the sum on the right hand side of (4.1) only runs over n-tuples which contain at least m symbols $\bar{x}$, that is, each term of the sum is contained in the set $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ defined by (4.2) and hence $|\Theta^Q\rangle \in \operatorname{span} \mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$. This proves that all vectors $|\Theta^Q\rangle \in \mathcal{B}$ are contained in $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$. Moreover, the vectors $|\Theta^Q\rangle$ are mutually orthogonal and normalized.

It remains to be shown that $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ is spanned by the vectors $|\Theta^Q\rangle \in \mathcal{B}$. Let thus $|\Psi\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ be fixed. Since $\{|x\rangle\}_{x\in\mathcal{X}}$ is a basis of $\mathcal{H}$, there exist coefficients $\alpha_{\mathbf{x}}$, for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, such that

$$|\Psi\rangle = \sum_{\mathbf{x}\in\mathcal{X}^n} \alpha_{\mathbf{x}} |x_1\rangle \otimes \cdots \otimes |x_n\rangle .$$

Because $|\Psi\rangle$ is invariant under permutations of the subsystems, the coefficients $\alpha_{\mathbf{x}}$ can only depend on the frequency distribution $\lambda_{\mathbf{x}}$. This implies that there exist coefficients β_Q such that

$$|\Psi\rangle = \sum_{Q\in\mathcal{Q}_n^{\mathcal{X}}} \beta_Q |\Theta^Q\rangle .$$

To conclude the proof, we need to verify that this sum can be restricted to frequency distributions Q such that $Q(\bar{x}) \geq \frac{m}{n}$. Observe that, for any $Q \in \mathcal{Q}_n^{\mathcal{X}}$ with $Q(\bar{x}) < \frac{m}{n}$, the vector $|\Theta^Q\rangle$ is orthogonal to any vector in $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ and thus also to any vector in $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$. The corresponding coefficient β_Q must thus be zero. □

Any vector $|\Psi\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ can be written as a linear combination of at most[1] $2^{nh(m/n)}$ vectors from the set $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ defined by (4.2).

Lemma 4.1.6. *Let* $|\Psi\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$. *Then there exists an orthonormal family* $\{|\Psi^s\rangle\}_{s\in\mathcal{S}}$ *of vectors from* $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$ *with cardinality* $|\mathcal{S}| \leq 2^{nh(m/n)}$ *such that* $|\Psi\rangle \in \mathrm{span}\{|\Psi^s\rangle\}_{s\in\mathcal{S}}$.

Proof. Let $\{|x\rangle\}_{x\in\mathcal{X}}$ be an orthonormal basis of $\mathcal{H}$ such that $|\bar{x}\rangle = |\theta\rangle$. For any n-tuple $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, we denote by $|\mathbf{x}\rangle$ the vector $|x_1\rangle \otimes \cdots \otimes |x_n\rangle$. Because $|\Psi\rangle \in \mathrm{span}\,\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$, there exist coefficients $\beta_{\mathbf{x}}$, for $\mathbf{x} \in \mathcal{X}^n$, such that

$$|\Psi\rangle = \sum_{\mathbf{x}\colon \lambda_{\mathbf{x}}(\bar{x}) \geq \frac{m}{n}} \beta_{\mathbf{x}} |\mathbf{x}\rangle \ . \tag{4.3}$$

Let $\mathcal{S}$ be the set of all subsets $s \subseteq \{1, \ldots, n\}$ of cardinality $|s| = m$. Moreover, for any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ with $\lambda_{\mathbf{x}}(\bar{x}) \geq \frac{m}{n}$, let $s(\mathbf{x}) \in \mathcal{S}$ be a set of m indices from $\{1, \ldots, n\}$ such that $i \in s(\mathbf{x}) \implies x_i = \bar{x}$. Finally, for any $s \in \mathcal{S}$, let

$$|\Psi^s\rangle := \sum_{\mathbf{x}\colon s(\mathbf{x})=s} \beta_{\mathbf{x}} |\mathbf{x}\rangle \ . \tag{4.4}$$

The sum in (4.3) can then be rewritten as $|\Psi\rangle = \sum_{s\in\mathcal{S}} |\Psi^s\rangle$, that is, $|\Psi\rangle \in \mathrm{span}\{|\Psi^s\rangle\}_{s\in\mathcal{S}}$. Moreover, Lemma B.1.3 implies $|\mathcal{S}| \leq 2^{nh(m/n)}$.

It remains to be shown that $\{|\Psi^s\rangle\}_{s\in\mathcal{S}}$ is an orthonormal family of vectors from $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$. Let thus $s \in \mathcal{S}$ be fixed and let π be a permutation such that $\pi(s) = \{1, \ldots, m\}$. Hence, for any $\mathbf{x}$ with $s(\mathbf{x}) = s$, the vector $\pi|\mathbf{x}\rangle$ has the form $|\theta\rangle^{\otimes m} \otimes |\tilde{\Psi}\rangle$, for some $|\tilde{\Psi}\rangle \in \mathcal{H}^{\otimes n-m}$. By the definition (4.4), the same holds for $\pi|\Psi^s\rangle$, i.e., $|\Psi^s\rangle \in \mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes m})$. Furthermore, because for distinct $s, s' \in \mathcal{S}$, the sum in (4.4) runs over disjoint sets of n-tuples $\mathbf{x}$, and because the vectors $|\mathbf{x}\rangle$ are mutually orthogonal, the states $|\Psi^{\mathbf{s}}\rangle$ are also mutually orthogonal. The assertion thus follows by normalizing the vectors $|\Psi^s\rangle$. □

4.2 Symmetric purification

An operator ρ_n on $\mathcal{H}^{\otimes n}$ is called *permutation-invariant* if $\pi\rho_n\pi^\dagger = \rho_n$, for any permutation $\pi \in \mathcal{S}_n$. For example, the pure state $\rho_n = |\Psi\rangle\langle\Psi|$, for some vector $|\Psi\rangle$ of the symmetric subspace of $\mathcal{H}^{\otimes n}$, is permutation-invariant. More generally, any mixture of symmetric pure states is permutation-invariant.

[1] h denotes the *binary Shannon entropy function* defined by $h(p) := -p\log(p) - (1-p)\log(1-p)$.

The converse, however, is not always true. Consider for example the fully mixed state ρ_2 on $\mathcal{H}^{\otimes 2}$ where $\dim(\mathcal{H}) = 2$. Because this operator can be written as $\rho_2 = \sigma^{\otimes 2}$, it is invariant under permutations. However, ρ_2 has rank 4, whereas the symmetric subspace of $\mathcal{H}^{\otimes 2}$ only has dimension 3. Consequently, ρ_2 is not a mixture of symmetric pure states.

Lemma 4.2.2 below establishes another connection between permutation-invariant operators and symmetric pure states. We show that any permutation-invariant operator ρ_n on $\mathcal{H}^{\otimes n}$ has a purification on the symmetric subspace of $(\mathcal{H} \otimes \mathcal{H})^{\otimes n}$.

To prove this result, we need a technical lemma which states that a fully entangled state on two subsystems is unchanged when the same unitary operation is applied to both subsystems.

Lemma 4.2.1. *Let $\{|x\rangle\}_{x\in\mathcal{X}}$ be an orthonormal family of vectors on a Hilbert space $\mathcal{H}$ and define*

$$|\Psi\rangle := \sum_{x\in\mathcal{X}} |x\rangle \otimes \overline{|x\rangle} \ ,$$

where, for any $x \in \mathcal{X}$, $\overline{|x\rangle}$ denotes the complex conjugate of $|x\rangle$ (with respect to some basis of $\mathcal{H}$). Let U be a unitary operation on the subspace spanned by $\{|x\rangle\}_{x\in\mathcal{X}}$ and let $\overline{U}$ be its complex conjugate. Then

$$(U \otimes \overline{U})|\Psi\rangle = |\Psi\rangle \ .$$

Proof. A simple calculation shows that, for any $x, x' \in \mathcal{X}$,

$$\begin{aligned} \big(\langle x| \otimes \overline{\langle x'|}\big)|\Psi\rangle &= \delta_{x,x'} \\ \big(\langle x| \otimes \overline{\langle x'|}\big)\big(U \otimes \overline{U}\big)|\Psi\rangle &= \delta_{x,x'} \ . \end{aligned}$$

The assertion follows because, obviously, $\{|x\rangle \otimes \overline{|x'\rangle}\}_{x,x'\in\mathcal{X}}$ is a basis of the subspace of $\mathcal{H} \otimes \mathcal{H}$ that contains $|\Psi\rangle$. □

Lemma 4.2.2. *Let $\rho_n \in \mathcal{P}(\mathcal{H}^{\otimes n})$ be permutation-invariant. Then there exists a purification of ρ_n on $\mathrm{Sym}((\mathcal{H} \otimes \mathcal{H})^{\otimes n})$.*

Proof. Let $\{|x\rangle\}_{x\in\mathcal{X}}$ be an (orthonormal) eigenbasis of ρ_n and let Λ be the set of eigenvalues of ρ_n. For any $\lambda \in \Lambda$, let $\mathcal{H}_\lambda$ be the corresponding eigenspace of ρ_n, i.e., $\rho_n|\phi\rangle = \lambda|\phi\rangle$, for any $|\phi\rangle \in \mathcal{H}_\lambda$.

Because ρ_n is invariant under permutations, we have $\pi^\dagger \rho_n \pi|\phi\rangle = \lambda|\phi\rangle$, for any $|\phi\rangle \in \mathcal{H}_\lambda$ and $\pi \in \mathcal{S}_n$. Applying the unitary operation π to both sides of this equality gives $\rho_n\pi|\phi\rangle = \lambda\pi|\phi\rangle$, that is, $\pi|\phi\rangle \in \mathcal{H}_\lambda$. This proves that the eigenspaces $\mathcal{H}_\lambda$ of ρ_n are invariant under permutations.

For any $|\phi\rangle \in \mathcal{H}^{\otimes n}$, we denote by $\overline{|\phi\rangle}$ the complex conjugate of $|\phi\rangle$ with respect to some product basis on $\mathcal{H}^{\otimes n}$. Moreover, for any eigenvalue $\lambda \in \Lambda$, let

$$|\Psi^\lambda\rangle := \sum_{x\in\mathcal{X}_\lambda} |x\rangle \otimes \overline{|x\rangle} \ ,$$

where $\mathcal{X}_\lambda := \{x \in \mathcal{X} : |x\rangle \in \mathcal{H}_\lambda\}$, i.e., $\{|x\rangle\}_{x \in \mathcal{X}_\lambda}$ is an orthonormal basis of the eigenspace $\mathcal{H}_\lambda$. Finally, we define the vector $|\Psi\rangle \in \mathcal{H}^{\otimes n} \otimes \mathcal{H}^{\otimes n}$ by

$$|\Psi\rangle := \sum_{\lambda \in \Lambda} \sqrt{\lambda} |\Psi^\lambda\rangle \ .$$

It is easy to verify that the operator obtained by taking the partial trace of $|\Psi\rangle\langle\Psi|$ satisfies

$$\mathrm{tr}_{\mathcal{H}^{\otimes n}}(|\Psi\rangle\langle\Psi|) = \sum_{\lambda \in \Lambda} \sum_{x \in \mathcal{X}_\lambda} \lambda |x\rangle\langle x| = \rho_n \ ,$$

i.e., $|\Psi\rangle\langle\Psi|$ is a purification of ρ_n. It thus remains to be shown that $|\Psi\rangle$ is symmetric.

Let $\pi \in \mathcal{S}_n$ be a fixed permutation. Note that its complex conjugate $\overline{\pi}$ is equal to π. (Recall that we defined the complex conjugate with respect to a product basis of $\mathcal{H}^{\otimes n}$.) Moreover, because π is unitary on $\mathcal{H}^{\otimes n}$ and, additionally, for any $\lambda \in \Lambda$, the subspace $\mathcal{H}_\lambda$ is invariant under π, the restriction of π to $\mathcal{H}_\lambda$ is unitary as well. Hence, by Lemma 4.2.1,

$$(\pi \otimes \pi)|\Psi^\lambda\rangle = (\pi \otimes \overline{\pi})|\Psi^\lambda\rangle = |\Psi^\lambda\rangle$$

and thus, by linearity,

$$(\pi \otimes \pi)|\Psi\rangle = \sum_{\lambda \in \Lambda} \sqrt{\lambda} (\pi \otimes \pi)|\Psi^\lambda\rangle = \sum_{\lambda \in \Lambda} \sqrt{\lambda} |\Psi^\lambda\rangle = |\Psi\rangle \ .$$

Because this holds for any permutation π on $\mathcal{H}^{\otimes n}$, we conclude $|\Psi\rangle \in \mathrm{Sym}((\mathcal{H} \otimes \mathcal{H})^{\otimes n})$. □

4.3 De Finetti representation

While any product state $\rho_n = \sigma^{\otimes n}$ on $\mathcal{H}^{\otimes n}$ is permutation-invariant, the converse is not true in general. Nevertheless, as we shall see, the properties of permutation-invariant states ρ_n are usually very similar to those of product states.

The quantum de Finetti representation theorem makes this connection explicit. In its basic version, it states that any density operator ρ_n on $\mathcal{H}^{\otimes n}$ which is *infinitely* exchangeable, i.e., ρ_n is the partial state of a permutation-invariant operator ρ_{n+k} on $n+k$ subsystems, for *all* $k \geq 0$, can be written as a mixture of product states $\sigma^{\otimes n}$.

In this section, we generalize the quantum de Finetti representation to the *finite* case, where ρ_n is only $(n+k)$-exchangeable, i.e., ρ_n is the partial state of a permutation-invariant operator ρ_{n+k} on $n+k$ subsystems, for some *fixed* $k \geq 0$. Theorem 4.3.2 below states that any pure density operator ρ_n

on $\mathcal{H}^{\otimes n}$ which is $(n+k)$-exchangeable is close to a mixture of states $\bar{\rho}_n^{|\theta\rangle}$ which have almost product form $|\theta\rangle^{\otimes n}$, for $|\theta\rangle \in \mathcal{H}$. More precisely, for any $|\theta\rangle$, $\bar{\rho}_n^{|\theta\rangle}$ is a pure state of the symmetric subspace of $\mathcal{H}^{\otimes n}$ along $|\theta\rangle^{\otimes n-r}$, for some small $r \geq 0$. Because of Lemma 4.2.2, this statement also holds for mixed states ρ_n.

The proof of Theorem 4.3.2 is based on the following lemma which states that the uniform mixture of product states $(|\theta\rangle\langle\theta|)^{\otimes n}$, for all normalized vectors $|\theta\rangle \in \mathcal{S}_1(\mathcal{H}) := \{|\theta\rangle \in \mathcal{H} : \||\theta\rangle\| = 1\}$, is equal to the fully mixed state on the symmetric subspace of $\mathcal{H}^{\otimes n}$.

Lemma 4.3.1. *Let $\mathcal{H}$ be a d-dimensional Hilbert space and let $n \geq 0$. Then*

$$\int_{\mathcal{S}_1(\mathcal{H})} (|\theta\rangle\langle\theta|)^{\otimes n} \omega(|\theta\rangle) = \binom{n+d-1}{n}^{-1} \cdot \mathrm{id}_{\mathrm{Sym}(\mathcal{H}^{\otimes n})} ,$$

where ω denotes the uniform probability measure on the unit sphere $\mathcal{S}_1(\mathcal{H})$.

Lemma 4.3.1 can be proven using techniques from representation theory, in particular, Schur's Lemma (see, e.g., [WG00]). In the following, however, we propose an alternative proof.

Proof. Let

$$T := \int_{\mathcal{S}_1(\mathcal{H})} (|\theta\rangle\langle\theta|)^{\otimes n} \omega(|\theta\rangle) .$$

We first show that $T = c \cdot \mathrm{id}_{\mathrm{Sym}(\mathcal{H}^{\otimes n})}$ for some constant c.

Because the space $\mathrm{Sym}(\mathcal{H}^{\otimes n})$ is spanned by vectors of the form $|\theta\rangle^{\otimes n}$ (cf. Lemma 4.1.3), it is sufficient to show that, for any $|u\rangle, |v\rangle \in \mathcal{S}_1(\mathcal{H})$,

$$\langle u|^{\otimes n} T |v\rangle^{\otimes n} = \langle u|^{\otimes n} c \cdot \mathrm{id}_{\mathrm{Sym}(\mathcal{H}^{\otimes n})} |v\rangle^{\otimes n} . \tag{4.5}$$

Let thus $|u\rangle, |v\rangle \in \mathcal{S}_1(\mathcal{H})$ be fixed and define $\alpha := \langle u|v\rangle$ and $|w\rangle := |v\rangle - \alpha|u\rangle$, i.e., $\langle u|w\rangle = 0$. Then

$$\begin{aligned} \langle u|^{\otimes n} T |v\rangle^{\otimes n} &= \int_{\mathcal{S}_1(\mathcal{H})} \langle u|\theta\rangle^n \langle\theta|v\rangle^n \omega(|\theta\rangle) \\ &= \int_{\mathcal{S}_1(\mathcal{H})} \langle u|\theta\rangle^n \big(\alpha\langle\theta|u\rangle + \langle\theta|w\rangle\big)^n \omega(|\theta\rangle) . \end{aligned} \tag{4.6}$$

Note that, for any $m \in \{0, \ldots, n\}$,

$$\int_{\mathcal{S}_1(\mathcal{H})} \langle u|\theta\rangle^n \langle\theta|u\rangle^{n-m} \langle\theta|w\rangle^m \omega(|\theta\rangle) = \int_{\mathcal{S}_1(\mathcal{H})} |\langle u|\theta\rangle|^{2(n-m)} \langle u|\theta\rangle^m \langle\theta|w\rangle^m \omega(|\theta\rangle) .$$

Because, for any fixed value of $\langle u|\theta\rangle$, the integral runs over all phases of $\langle\theta|w\rangle$ (recall that $|u\rangle$ and $|w\rangle$ are orthogonal) and because the probability measure ω is invariant under unitary operations, this expression equals zero for any $m > 0$. The integral on the right hand side of (4.6) can thus be rewritten as

$$\begin{aligned}\langle u|^{\otimes n} T |v\rangle^{\otimes n} &= \int_{\mathcal{S}_1(\mathcal{H})} \alpha^n |\langle u|\theta\rangle|^{2n} \omega(|\theta\rangle) \\ &= \langle u|v\rangle^n \int_{\mathcal{S}_1(\mathcal{H})} |\langle u|\theta\rangle|^{2n} \omega(|\theta\rangle) \ .\end{aligned} \tag{4.7}$$

Using again the fact that the probability measure ω is invariant under unitary operations, we conclude that the integral on the right hand side cannot depend on the vector $|u\rangle$, i.e., it is equal to a constant c. This implies (4.5) and thus proves that $T = c \cdot \mathrm{id}_{\mathrm{Sym}(\mathcal{H}^{\otimes n})}$.

To determine the value of c,[2] observe that

$$\mathrm{tr}(T) = \int_{\mathcal{S}_1(\mathcal{H})} \mathrm{tr}\big((|\theta\rangle\langle\theta|)^{\otimes n}\big)\omega(|\theta\rangle) = \int_{\mathcal{S}_1(\mathcal{H})} \omega(|\theta\rangle) = 1 \ , \tag{4.8}$$

where the last equality holds because ω is a probability measure on $\mathcal{S}_1(\mathcal{H})$. On the other hand, we have $\mathrm{tr}(T) = c \cdot \dim(\mathrm{Sym}(\mathcal{H}^{\otimes n}))$. Hence, $c^{-1} = \dim(\mathrm{Sym}(\mathcal{H}^{\otimes n})) = \binom{n+d-1}{n}$, which concludes the proof. □

We are now ready to state and prove a de Finetti style representation theorem. Note that Theorem 4.3.2 is restricted to *pure* symmetric states. The statement for general permutation-invariant states then follows because any such state has a symmetric purification (see Lemma 4.2.2).

Theorem 4.3.2. *Let ρ_{n+k} be a pure density operator on $\mathrm{Sym}(\mathcal{H}^{\otimes n+k})$ and let $0 \leq r \leq n$. Then there exists a measure ν on $\mathcal{S}_1(\mathcal{H})$ and, for each $|\theta\rangle \in \mathcal{S}_1(\mathcal{H})$, a pure density operator $\bar{\rho}_n^{|\theta\rangle}$ on $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$ such that*

$$\Big\| \mathrm{tr}_k(\rho_{n+k}) - \int_{\mathcal{S}_1(\mathcal{H})} \bar{\rho}_n^{|\theta\rangle} \, \nu(|\theta\rangle) \Big\|_1 \leq 2 e^{-\frac{k(r+1)}{2(n+k)} + \frac{1}{2}\dim(\mathcal{H})\ln k} \ .$$

Proof. Because the density operator ρ_{n+k} is pure, we have $\rho_{n+k} = |\Psi\rangle\langle\Psi|$ for some $|\Psi\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n+k})$. For any $|\theta\rangle \in \mathcal{S}_1(\mathcal{H})$, let

$$|\Psi^{|\theta\rangle}\rangle := \sqrt{\tbinom{k+d-1}{k}} \cdot \big(\mathrm{id}_{\mathcal{H}}^{\otimes n} \otimes \langle\theta|^{\otimes k}\big) \cdot |\Psi\rangle \ ,$$

where $d := \dim(\mathcal{H})$. Because $\mathrm{Sym}(\mathcal{H}^{\otimes n+k})$ is a subspace of $\mathrm{Sym}(\mathcal{H}^{\otimes n}) \otimes \mathrm{Sym}(\mathcal{H}^{\otimes k})$ (see Remark 4.1.2), $|\Psi^{|\theta\rangle}\rangle$ is contained in $\mathrm{Sym}(\mathcal{H}^{\otimes n})$. Let $\rho_n^{|\theta\rangle} :=$

[2] Alternatively, the constant c can be computed by an explicit evaluation of the integral on the right hand side of (4.7). Remarkably, this can be used to prove Lemma 4.1.3: Observe first that, by the arguments given in the proof, c^{-1} must be equal to the dimension of the space spanned by the vectors of the form $|\theta\rangle^{\otimes n}$. On the other hand, the explicit computation of c shows that c^{-1} equals $\binom{n+d-1}{n}$, which is the dimension of $\mathrm{Sym}(\mathcal{H}^{\otimes n})$. Because the space spanned by the vectors $|\theta\rangle^{\otimes n}$ is a subspace of $\mathrm{Sym}(\mathcal{H}^{\otimes n})$, it follows that these spaces are equal.

$|\Psi^{|\theta\rangle}\rangle\langle\Psi^{|\theta\rangle}|$, let $P^{|\theta\rangle}$ be the projector onto the subspace $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$, and define

$$\bar{\rho}_n^{|\theta\rangle} := \tfrac{1}{p(|\theta\rangle)} P^{|\theta\rangle} \rho_n^{|\theta\rangle} P^{|\theta\rangle} \ ,$$

where $p(|\theta\rangle) := \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle}P^{|\theta\rangle})$, i.e., $\bar{\rho}_n^{|\theta\rangle}$ is normalized and, because $\rho_n^{|\theta\rangle}$ has rank one, it is also pure. Finally, let ν be the measure defined by $\nu := p \cdot \omega$, where ω is the uniform probability measure on $\mathcal{S}_1(\mathcal{H})$. It then suffices to show that

$$\delta := \Bigl\|\mathrm{tr}_k(\rho_{n+k}) - \int_{\mathcal{S}_1(\mathcal{H})} P^{|\theta\rangle}\rho_n^{|\theta\rangle}P^{|\theta\rangle}\,\omega(|\theta\rangle)\Bigr\|_1 \leq 2e^{-\frac{k(r+1)}{2(n+k)}+\frac{1}{2}d\ln k} \ . \qquad (4.9)$$

By the definition of $\rho_n^{|\theta\rangle}$, we have

$$\rho_n^{|\theta\rangle} = |\Psi^{|\theta\rangle}\rangle\langle\Psi^{|\theta\rangle}| = \tbinom{k+d-1}{k} \cdot \mathrm{tr}_k\bigl(\mathrm{id}_{\mathcal{H}}^{\otimes n} \otimes (|\theta\rangle\langle\theta|)^{\otimes k} \cdot |\Psi\rangle\langle\Psi|\bigr) \ , \qquad (4.10)$$

and thus, by Lemma 4.3.1,

$$\begin{aligned}\int_{\mathcal{S}_1(\mathcal{H})} \rho_n^{|\theta\rangle}\,\omega(|\theta\rangle) &= \tbinom{k+d-1}{k} \cdot \int_{\mathcal{S}_1(\mathcal{H})} \mathrm{tr}_k\bigl(\mathrm{id}_{\mathcal{H}}^{\otimes n} \otimes (|\theta\rangle\langle\theta|)^{\otimes k} \cdot |\Psi\rangle\langle\Psi|\bigr)\,\omega(|\theta\rangle) \\ &= \mathrm{tr}_k\bigl(\mathrm{id}_{\mathcal{H}}^{\otimes n} \otimes \mathrm{id}_{\mathrm{Sym}(\mathcal{H}^{\otimes k})} \cdot |\Psi\rangle\langle\Psi|\bigr) \ .\end{aligned}$$

Since $\mathrm{Sym}(\mathcal{H}^{\otimes n+k})$ is a subspace of $\mathcal{H}^{\otimes n} \otimes \mathrm{Sym}(\mathcal{H}^{\otimes k})$, the vector $|\Psi\rangle$ is contained in $\mathcal{H}^{\otimes n} \otimes \mathrm{Sym}(\mathcal{H}^{\otimes k})$. The operation $\mathrm{id}_{\mathcal{H}}^{\otimes n} \otimes \mathrm{id}_{\mathrm{Sym}(\mathcal{H}^{\otimes k})}$ in the above expression thus leaves $|\Psi\rangle\langle\Psi|$ unchanged. Because $\mathrm{tr}_k(|\Psi\rangle\langle\Psi|) = \mathrm{tr}_k(\rho_{n+k})$, we conclude

$$\int_{\mathcal{S}_1(\mathcal{H})} \rho_n^{|\theta\rangle}\,\omega(|\theta\rangle) = \mathrm{tr}_k(\rho_{n+k}) \ . \qquad (4.11)$$

Using this representation of $\mathrm{tr}_k(\rho_{n+k})$ and the triangle inequality, the distance δ defined by (4.9) can be bounded by

$$\delta \leq \int_{\mathcal{S}_1(\mathcal{H})} \bigl\|\rho_n^{|\theta\rangle} - P^{|\theta\rangle}\rho_n^{|\theta\rangle}P^{|\theta\rangle}\bigr\|_1\,\omega(|\theta\rangle) \ .$$

Because the operators $P^{|\theta\rangle}$ are projectors, we can apply Lemma A.2.8 to bound the distance between $\rho_n^{|\theta\rangle}$ and $P^{|\theta\rangle}\rho_n^{|\theta\rangle}P^{|\theta\rangle}$, which gives

$$\delta \leq 2\int_{\mathcal{S}_1(\mathcal{H})} \sqrt{\mathrm{tr}(\rho_n^{|\theta\rangle})}\sqrt{\mathrm{tr}(\rho_n^{|\theta\rangle}) - \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle})}\,\omega(|\theta\rangle) \ .$$

To bound the integral on the right hand side, we use the Cauchy-Schwartz inequality for the scalar product defined by $\langle f|g\rangle := \int_{\mathcal{S}_1(\mathcal{H})} f(|\theta\rangle)g(|\theta\rangle)\,\omega(|\theta\rangle)$, i.e.,

$$\delta \leq 2\sqrt{\int_{\mathcal{S}_1(\mathcal{H})} \mathrm{tr}(\rho_n^{|\theta\rangle})\,\omega(|\theta\rangle)}\sqrt{\int_{\mathcal{S}_1(\mathcal{H})} \bigl(\mathrm{tr}(\rho_n^{|\theta\rangle}) - \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle})\bigr)\,\omega(|\theta\rangle)} \ .$$

Because of (4.11), the first integral on the right hand side equals $\mathrm{tr}(\rho_{n+k}) = 1$, that is,

$$\delta \leq 2\sqrt{\int_{\mathcal{S}_1(\mathcal{H})} \big(\mathrm{tr}(\rho_n^{|\theta\rangle}) - \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle})\big)\, \omega(|\theta\rangle)} \ . \tag{4.12}$$

Let $\bar{P}^{|\theta\rangle}$ be the projector orthogonal to $P^{|\theta\rangle}$, i.e., $\bar{P}^{|\theta\rangle} := \mathrm{id}_{\mathrm{Sym}(\mathcal{H}^{\otimes n})} - P^{|\theta\rangle}$. With (4.10), the term in the integral can be rewritten as

$$\begin{aligned} \mathrm{tr}(\rho_n^{|\theta\rangle}) - \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle}) &= \mathrm{tr}(\bar{P}^{|\theta\rangle}\rho_n^{|\theta\rangle}) \\ &= \tbinom{k+d-1}{k} \cdot \mathrm{tr}\big(\bar{P}^{|\theta\rangle} \otimes (|\theta\rangle\langle\theta|)^{\otimes k} \cdot |\Psi\rangle\langle\Psi|\big) \ . \end{aligned} \tag{4.13}$$

Let $|\theta\rangle \in \mathcal{H}$ be fixed and let $\{|x\rangle\}_{x\in\mathcal{X}}$ be an orthonormal basis of $\mathcal{H}$ with $|\bar{x}\rangle = |\theta\rangle$, for some $\bar{x} \in \mathcal{X}$. Moreover, for all frequency distributions $Q \in \mathcal{Q}_n^{\mathcal{X}}$ and $\bar{Q} \in \mathcal{Q}_{n+k}^{\mathcal{X}}$, let $|\Theta_n^Q\rangle$ and $|\bar{\Theta}_{n+k}^{\bar{Q}}\rangle$ be the vectors in $\mathrm{Sym}(\mathcal{H}^{\otimes n})$ and $\mathrm{Sym}(\mathcal{H}^{\otimes n+k})$, respectively, defined by (4.1).

According to Lemma 4.1.5, the family of vectors $|\Theta_n^Q\rangle$, for all $Q \in \mathcal{Q}_n^{\mathcal{X}}$, is an orthonormal basis of $\mathrm{Sym}(\mathcal{H}^{\otimes n})$. Moreover, the subfamily where $Q(\bar{x}) \geq \frac{n-r}{n}$ is a basis of $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$. Consequently, the projector $\bar{P}^{|\theta\rangle}$ on the space orthogonal to $\mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$ can be written as

$$\bar{P}^{|\theta\rangle} = \sum_{Q\colon Q(\bar{x}) < \frac{n-r}{n}} |\Theta_n^Q\rangle\langle\Theta_n^Q| \ .$$

Identity (4.13) then reads

$$\mathrm{tr}(\rho_n^{|\theta\rangle}) - \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle}) = \tbinom{k+d-1}{k} \sum_{Q\colon Q(\bar{x}) < \frac{n-r}{n}} \Big| \big(\langle\Theta_n^Q| \otimes \langle\theta|^{\otimes k}\big) \cdot |\Psi\rangle \Big|^2 \ . \tag{4.14}$$

Because the family of vectors $|\bar{\Theta}_{n+k}^{\bar{Q}}\rangle$, for $\bar{Q} \in \mathcal{Q}_{n+k}^{\mathcal{X}}$, is a basis of the symmetric subspace $\mathrm{Sym}(\mathcal{H}^{\otimes n+k})$ (see again Lemma 4.1.5) there exist coefficients $\alpha_{\bar{Q}}$ such that

$$|\Psi\rangle = \sum_{\bar{Q}} \alpha_{\bar{Q}} |\bar{\Theta}_{n+k}^{\bar{Q}}\rangle \ , \tag{4.15}$$

where the sum runs over all $\bar{Q} \in \mathcal{Q}_{n+k}^{\mathcal{X}}$.

It is easy to verify that, for any $Q \in \mathcal{Q}_n^{\mathcal{X}}$ and $\bar{Q} \in \mathcal{Q}_{n+k}^{\mathcal{X}}$, the scalar product $(\langle\Theta_n^Q| \otimes \langle\theta|^{\otimes k}) \cdot |\bar{\Theta}_{n+k}^{\bar{Q}}\rangle$ equals zero unless

$$(n+k)\bar{Q}(x) = \begin{cases} nQ(x) + k & \text{if } x = \bar{x} \\ nQ(x) & \text{otherwise} \end{cases} \tag{4.16}$$

holds for all $x \in \mathcal{X}$, in which case

$$\big(\langle\Theta_n^Q| \otimes \langle\theta|^{\otimes k}\big) \cdot |\bar{\Theta}_{n+k}^{\bar{Q}}\rangle = \sqrt{\frac{\frac{n!}{\prod_x (nQ(x))!}}{\frac{(n+k)!}{\prod_x ((n+k)\bar{Q}(x))!}}} = \sqrt{\frac{n!(nQ(\bar{x})+k)!}{(n+k)!(nQ(\bar{x}))!}} \ . \tag{4.17}$$

Let $Q \in \mathcal{Q}_n^{\mathcal{X}}$ with $Q(\bar{x}) < \frac{n-r}{n}$ and let $\bar{Q} \in \mathcal{Q}_{n+k}^{\mathcal{X}}$ such that (4.16) holds. Then, from (4.15) and (4.17),

$$\left|\left(\langle\Theta_n^Q| \otimes \langle\theta|^{\otimes k}\right) \cdot |\Psi\rangle\right|^2 = |\alpha_{\bar{Q}}|^2 \frac{n!(nQ(\bar{x}) + k)!}{(n+k)!(nQ(\bar{x}))!} \leq |\alpha_{\bar{Q}}|^2 D_{n,k,r}$$

where $D_{n,k,r} := \frac{n!(n+k-r-1)!}{(n+k)!(n-r-1)!}$. Note that $Q(\bar{x}) < \frac{n-r}{n}$ implies $\bar{Q}(\bar{x}) < \frac{n+k-r}{n+k}$. Consequently, from (4.14),

$$\mathrm{tr}(\rho_n^{|\theta\rangle}) - \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle}) \leq \binom{k+d-1}{k} \cdot D_{n,k,r} \sum_{\bar{Q}:\, \bar{Q}(\bar{x}) < \frac{n+k-r}{n+k}} |\alpha_{\bar{Q}}|^2 \leq \binom{k+d-1}{k} \cdot D_{n,k,r} \ ,$$

where the last inequality follows from the fact that $\sum_{\bar{Q}} |\alpha_{\bar{Q}}|^2 = \||\Psi\rangle\|^2 = \mathrm{tr}(\rho_{n+k}) = 1$. The term $D_{n,k,r}$ can be bounded by

$$\begin{aligned} D_{n,k,r} &= \frac{(n-r)(n-r+1)\cdots(n+k-r-1)}{(n+1)(n+2)\cdots(n+k)} \\ &\leq \left(\frac{n+k-r-1}{n+k}\right)^k \\ &= \left(1 - \frac{r+1}{n+k}\right)^k \ . \end{aligned}$$

Defining $\beta := \frac{r+1}{n+k}$ and using the fact that, for any $\beta \in [0,1]$, $(1-\beta)^{1/\beta} \leq e^{-1}$, we find

$$D_{n,k,r} \leq (1-\beta)^k = \left((1-\beta)^{1/\beta}\right)^{\beta k} \leq e^{-\beta k} \ .$$

Finally, because for any $k \geq 2$ (note that, for $k < 2$, the assertion is trivial) $\binom{k+d-1}{k} \leq k^d$, we have

$$\mathrm{tr}(\rho_n^{|\theta\rangle}) - \mathrm{tr}(P^{|\theta\rangle}\rho_n^{|\theta\rangle}) \leq k^d e^{-k\frac{r+1}{n+k}} \ .$$

Inserting this into (4.12), the bound (4.9) follows because $\omega(|\theta\rangle)$ is a probability measure on $\mathcal{S}_1(\mathcal{H})$. □

If the symmetric state ρ_{n+k} on $\mathrm{Sym}(\mathcal{H}^{\otimes n+k})$ has some additional structure then the set of states that contribute to the mixture in the expression of Theorem 4.3.2 can be restricted. Remark 4.3.3 below treats the case where the subspaces $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ are bipartite systems and where the partial state on $\mathcal{H}_A^{\otimes n+k}$ has product form.

Remark 4.3.3. Let $\mathcal{H} := \mathcal{H}_A \otimes \mathcal{H}_B$ be a bipartite Hilbert space, let $\rho_{A^{n+k}B^{n+k}}$ be a pure density operator on $\mathrm{Sym}(\mathcal{H}^{\otimes n})$ such that $\rho_{A^{n+k}} = \sigma_A^{\otimes n+k}$, let $0 \leq r \leq n$, and let ν be the measure defined by Theorem 4.3.2. Then, for any $\delta \geq 0$, the set

$$\overline{\Gamma^\delta} := \left\{|\theta\rangle \in \mathcal{S}_1(\mathcal{H}) : \|\mathrm{tr}_B(|\theta\rangle\langle\theta|) - \sigma_A\|_1 > \delta\right\}$$

has at most weight $\nu(\overline{\Gamma^\delta}) \leq e^{-\frac{1}{4}k\delta^2 + \dim(\mathcal{H})\ln k}$.

Proof. Let $|\Psi\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n+k})$ and $\rho_n^{|\theta\rangle} \in \mathcal{P}(\mathrm{Sym}(\mathcal{H}^{\otimes n}))$ as defined in the proof of Theorem 4.3.2. It then suffices to show that

$$\int_{\overline{\Gamma^\delta}} \mathrm{tr}(\rho_n^{|\theta\rangle})\omega(|\theta\rangle) \leq e^{-\frac{1}{4}k\delta^2 + d\ln k} \ , \tag{4.18}$$

where ω is the uniform probability measure on the unit sphere $\mathcal{S}_1(\mathcal{H})$ and $d := \dim(\mathcal{H})$.

Let $|\theta\rangle \in \overline{\Gamma^\delta}$ be fixed, i.e., $\|\mathrm{tr}_B(|\theta\rangle\langle\theta|) - \sigma_A\|_1 > \delta$. Then, by (4.10),

$$\begin{aligned}\mathrm{tr}(\rho_n^{|\theta\rangle}) &= \tbinom{k+d-1}{k} \cdot \mathrm{tr}\big(\mathrm{id}_{\mathcal{H}}^{\otimes n} \otimes (|\theta\rangle\langle\theta|)^{\otimes k} \cdot |\Psi\rangle\langle\Psi|\big) \\ &= \tbinom{k+d-1}{k} \cdot \mathrm{tr}\big((|\theta\rangle\langle\theta|)^{\otimes k} \cdot \rho_{A^k B^k}\big) \ ,\end{aligned}$$

where $\rho_{A^kB^k} := \mathrm{tr}_n(\rho_{A^{n+k}B^{n+k}}) = \mathrm{tr}_n(|\Psi\rangle\langle\Psi|)$. Since the fidelity cannot decrease when taking the partial trace (cf. Lemma A.1.5) we get

$$\begin{aligned}\mathrm{tr}\big((|\theta\rangle\langle\theta|)^{\otimes k}\rho_{A^kB^k}\big) &= F\big(\rho_{A^kB^k}, (|\theta\rangle\langle\theta|)^{\otimes k}\big)^2 \\ &\leq F\big(\rho_{A^k}, \mathrm{tr}_B(|\theta\rangle\langle\theta|)^{\otimes k}\big)^2 \\ &= F\big(\sigma_A^{\otimes k}, \mathrm{tr}_B(|\theta\rangle\langle\theta|)^{\otimes k}\big)^2 \\ &= F\big(\sigma_A, \mathrm{tr}_B(|\theta\rangle\langle\theta|)\big)^{2k} \ .\end{aligned}$$

Because, by Lemma A.2.4,

$$F\big(\sigma_A, \mathrm{tr}_B(|\theta\rangle\langle\theta|)\big)^2 \leq 1 - \frac{1}{4}\big\|\sigma_A - \mathrm{tr}_B(|\theta\rangle\langle\theta|)\big\|_1^2 < 1 - \frac{\delta^2}{4} \ ,$$

we conclude

$$\mathrm{tr}(\rho_n^{|\theta\rangle}) \leq \tbinom{k+d-1}{k} \cdot \big(1 - \frac{\delta^2}{4}\big)^k \leq k^d e^{k\ln(1-\frac{\delta^2}{4})} \leq e^{-\frac{1}{4}k\delta^2 + d\ln k} \ ,$$

where we have used $\ln(1-a) \leq -a$, for $a \in [0,1]$. Inequality (4.18) then follows because ω is a probability measure. □

4.4 Smooth min-entropy of symmetric states

Let $|\theta\rangle \in \mathcal{H}$, let $\mathcal{E}$ be a quantum operation from $\mathcal{H}$ to $\mathcal{H}_X \otimes \mathcal{H}_B$, and define $\rho_{X^nB^n} := \mathcal{E}^{\otimes n}(|\Psi\rangle\langle\Psi|)$, for $|\Psi\rangle := |\theta\rangle^{\otimes n}$. Obviously, $\rho_{X^nB^n}$ has product form, i.e., $\rho_{X^nB^n} = \sigma_{XB}^{\otimes n}$, where $\sigma_{XB} = \mathcal{E}(|\theta\rangle\langle\theta|)$. Hence, as demonstrated in Section 3.3 (Corollary 3.3.7), the smooth min-entropy of such a product state can be expressed in terms of the von Neumann entropy, that is,

$$\frac{1}{n}H_{\min}(\rho_{X^nB^n}|B^n) \gtrapprox H(\sigma_{XB}) - H(\sigma_B) \ . \tag{4.19}$$

Theorem 4.4.1 below states that this still holds if the product state $|\Psi\rangle := |\theta\rangle^{\otimes n}$ is replaced by a state in the symmetric subspace of $\mathcal{H}^{\otimes n}$ along $|\theta\rangle^{\otimes n-r}$, for some $r \ll n$.

Theorem 4.4.1. *Let $0 \leq r \leq \frac{1}{2}n$, let $|\theta\rangle \in \mathcal{H}$ and $|\Psi\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$ be normalized, and let $\mathcal{E}$ be a trace-preserving CPM from $\mathcal{H}$ to $\mathcal{H}_X \otimes \mathcal{H}_B$ which is classical on $\mathcal{H}_X$. Define $\rho_{X^n B^n} := \mathcal{E}^{\otimes n}(|\Psi\rangle\langle\Psi|)$ and $\sigma_{XB} := \mathcal{E}(|\theta\rangle\langle\theta|)$. Then, for any $\varepsilon \geq 0$,*

$$\frac{1}{n} H_{\min}^{\varepsilon}(\rho_{X^n B^n} | B^n) \geq H(\sigma_{XB}) - H(\sigma_B) - \delta \ ,$$

where $\delta := \bigl(\frac{5}{2} H_{\max}(\rho_X) + 4\bigr) \sqrt{\frac{2 \log(4/\varepsilon)}{n}} + h(r/n)$.

Proof. According to Lemma 4.1.6, there exists a family $\{|\Psi^s\rangle\}_{s \in \mathcal{S}}$ of orthonormal vectors from $\mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$ of size $|\mathcal{S}| \leq 2^{nh(r/n)}$ such that

$$|\Psi\rangle = \sum_{s \in \mathcal{S}} \gamma_s |\Psi^s\rangle \ , \tag{4.20}$$

where γ_s are coefficients with $\sum_{s \in \mathcal{S}} |\gamma_s|^2 = 1$.

Let $\{E_w\}_{w \in \mathcal{W}}$ be the family of operators from $\mathcal{H}$ to $\mathcal{H}_X \otimes \mathcal{H}_B$ defined by the CPM $\mathcal{E}$, i.e., $\mathcal{E}(\sigma) = \sum_{w \in \mathcal{W}} E_w \sigma E_w^\dagger$, for any operator σ on $\mathcal{H}$. Moreover, let $\mathcal{H}_W$ be a Hilbert space with orthonormal basis $\{|w\rangle\}_{w \in \mathcal{W}}$ and let U be the operator from $\mathcal{H}$ to $\mathcal{H}_X \otimes \mathcal{H}_B \otimes \mathcal{H}_W$ defined by

$$U := \sum_{w \in \mathcal{W}} E_w \otimes |w\rangle \ .$$

Because $\mathcal{E}$ is trace-preserving, i.e., $\sum_w E_w^\dagger E_w = \mathrm{id}_{\mathcal{H}}$, we have $U^\dagger U = \mathrm{id}_{\mathcal{H}}$, that is, U is unitary. Furthermore, for any operator σ on $\mathcal{H}$,

$$\mathrm{tr}_W(U \sigma U^\dagger) = \mathcal{E}(\sigma) \ . \tag{4.21}$$

Let $|\Phi\rangle := U^{\otimes n} |\Psi\rangle$ and, similarly, for any $s \in \mathcal{S}$, let $|\Phi^s\rangle := U^{\otimes n} |\Psi^s\rangle$. Then, using (4.20),

$$|\Phi\rangle = \sum_{s \in \mathcal{S}} \gamma_s |\Phi^s\rangle \ .$$

Because U is unitary and the vectors $|\Psi^s\rangle$ are orthonormal, the vectors $|\Phi^s\rangle$ are orthonormal as well. Moreover, using (4.21),

$$\rho_{X^n B^n} = \mathcal{E}^{\otimes n}(|\Psi\rangle\langle\Psi|) = \mathrm{tr}_{W^n}(U^{\otimes n} |\Psi\rangle\langle\Psi| (U^\dagger)^{\otimes n}) = \mathrm{tr}_{W^n}(|\Phi\rangle\langle\Phi|) \ .$$

Let $\tilde{\rho}^s_{X^n B^n} := \mathrm{tr}_{W^n}(|\Phi^s\rangle\langle\Phi^s|)$ and define the operator $\tilde{\rho}_{X^n B^n S}$ on $\mathcal{H}_X^{\otimes n} \otimes \mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_S$ by

$$\tilde{\rho}_{X^n B^n S} := \sum_{s \in \mathcal{S}} |\gamma_s|^2 \tilde{\rho}^s_{X^n B^n} \otimes |s\rangle\langle s| \ ,$$

where $\mathcal{H}_S$ is a Hilbert space with orthonormal basis $\{|s\rangle\}_{s \in \mathcal{S}}$. Lemma 3.2.11 then allows us to express the smooth min-entropy of $\rho_{X^n B^n}$ in terms of the smooth min-entropy of $\tilde{\rho}_{X^n B^n S}$. Moreover, by Lemma 3.2.8, the smooth min-entropy of $\tilde{\rho}_{X^n B^n S}$ is lower bounded by the min-entropy of the operators $\tilde{\rho}^s_{X^n B^n}$, that is,

$$\begin{aligned} H^{\varepsilon}_{\min}(\rho_{X^n B^n}|\tilde{\rho}_{B^n}) &\geq H^{\tilde{\varepsilon}}_{\min}(\tilde{\rho}_{X^n B^n S}|\tilde{\rho}_{B^n S}) - H_{\max}(\tilde{\rho}_S) \\ &\geq \min_{s \in \mathcal{S}} H^{\tilde{\varepsilon}}_{\min}(\tilde{\rho}^s_{X^n B^n}|\tilde{\rho}^s_{B^n}) - H_{\max}(\tilde{\rho}_S) \ , \end{aligned}$$

where $\tilde{\varepsilon} = \frac{\varepsilon^2}{6|\mathcal{S}|}$. Using the fact that $|\mathcal{S}| \leq 2^{nh(r/n)}$, we find

$$H^{\varepsilon}_{\min}(\rho_{X^n B^n}|\tilde{\rho}_{B^n}) \geq \min_{s \in \mathcal{S}} H^{\tilde{\varepsilon}}_{\min}(\tilde{\rho}^s_{X^n B^n}|\tilde{\rho}^s_{B^n}) - nh(r/n) \tag{4.22}$$

and

$$\log(1/\tilde{\varepsilon}) \leq \log(2/\varepsilon) + \log 6 + nh(r/n) \ . \tag{4.23}$$

Let us now compute the min-entropies of the operators $\tilde{\rho}^s_{X^n B^n}$, for $s \in \mathcal{S}$. Since $|\Psi^s\rangle \in \mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$, the vector $|\Psi^s\rangle$, after some appropriate reordering of the subsystems, has the form $|\Psi^s\rangle = |\theta\rangle^{\otimes n-r} \otimes |\hat{\Psi}^s\rangle$, for some $|\hat{\Psi}^s\rangle \in \mathcal{H}^{\otimes r}$. Hence, the same holds for the vector $|\Phi^s\rangle$, i.e.,

$$|\Phi^s\rangle = U^{\otimes n}|\Psi^s\rangle = \bigl(U|\theta\rangle\bigr)^{\otimes n-r} \otimes U^{\otimes r}|\hat{\Psi}^s\rangle \ .$$

Consequently, from (4.21) and the definition of σ_{XB},

$$\begin{aligned} \tilde{\rho}^s_{X^n B^n} &= \bigl(\mathrm{tr}_W(U|\theta\rangle\langle\theta|U^\dagger)\bigr)^{\otimes n-r} \otimes \mathrm{tr}_{W^r}\bigl(U^{\otimes r}|\hat{\Psi}^s\rangle\langle\hat{\Psi}^s|(U^\dagger)^{\otimes r}\bigr) \\ &= \sigma^{\otimes n-r}_{XB} \otimes \hat{\rho}^s_{X^r B^r} \ , \end{aligned}$$

where $\hat{\rho}^s_{X^r B^r} := \mathcal{E}^{\otimes r}(|\hat{\Psi}^s\rangle\langle\hat{\Psi}^s|)$. Because $\mathcal{E}$ is classical on $\mathcal{H}_X$, $\hat{\rho}_{X^r B^r}$ is also classical on $\mathcal{H}^{\otimes r}_X$. Using the superadditivity of the smooth min-entropy (Lemma 3.2.6) and the fact that the min-entropy of a classical subsystem cannot be negative (Lemma 3.1.9) we find

$$\begin{aligned} H^{\tilde{\varepsilon}}_{\min}(\tilde{\rho}^s_{X^n B^n}|\tilde{\rho}^s_{B^n}) &\geq H^{\tilde{\varepsilon}}_{\min}\bigl(\sigma^{\otimes n-r}_{XB}\big|\sigma^{\otimes n-r}_{B}\bigr) + H_{\min}(\hat{\rho}^s_{X^r B^r}|\hat{\rho}_{B^r}) \\ &\geq H^{\tilde{\varepsilon}}_{\min}\bigl(\sigma^{\otimes n-r}_{XB}\big|\sigma^{\otimes n-r}_{B}\bigr) \ . \end{aligned} \tag{4.24}$$

Furthermore, because σ_{XB} is classical on $\mathcal{H}_X$, we can use Corollary 3.3.7 to bound the smooth min-entropy of the product state in terms of the von Neumann entropy,

$$\begin{aligned} H^{\tilde{\varepsilon}}_{\min}\bigl(\sigma^{\otimes n-r}_{XB}\big|\sigma^{\otimes n-r}_{B}\bigr) &\geq (n-r)\bigl(H(\sigma_{XB}) - H(\sigma_B) - \delta'\bigr) \\ &\geq n\bigl(H(\sigma_{XB}) - H(\sigma_B)\bigr) - rH_{\max}(\rho_X) - (n-r)\delta' \end{aligned}$$

with $\delta' := \bigl(2H_{\max}(\rho_X) + 3\bigr)\sqrt{\frac{\log(1/\tilde{\varepsilon})+1}{n-r}}$. Together with (4.22) and (4.24) we conclude

$$\frac{1}{n}H^{\varepsilon}_{\min}(\rho_{X^n B^n}|\tilde{\rho}_{B^n}) \geq H(\sigma_{XB}) - H(\sigma_B) - h(r/n) - r/nH_{\max}(\rho_X) - \frac{n-r}{n}\cdot\delta' \ . \tag{4.25}$$

Moreover, from (4.23),

$$\sqrt{n-r}\cdot\delta' \leq \bigl(2H_{\max}(\rho_X)+3\bigr)\sqrt{\log(2/\varepsilon)+nh(r/n)+\log 6+1}\ ,$$

and hence, using the fact that $c \leq \sqrt{c}$, for any $c \leq 1$,

$$\frac{n-r}{n}\cdot\delta' \leq \sqrt{\frac{n-r}{n}}\cdot\delta' \leq \bigl(2H_{\max}(\rho_X)+3\bigr)\sqrt{\frac{\log(2/\varepsilon)+4}{n}+h(r/n)}\ .$$

Finally, because $\frac{2r}{n} \leq h(r/n)$ and $h(r/n) \leq \sqrt{h(r/n)}$, we find

$$\begin{aligned} h(r/n)+\frac{r}{n}H_{\max}(\rho_X)+\frac{n-r}{n}\cdot\delta' & \\ &\leq \bigl(\frac{5}{2}H_{\max}(\rho_X)+4\bigr)\sqrt{\frac{2\log(2/\varepsilon)+4}{n}+h(r/n)}\ . \end{aligned}$$

Inserting this into (4.25) concludes the proof. □

4.5 Statistics of symmetric states

Let $z_1,\ldots,z_n$ be the outcomes of n independent measurements of a state $|\theta\rangle \in \mathcal{H}$ with respect to a POVM $\mathcal{M}=\{M_z\}_{z\in\mathcal{Z}}$. The law of large numbers tells us that, for large n, the statistics $\lambda_{\mathbf{z}}$ of the n-tuple $\mathbf{z}=(z_1,\ldots,z_n)$ is close to the probability distribution P_Z defined by $P_Z(z) := \mathrm{tr}(M_z|\theta\rangle\langle\theta|)$, for $z \in \mathcal{Z}$. Theorem 4.5.2 below states that the same is true if the n-tuple $\mathbf{z}$ is the outcome of a product measurement $\mathcal{M}^{\otimes n}$ applied to a state $|\Psi\rangle$ of the symmetric subspace of $\mathcal{H}^{\otimes n}$ along $|\theta\rangle^{\otimes n-r}$, for some small $r \ll n$.

For the proof of this result, we need the following technical lemma.

Lemma 4.5.1. *Let $|\psi\rangle = \sum_{x\in\mathcal{X}}|\psi^x\rangle$ and let $\rho \in \mathcal{P}(\mathcal{H})$. Then*

$$\langle\psi|\rho|\psi\rangle \leq |\mathcal{X}|\sum_{x\in\mathcal{X}}\langle\psi^x|\rho|\psi^x\rangle\ .$$

Proof. Let $\rho = \sum_{y\in\mathcal{Y}} p_y|y\rangle\langle y|$ be a spectral decomposition of ρ. For any $y \in \mathcal{Y}$,

$$|\langle y|\psi\rangle|^2 = \bigl|\sum_{x\in\mathcal{X}}\langle y|\psi^x\rangle\bigr|^2 \leq \bigl(\sum_{x\in\mathcal{X}}|\langle y|\psi^x\rangle|\bigr)^2 \leq |\mathcal{X}|\sum_{x\in\mathcal{X}}|\langle y|\psi^x\rangle|^2\ ,$$

where we have used the Cauchy-Schwartz inequality in the last step. Consequently,

$$\begin{aligned}\langle\psi|\rho|\psi\rangle &= \sum_{y\in\mathcal{Y}} p_y |\langle y|\psi\rangle|^2 \\ &\leq |\mathcal{X}| \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} p_y |\langle y|\psi^x\rangle|^2 \\ &= |\mathcal{X}| \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_y \langle\psi^x|y\rangle\langle y|\psi^x\rangle \\ &= |\mathcal{X}| \sum_{x\in\mathcal{X}} \langle\psi^x|\rho|\psi^x\rangle \ . \end{aligned}$$

□

Theorem 4.5.2. *Let $0 \leq r \leq \frac{1}{2}n$, let $|\theta\rangle \in \mathcal{H}$ and $|\Psi\rangle \in \mathrm{Sym}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$ be normalized, let $\mathcal{M} = \{M_z\}_{z\in\mathcal{Z}}$ be a POVM on $\mathcal{H}$, and let P_Z be the probability distribution of the outcomes of the measurement $\mathcal{M}$ applied to $|\theta\rangle\langle\theta|$. Then*

$$\Pr_{\mathbf{z}}\Big[\|\lambda_{\mathbf{z}} - P_Z\|_1 > 2\sqrt{\frac{\log(1/\varepsilon)}{n} + h(r/n) + \frac{|\mathcal{Z}|}{n}\log(\frac{n}{2}+1)}\Big] \leq \varepsilon \ ,$$

where the probability is taken over the outcomes $\mathbf{z} = (z_1, \dots, z_n)$ of the product measurement $\mathcal{M}^{\otimes n}$ applied to $|\Psi\rangle\langle\Psi|$.

Proof. According to Lemma 4.1.6, the vector $|\Psi\rangle$ can be written as a superposition of orthonormal vectors $|\Psi^s\rangle \in \mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$, that is,

$$|\Psi\rangle = \sum_{s\in\mathcal{S}} \gamma_s |\Psi^s\rangle \ , \tag{4.26}$$

where $\mathcal{S}$ is a set of size $|\mathcal{S}| \leq 2^{nh(\frac{r}{n})}$ and where γ_s are coefficients such that $\sum_{s\in\mathcal{S}} |\gamma_s|^2 = 1$.

Let now $s \in \mathcal{S}$ be fixed. Because $|\Psi^s\rangle \in \mathcal{V}(\mathcal{H}^{\otimes n}, |\theta\rangle^{\otimes n-r})$, there exists a permutation π which maps $|\Psi^s\rangle$ to a vector which, on the first $n-r$ subsystems, has the form $|\theta\rangle^{\otimes n-r}$. We can thus assume without loss of generality that $|\Psi^s\rangle = |\theta\rangle^{\otimes n-r} \otimes |\hat{\Psi}\rangle$, for some $|\hat{\Psi}\rangle \in \mathcal{H}^{\otimes r}$.

Let $\mathbf{z} = (z_1, \dots, z_n)$ be the outcome of the measurement $\mathcal{M}^{\otimes n}$ applied to $|\Psi^s\rangle\langle\Psi^s|$ and define $\mathbf{z}' := (z_1, \dots, z_{n-r})$ and $\mathbf{z}'' := (z_{n-r+1}, \dots, z_n)$. Clearly, $\mathbf{z}'$ is distributed according to the product distribution P_Z^{n-r}. Hence, with high probability, $\mathbf{z}'$ is a typical sequence, that is, by Corollary B.3.3,

$$\Pr_{\mathbf{z}'}\Big[\|\lambda_{\mathbf{z}'} - P_Z\|_1 > \sqrt{2(\ln 2)\big(\delta + \frac{|\mathcal{Z}|\log(n-r+1)}{n-r}\big)}\Big] \leq 2^{-(n-r)\delta} \ , \tag{4.27}$$

for any $\delta \geq 0$. Moreover, because $\lambda_{\mathbf{z}} = \frac{n-r}{n}\lambda_{\mathbf{z}'} + \frac{r}{n}\lambda_{\mathbf{z}''}$, we can apply the triangle inequality which gives

$$\big\|\lambda_{\mathbf{z}} - P_Z\big\|_1 \leq \frac{n-r}{n}\big\|\lambda_{\mathbf{z}'} - P_Z\big\|_1 + \frac{r}{n}\big\|\lambda_{\mathbf{z}''} - P_Z\big\|_1 \leq \big\|\lambda_{\mathbf{z}'} - P_Z\big\|_1 + \frac{r}{n} \ .$$

Using this inequality and the assumption $r \leq \frac{1}{2}n$, (4.27) implies that

$$\Pr_{\mathbf{z} \leftarrow |\Psi^s\rangle}[\mathbf{z} \in \mathcal{W}_\delta] \leq 2^{-\frac{n\delta}{2}} , \tag{4.28}$$

where we write $\mathbf{z} \leftarrow |\Psi^s\rangle$ to indicate that $\mathbf{z}$ is distributed according to the outcomes of the measurement applied to $|\Psi^s\rangle$ and where $\mathcal{W}_\delta$ is the subset of $\mathcal{Z}^n$ defined by

$$\mathcal{W}_\delta := \big\{\mathbf{z} \in \mathcal{Z}^n : \|\lambda_{\mathbf{z}} - P_Z\|_1 > \sqrt{2(\ln 2)\big(\delta + \frac{2|\mathcal{Z}|}{n}\log(\frac{n}{2}+1)\big)} + \frac{r}{n}\big\} .$$

Let $M_{\mathbf{z}} := M_{z_1} \otimes \cdots \otimes M_{z_n}$, for $\mathbf{z} = (z_1, \ldots, z_n) \in \mathcal{Z}^n$, be the linear operators defined by the POVM $\mathcal{M}^{\otimes n}$. Then, using Lemma 4.5.1, (4.26), and (4.28) we get

$$\begin{aligned}
\Pr_{\mathbf{z} \leftarrow |\Psi\rangle}[\mathbf{z} \in \mathcal{W}_\delta] &= \sum_{\mathbf{z} \in \mathcal{W}_\delta} \langle \Psi | M_{\mathbf{z}} | \Psi \rangle \\
&\leq \sum_{\mathbf{z} \in \mathcal{W}_\delta} |\mathcal{S}| \sum_{s \in \mathcal{S}} |\gamma_s|^2 \langle \Psi^s | M_{\mathbf{z}} | \Psi^s \rangle \\
&= |\mathcal{S}| \sum_{s \in \mathcal{S}} |\gamma_s|^2 \Pr_{\mathbf{z} \leftarrow |\Psi^s\rangle}[\mathbf{z} \in \mathcal{W}_\delta] \\
&\leq 2^{nh(r/n)} 2^{-\frac{n\delta}{2}} \\
&= 2^{-n(\frac{\delta}{2} - h(r/n))} .
\end{aligned}$$

Hence, with $\delta := \frac{2\log(1/\varepsilon)}{n} + 2h(r/n)$,

$$\Pr_{\mathbf{z}}\Big[\|\lambda_{\mathbf{z}} - P_Z\|_1 > \sqrt{4(\ln 2)\big(\frac{\log(1/\varepsilon)}{n} + h(r/n) + \frac{|\mathcal{Z}|}{n}\log(\frac{n}{2}+1)\big)} + \frac{r}{n}\Big] \leq \varepsilon .$$

The assertion then follows from the fact that $\sqrt{c} + \frac{r}{n} \leq \sqrt{c + \frac{2r}{n}}$, for any $c \geq 0$ with $c + \frac{2r}{n} \leq 1$, and from $\frac{2r}{n} \leq h(r/n)$. □

Chapter 5

Privacy Amplification

A fundamental problem in cryptography is to distill a secret key from only partially secret data, on which an adversary might have information encoded into the state of a quantum system. In this chapter, we propose a general solution to this problem, which is called *privacy amplification*: We show that the key computed as the output of a hash function (chosen at random from a two-universal[1] family of functions) is secure under the sole condition that its length is smaller than the adversary's uncertainty on the input, measured in terms of (smooth) min-entropy.

We start with the derivation of various technical results (Sections 5.1–5.4). These are used for the proof of the main statement, which is first formulated in terms of min-entropy (Section 5.5) and then generalized to *smooth* min-entropy (Section 5.6).

5.1 Bounding the norm of hermitian operators

In this section, we derive an upper bound on the trace norm for hermitian operators (Lemma 5.1.3). The bound only involves matrix multiplications, which makes it easy to evaluate.

Lemma 5.1.1. *Let S and T be hermitian operators on $\mathcal{H}$. Then*

$$\mathrm{tr}(ST) \leq \sqrt{\mathrm{tr}(S^2)\mathrm{tr}(T^2)} .$$

Proof. Let $S = \sum_{y \in \mathcal{Y}} \beta_y |y\rangle\langle y|$ and $T = \sum_{z \in \mathcal{Z}} \gamma_z |z\rangle\langle z|$ be spectral decompositions of S and T, respectively. With the definition $a_{y,z} := |\langle y|z\rangle|^2$, we have

$$\mathrm{tr}(ST) = \sum_{y,z} \beta_y \gamma_z \mathrm{tr}\big(|y\rangle\langle y| \cdot |z\rangle\langle z|\big) = \sum_{y,z} \beta_y \gamma_z a_{y,z} .$$

[1]See Section 5.4 for a definition.

On the other hand, $\mathrm{tr}(S^2) = \sum_y \beta_y^2$ and $\mathrm{tr}(T^2) = \sum_z \gamma_z^2$. It thus suffices to show that

$$\sum_{y,z} \beta_y \gamma_z a_{y,z} \leq \sqrt{\Bigl(\sum_y \beta_y^2\Bigr)\Bigl(\sum_z \gamma_z^2\Bigr)} \ . \tag{5.1}$$

It is easy to verify that $(a_{y,z})_{y \in \mathcal{Y}, z \in \mathcal{Z}}$ is a bistochastic matrix. Hence, according to Birkhoff's theorem (cf. Theorem B.2.2) there exist nonnegative coefficients μ_π parameterized by the bijections π from $\mathcal{Z}$ to $\mathcal{Y}$ such that $\sum_\pi \mu_\pi = 1$ and, for any $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, $a_{y,z} = \sum_\pi \mu_\pi \delta_{y,\pi(z)}$. We thus have

$$\sum_{y,z} \beta_y \gamma_z a_{y,z} = \sum_\pi \mu_\pi \sum_{y,z} \beta_y \gamma_z \delta_{y,\pi(z)} \ . \tag{5.2}$$

Furthermore, by the Cauchy-Schwartz inequality, for any fixed bijection π,

$$\sum_z \beta_{\pi(z)} \gamma_z \leq \sqrt{\Bigl(\sum_z \beta_{\pi(z)}^2\Bigr)\Bigl(\sum_z \gamma_z^2\Bigr)} \ .$$

This can be rewritten as

$$\sum_{y,z} \beta_y \gamma_z \delta_{y,\pi(z)} \leq \sqrt{\Bigl(\sum_y \beta_y^2\Bigr)\Bigl(\sum_z \gamma_z^2\Bigr)} \ .$$

Inserting this into (5.2) implies (5.1) and thus concludes the proof. □

Lemma 5.1.2. *Let S be a hermitian operator on $\mathcal{H}$ and let σ be a nonnegative operator on $\mathcal{H}$. Then*

$$\mathrm{tr}|\sqrt{\sigma} S \sqrt{\sigma}| \leq \sqrt{\mathrm{tr}(S^2)\mathrm{tr}(\sigma^2)} \ .$$

Proof. Let $\{|v\rangle\}_{v \in \mathcal{V}}$ be an eigenbasis of $\sqrt{\sigma} S \sqrt{\sigma}$ and let $S = \sum_{x \in \mathcal{X}} \alpha_x |x\rangle\langle x|$ be a spectral decomposition of S. Then

$$\begin{aligned} \mathrm{tr}|\sqrt{\sigma} S \sqrt{\sigma}| &= \sum_v \bigl|\langle v|\sqrt{\sigma} S \sqrt{\sigma}|v\rangle\bigr| \\ &= \sum_v \Bigl|\sum_x \alpha_x \langle v|\sqrt{\sigma}|x\rangle\langle x|\sqrt{\sigma}|v\rangle\Bigr| \\ &\leq \sum_v \sum_x |\alpha_x| \langle v|\sqrt{\sigma}|x\rangle\langle x|\sqrt{\sigma}|v\rangle \\ &= \sum_v \langle v|\sqrt{\sigma}|S|\sqrt{\sigma}|v\rangle \\ &= \mathrm{tr}(\sqrt{\sigma}|S|\sqrt{\sigma}) \ . \end{aligned}$$

Furthermore, by Lemma 5.1.1,

$$\mathrm{tr}(\sqrt{\sigma}|S|\sqrt{\sigma}) = \mathrm{tr}(|S|\sigma) \leq \sqrt{\mathrm{tr}(|S|^2)\mathrm{tr}(\sigma^2)} = \sqrt{\mathrm{tr}(S^2)\mathrm{tr}(\sigma^2)} \ ,$$

which concludes the proof. □

Lemma 5.1.3. *Let S be a hermitian operator on $\mathcal{H}$ and let σ be a nonnegative operator on $\mathcal{H}$. Then*

$$\|S\|_1 \leq \sqrt{\mathrm{tr}(\sigma)\mathrm{tr}(S\sigma^{-1/2}S\sigma^{-1/2})} \ .$$

Proof. The assertion follows directly from Lemma 5.1.2 with $\bar{\sigma} := \sqrt{\sigma}$ and $\bar{S} := \bar{\sigma}^{-1/2}S\bar{\sigma}^{-1/2}$, that is, $\sigma = \bar{\sigma}^2$ and $S = \sqrt{\bar{\sigma}}\bar{S}\sqrt{\bar{\sigma}}$. □

5.2 Distance from uniform

According to the discussion on universal security in Section 2.2.2, the security of a key is defined with respect to its L_1-distance from a perfect key which is uniformly distributed and independent of the adversary's state (see (2.6)). This motivates the following definition.

Definition 5.2.1. Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then the L_1*-distance from uniform of ρ_{AB} given B* is

$$d(\rho_{AB}|B) := \big\|\rho_{AB} - \rho_U \otimes \rho_B\big\|_1 \ ,$$

where $\rho_U := \frac{1}{\dim(\mathcal{H}_A)}\mathrm{id}_A$ is the fully mixed state on $\mathcal{H}_A$.

For an operator ρ_{XZ} defined by a classical probability distribution P_{XZ}, $d(\rho_{XZ}|Z)$ is the expectation (over z chosen according to P_Z) of the L_1-distance between the conditional distribution $P_{X|Z=z}$ and the uniform distribution. This property is generalized by the following lemma.

Lemma 5.2.2. *Let ρ_{ABZ} be classical with respect to an orthonormal basis $\{|z\rangle\}_{z\in\mathcal{Z}}$ of $\mathcal{H}_Z$ and let ρ_{AB}^z, for $z \in \mathcal{Z}$, be the corresponding (non-normalized) conditional operators. Then*

$$d(\rho_{ABZ}|BZ) = \sum_{z\in\mathcal{Z}} d(\rho_{AB}^z|B) \ .$$

Proof. Let ρ_U be the fully mixed state on $\mathcal{H}_A$. Then, by Lemma A.2.2,

$$\begin{aligned} d(\rho_{ABZ}|BZ) &= \|\rho_{ABZ} - \rho_U \otimes \rho_{BZ}\|_1 \\ &= \sum_{z\in\mathcal{Z}} \|\rho_{AB}^z - \rho_U \otimes \rho_B^z\|_1 \\ &= \sum_{z\in\mathcal{X}} d(\rho_{AB}^z|B) \ . \end{aligned}$$

□

To derive our result on the security of privacy amplification, it is convenient to consider an alternative measure for the distance from uniform. Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$ and $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$. The *(conditional) L_2-distance from uniform of ρ_{AB} relative to σ_B* is defined by

$$d_2(\rho_{AB}|\sigma_B) := \mathrm{tr}\Big(\big((\rho_{AB} - \rho_U \otimes \rho_B)(\mathrm{id}_A \otimes \sigma_B^{-1/2})\big)^2\Big) ,$$

where ρ_U is the fully mixed state on $\mathcal{H}_A$. Note that $d_2(\rho_{AB}|\sigma_B)$ can equivalently be written as

$$d_2(\rho_{AB}|\sigma_B) = \mathrm{tr}\Big(\big((\mathrm{id}_A \otimes \sigma_B^{-1/4})(\rho_{AB} - \rho_U \otimes \rho_B)(\mathrm{id}_A \otimes \sigma_B^{-1/4})\big)^2\Big) , \quad (5.3)$$

which proves that $d_2(\rho_{AB}|\sigma_B)$ cannot be negative.

The L_2-distance from uniform can be used to bound the L_1-distance from uniform.

Lemma 5.2.3. *Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then, for any $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$,*

$$d(\rho_{AB}|B) \leq \sqrt{\dim(\mathcal{H}_A)\mathrm{tr}(\sigma_B)d_2(\rho_{AB}|\sigma_B)} .$$

Proof. The assertion follows directly from Lemma 5.1.3 with $S := \rho_{AB} - \rho_U \otimes \rho_B$ and $\sigma := \mathrm{id}_A \otimes \sigma_B$, where ρ_U is the fully mixed state on $\mathcal{H}_A$. □

The following lemma provides an expression for the L_2-distance from uniform for the case where the first subsystem is classical.

Lemma 5.2.4. *Let $\rho_{XB} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_B)$ be classical with respect to an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of $\mathcal{H}_X$, let ρ_B^x, for $x \in \mathcal{X}$, be the corresponding (non-normalized) conditional operators, and let $\sigma \in \mathcal{P}(\mathcal{H}_B)$. Then*

$$d_2(\rho_{XB}|\sigma_B) = \sum_x \mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})^2\big) - \frac{1}{|\mathcal{X}|}\mathrm{tr}\big((\sigma_B^{-1/4}\rho_B\sigma_B^{-1/4})^2\big) .$$

Proof. Let ρ_U be the fully mixed state on $\mathcal{H}_X$. Because ρ_{XB} is classical on $\mathcal{H}_X$, we have

$$\rho_{XB} - \rho_U \otimes \rho_B = \sum_x |x\rangle\langle x| \otimes (\rho_B^x - \frac{1}{|\mathcal{X}|}\rho_B) ,$$

and thus

$$\begin{aligned}(\mathrm{id}_X \otimes \sigma_B^{-1/4})&(\rho_{XB} - \rho_U \otimes \rho_B)(\mathrm{id}_X \otimes \sigma_B^{-1/4}) \\ &= \sum_x |x\rangle\langle x| \otimes \Big(\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4} - \frac{1}{|\mathcal{X}|}\sigma_B^{-1/4}\rho_B\sigma_B^{-1/4}\Big) .\end{aligned}$$

Hence, since $\{|x\rangle\}_{x\in\mathcal{X}}$ is an orthonormal basis,

$$\begin{aligned}\mathrm{tr}\Bigl(\bigl((\mathrm{id}_X\otimes\sigma_B^{-1/4})(\rho_{XB}-\rho_U\otimes\rho_B)(\mathrm{id}_X\otimes\sigma_B^{-1/4})\bigr)^2\Bigr)\\ &=\sum_x \mathrm{tr}\Bigl(\bigl(\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4}-\frac{1}{|\mathcal{X}|}\sigma_B^{-1/4}\rho_B\sigma_B^{-1/4}\bigr)^2\Bigr)\\ &=\sum_x \mathrm{tr}\bigl((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})^2\bigr)-\frac{1}{|\mathcal{X}|}\mathrm{tr}\bigl((\sigma_B^{-1/4}\rho_B\sigma_B^{-1/4})^2\bigr)\ ,\end{aligned}$$

where the second equality holds because $\sum_x \rho_B^x = \rho_B$. The assertion then follows from (5.3). □

5.3 Collision entropy

Definition 5.3.1 below can be seen as a generalization of the well-known classical (conditional) collision entropy to quantum states.

Definition 5.3.1. Let $\rho_{AB}\in\mathcal{P}(\mathcal{H}_A\otimes\mathcal{H}_B)$ and $\sigma_B\in\mathcal{P}(\mathcal{H}_B)$. Then the *collision entropy of ρ_{AB} relative to σ_B* is

$$H_2(\rho_{AB}|\sigma_B):=-\log\frac{1}{\mathrm{tr}(\rho_{AB})}\mathrm{tr}\Bigl(\bigl(\rho_{AB}(\mathrm{id}_A\otimes\sigma_B^{-1/2})\bigr)^2\Bigr)\ .$$

Remark 5.3.2. It follows immediately from Lemma B.5.3 that

$$H_{\min}(\rho_{AB}|\sigma_B)\leq H_2(\rho_{AB}|\sigma_B)\ .$$

Remark 5.3.3. If $\rho_{XB}\in\mathcal{P}(\mathcal{H}_X\otimes\mathcal{H}_B)$ is classical with respect to an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of $\mathcal{H}_X$ such that the (non-normalized) conditional operators ρ_B^x on $\mathcal{H}_B$, for $x\in\mathcal{X}$, are orthogonal then

$$2^{-H_2(\rho_{XB}|\sigma_B)}=\frac{1}{\mathrm{tr}(\rho_{XB})}\sum_x\mathrm{tr}\bigl((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})^2\bigr)\ .$$

5.4 Two-universal hashing

Definition 5.4.1. Let $\mathcal{F}$ be a family of functions from $\mathcal{X}$ to $\mathcal{Z}$ and let P_F be a probability distribution on $\mathcal{F}$. The pair $(\mathcal{F},P_F)$ is called *two-universal* if $\Pr_f[f(x)=f(x')]\leq\frac{1}{|\mathcal{Z}|}$, for any distinct $x,x'\in\mathcal{X}$ and f chosen at random from $\mathcal{F}$ according to the distribution P_F.

In accordance with the standard literature on two-universal hashing, we will, for simplicity, assume that P_F is the uniform distribution on $\mathcal{F}$. In particular, the family $\mathcal{F}$ is said to be *two-universal* if $(\mathcal{F},P_F)$, for P_F uniform, is two-universal. It is, however, easy to see that all statements proven below also hold with respect to the general definition where P_F is arbitrary.

We will use the following lemma on the existence of two-universal function families.

Lemma 5.4.2. *Let $0 \leq \ell \leq n$. Then there exists a two-universal family of hash functions from $\{0,1\}^n$ to $\{0,1\}^\ell$.*

Proof. For the proof of this statement we refer to [CW79] or [WC81], where explicit constructions of hash function families are given. □

Consider an operator ρ_{XB} which is classical with respect to an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of $\mathcal{H}_X$ and assume that f is a function from $\mathcal{X}$ to $\mathcal{Z}$. The density operator describing the classical function output together with the quantum system $\mathcal{H}_B$ is then given by

$$\rho_{f(X)B} := \sum_{z\in\mathcal{Z}} |z\rangle\langle z| \otimes \rho_B^z \qquad \text{for } \rho_B^z := \sum_{x\in f^{-1}(z)} \rho_B^x \ , \tag{5.4}$$

where $\{|z\rangle\}_{z\in\mathcal{Z}}$ is an orthonormal basis of $\mathcal{H}_Z$.

Assume now that the function f is randomly chosen from a family of functions $\mathcal{F}$ according to a probability distribution P_F. The function output $f(x)$, the state of the quantum system, and the choice of the function f is then described by the operator

$$\rho_{F(X)BF} := \sum_{f\in\mathcal{F}} P_F(f)\rho_{f(X)B} \otimes |f\rangle\langle f| \tag{5.5}$$

on $\mathcal{H}_Z \otimes \mathcal{H}_B \otimes \mathcal{H}_F$, where $\mathcal{H}_F$ is a Hilbert space with orthonormal basis $\{|f\rangle\}_{f\in\mathcal{F}}$.

The following lemma provides an upper bound on the expected L_2-distance from uniform of a key computed by two-universal hashing.

Lemma 5.4.3. *Let $\rho_{XB} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_B)$ be classical on $\mathcal{H}_X$, let $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$, and let $\mathcal{F}$ be a two-universal family of hash functions from $\mathcal{X}$ to $\mathcal{Z}$. Then*

$$\mathbb{E}_f\Big[d_2(\rho_{f(X)B}|\sigma_B)\Big] \leq \mathrm{tr}(\rho_{XB})2^{-H_2(\rho_{XB}|\sigma_B)},$$

for $\rho_{f(X)B} \in \mathcal{P}(\mathcal{H}_Z \otimes \mathcal{H}_B)$ defined by (5.4) and f chosen uniformly from $\mathcal{F}$.

Proof. Since $\rho_{f(X)B}$ is classical on $\mathcal{H}_Z$, we have, according to Lemma 5.2.4,

$$d_2(\rho_{f(X)B}|\sigma_B) = \sum_z \mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^z\sigma_B^{-1/4})^2\big) - \frac{1}{|\mathcal{Z}|}\mathrm{tr}\big((\sigma_B^{-1/4}\rho_B\sigma_B^{-1/4})^2\big) \ , \tag{5.6}$$

where ρ_B^z, for $z \in \mathcal{Z}$, are the conditional operators defined by (5.4). The first term on the right hand side of (5.6) can be rewritten as

$$\sum_z \mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^z\sigma_B^{-1/4})^2\big)$$
$$= \sum_z \sum_{\substack{x \in f^{-1}(z) \\ x' \in f^{-1}(z)}} \mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})(\sigma_B^{-1/4}\rho_B^{x'}\sigma_B^{-1/4})\big)$$
$$= \sum_{x,x'} \delta_{f(x),f(x')} \mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})(\sigma_B^{-1/4}\rho_B^{x'}\sigma_B^{-1/4})\big) \ .$$

Similarly, for the second term of (5.6) we find

$$\frac{1}{|\mathcal{Z}|}\mathrm{tr}\big((\sigma_B^{-1/4}\rho_B\sigma_B^{-1/4})^2\big) = \sum_{x,x'} \frac{1}{|\mathcal{Z}|}\mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})(\sigma_B^{-1/4}\rho_B^{x'}\sigma_B^{-1/4})\big) \ .$$

Hence,

$$\mathop{\mathbb{E}}_f\Big[d_2(\rho_{f(X)B}|\sigma_B)\Big]$$
$$= \sum_{x,x'} \mathop{\mathbb{E}}_f\big[\delta_{f(x),f(x')} - \frac{1}{|\mathcal{Z}|}\big] \cdot \mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})(\sigma_B^{-1/4}\rho_B^{x'}\sigma_B^{-1/4})\big) \ . \quad (5.7)$$

Because f is chosen at random from a two-universal family of hash functions from $\mathcal{X}$ to $\mathcal{Z}$, we have, for any $x \neq x'$,

$$\mathop{\mathbb{E}}_f\big[\delta_{f(x),f(x')} - \frac{1}{|\mathcal{Z}|}\big] = \Pr_f[f(x) = f(x')] - \frac{1}{|\mathcal{Z}|} \leq 0 \ ,$$

Since the trace $\mathrm{tr}(\sigma\sigma')$ of two nonnegative operators $\sigma, \sigma' \in \mathcal{P}(\mathcal{H})$ cannot be negative (cf. Lemma B.5.2) the trace on the right hand side of (5.7) cannot be negative, for any $x, x' \in \mathcal{X}$. Consequently, when omitting all terms with $x \neq x'$, the sum can only get larger, that is,

$$\mathop{\mathbb{E}}_f\Big[d_2(\rho_{f(X)B}|\sigma_B)\Big] \leq \sum_x \mathrm{tr}\big((\sigma_B^{-1/4}\rho_B^x\sigma_B^{-1/4})^2\big) \ .$$

The assertion then follows from Remark 5.3.3. □

5.5 Security of privacy amplification

We are now ready to state our main result on privacy amplification in the context of quantum adversaries. Let X be a string and assume that an adversary controls a quantum system $\mathcal{H}_B$ whose state is correlated with X. Theorem 5.5.1 provides a bound on the security of a key $f(X)$ computed from X by two-universal hashing. The bound only depends on the uncertainty of the adversary on X, measured in terms of collision entropy, min-entropy (cf. Corollary 5.5.2), or smooth min-entropy (Corollary 5.6.1), where the latter is (nearly) optimal (see Section 5.6).

Theorem 5.5.1. *Let $\rho_{XB} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_B)$ be classical with respect to an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of $\mathcal{H}_X$, let $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$, and let $\mathcal{F}$ be a two-universal family of hash function from $\mathcal{X}$ to $\{0,1\}^\ell$. Then*

$$d(\rho_{F(X)BF}|BF) \leq \sqrt{\operatorname{tr}(\rho_{XB}) \cdot \operatorname{tr}(\sigma_B)} \cdot 2^{-\frac{1}{2}(H_2(\rho_{XB}|\sigma_B)-\ell)} ,$$

for $\rho_{F(X)BF} \in \mathcal{P}(\mathcal{H}_Z \otimes \mathcal{H}_B \otimes \mathcal{H}_F)$ defined by (5.5).

Proof. We use Lemma 5.2.2 to write the L_1-distance from uniform as an expectation value,

$$d(\rho_{F(X)BF}|BF) = \sum_{f\in\mathcal{F}} P_F(f) \cdot d(\rho_{f(X)B}|B) = \mathbb{E}_f\big[d(\rho_{f(X)B}|B)\big] .$$

With Lemma 5.2.3, the term in the expectation can be bounded in terms of the L_2-distance from uniform, that is, for any $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$,

$$\begin{aligned} d(\rho_{F(X)BF}|BF) &\leq \sqrt{2^\ell \operatorname{tr}(\sigma_B)}\, \mathbb{E}_f\Big[\sqrt{d_2(\rho_{f(X)B}|\sigma_B)}\Big] \\ &\leq \sqrt{2^\ell \operatorname{tr}(\sigma_B)} \sqrt{\mathbb{E}_f\big[d_2(\rho_{f(X)B}|\sigma_B)\big]} , \end{aligned}$$

where we have used Jensen's inequality. Finally, we apply Lemma 5.4.3 to bound the L_2-distance from uniform in terms of the collision entropy, which gives

$$d(\rho_{F(X)BF}|BF) \leq \sqrt{2^\ell \operatorname{tr}(\sigma_B)} \sqrt{\operatorname{tr}(\rho_{XB}) 2^{-H_2(\rho_{XB}|\sigma_B)}} .$$ □

Corollary 5.5.2. *Let $\rho_{XB} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_B)$ be classical with respect to an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of $\mathcal{H}_X$ and let $\mathcal{F}$ be a two-universal family of hash functions from $\mathcal{X}$ to $\{0,1\}^\ell$. Then*

$$d(\rho_{F(X)BF}|BF) \leq \sqrt{\operatorname{tr}(\rho_{XB})} \cdot 2^{-\frac{1}{2}(H_{\min}(\rho_{XB}|B)-\ell)} ,$$

for $\rho_{F(X)BF} \in \mathcal{P}(\mathcal{H}_Z \otimes \mathcal{H}_B \otimes \mathcal{H}_F)$ defined by (5.5).

Proof. The assertion follows directly from Theorem 5.5.1 and Remark 5.3.2. □

5.6 Characterization using smooth min-entropy

The characterization of privacy amplification in terms of the collision entropy or min-entropy is not optimal.[2] Because of Remark 5.3.2, the same problem arises if we replace the collision entropy by the min-entropy (as in Corollary 5.5.2). However, as we shall see, the statement of Theorem 5.5.1 still holds if the uncertainty is measured in terms of *smooth* min-entropy. That is, the key generated from X by two-universal hashing is secure if its length is slightly smaller than roughly $H_{\min}^{\varepsilon}(\rho_{XB}|B)$, where ρ_{XB} is the joint state of the initial string X and the adversary's knowledge. This is essentially optimal, i.e., $H_{\min}^{\varepsilon}(\rho_{XB}|B)$ is also an upper bound on the maximum number of key bits that can be generated from X.[3]

Corollary 5.6.1. *Let $\rho_{XB} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_B)$ be a density operator which is classical with respect to an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of $\mathcal{H}_X$, let $\mathcal{F}$ be a two-universal family of hash functions from $\mathcal{X}$ to $\{0,1\}^\ell$, and let $\varepsilon \geq 0$. Then*

$$d(\rho_{F(X)BF}|BF) \leq 2\varepsilon + 2^{-\frac{1}{2}(H_{\min}^{\varepsilon}(\rho_{XB}|B)-\ell)} ,$$

for $\rho_{F(X)BF} \in \mathcal{P}(\mathcal{H}_Z \otimes \mathcal{H}_B \otimes \mathcal{H}_F)$ defined by (5.5)*.*

Proof. Consider an arbitrary operator $\bar{\rho}_{XB} \in \mathcal{B}^{\varepsilon}(\rho_{XB})$ and let $\bar{\rho}_{F(X)BF} \in \mathcal{P}(\mathcal{H}_Z \otimes \mathcal{H}_B \otimes \mathcal{H}_F)$ be the corresponding operator defined by (5.5). Because the L_1-distance cannot increase when applying a trace-preserving quantum operation (cf. Lemma A.2.1), we have $\bar{\rho}_{F(X)BF} \in \mathcal{B}^{\varepsilon}(\rho_{F(X)BF})$. Hence, by the triangle inequality,

$$\begin{aligned}
d(\rho_{F(Z)BF}|BF) &= \left\|\rho_{F(X)BF} - \rho_U \otimes \rho_{BF}\right\|_1 \\
&\leq \left\|\rho_{F(X)BF} - \bar{\rho}_{F(X)BF}\right\|_1 + \left\|\bar{\rho}_{F(X)BF} - \rho_U \otimes \bar{\rho}_{BF}\right\|_1 + \left\|\bar{\rho}_{BF} - \rho_{BF}\right\|_1 \\
&\leq 2\varepsilon + \left\|\bar{\rho}_{F(X)BF} - \rho_U \otimes \bar{\rho}_{BF}\right\|_1 = 2\varepsilon + d(\bar{\rho}_{F(Z)BF}|BF) ,
\end{aligned}$$

where ρ_U is the fully mixed state on $\mathcal{H}_Z$. Corollary 5.5.2, applied to $\bar{\rho}_{XB}$, gives

$$d(\rho_{F(X)BF}|BF) \leq 2\varepsilon + \sqrt{\mathrm{tr}(\bar{\rho}_{XB})} \cdot 2^{-\frac{1}{2}(H_{\min}(\bar{\rho}_{XB}|B)-\ell)} .$$

Because this holds for any $\bar{\rho}_{XB} \in \mathcal{B}^{\varepsilon}(\bar{\rho}_{XB})$, the assertion follows by the definition of smooth min-entropy. □

[2] This also holds for the classical result, as observed in [BBCM95]. In fact, depending on the probability distribution P_X of the initial string X, it might be possible to extract a key whose length exceeds the collision entropy of P_X.

[3] To see this, let F be an arbitrary hash function. It follows from Lemma 3.1.9 that the smooth min-entropy cannot increase when applying a function on X, i.e., $H_{\min}^{\varepsilon}(\rho_{XB}|B) \geq H_{\min}^{\varepsilon}(\rho_{F(X)BF}|BF)$. Moreover, it is easy to verify that the smooth min-entropy of a secret key given the adversary's information is roughly equal to its length. Hence, if $F(X)$ is a secret key of length ℓ, we have $H_{\min}^{\varepsilon}(\rho_{F(X)BF}|BF) \geq \ell$. Combining this with the above gives $H_{\min}^{\varepsilon}(\rho_{XB}|B) \geq \ell$.

Chapter 6

Security of QKD

In this chapter, we use the techniques developed in Chapters 3–5 to prove the security of QKD.[1] (The reader is referred to Section 1.6 for a high-level description of the material presented in the following, including a sketch of the security proof.) Typically, a QKD protocol is built from several subprotocols, e.g., for parameter estimation, information reconciliation, or privacy amplification. We first describe and analyze these subprotocols (Sections 6.2–6.4) and then put the parts together to get a general security criterion for quantum key *distillation* (Section 6.5), which directly implies the security of quantum key *distribution* (QKD) (Section 6.6).

6.1 Preliminaries

6.1.1 Two-party protocols

A *protocol* $\mathcal{P}$ between two parties, Alice and Bob, is specified by a sequence of operations, called *(protocol) steps*, to be performed by each of the parties. In the first protocol step, Alice and Bob might take (classical or quantum) *inputs* A and B, respectively (e.g., some correlated data). In each of the following steps, Alice and Bob either perform local computations or exchange messages (using a classical or a quantum communication channel). Finally, in the last protocol step, Alice and Bob generate *outputs* A' and B', respectively (e.g., a pair of secret keys).

We will mostly (except for Section 6.6) be concerned with the analysis of protocols $\mathcal{P}$ that only use communication over a classical and authentic channel. In this case, Alice and Bob's outputs as well as the transcript of the communication do not depend on the attack of a potential adversary. Let ρ_{AB} and $\rho_{A'B'C}$ be the density operators describing Alice and Bob's inputs A and B as well as their outputs A' and B' together with the communication

[1] As discussed in Chapter 1, we actually consider quantum key *distillation*, which is somewhat more general than quantum key *distribution* (QKD).

transcript C, respectively. The mapping that brings ρ_{AB} to $\rho_{A'B'C}$, in the following denoted by $\mathcal{E}^{\mathcal{P}}_{A'B'C\leftarrow AB}$, is then uniquely defined by the protocol $\mathcal{P}$. Moreover, because it must be physically realizable, $\mathcal{E}^{\mathcal{P}}_{A'B'C\leftarrow AB}$ is a CPM (see Section 2.1.1).

To analyze the security of a protocol $\mathcal{P}$, we need to include Eve's information in our description. Let ρ_{ABE} be the state of Alice and Bob's inputs as well as Eve's initial information. Similarly, let $\rho_{A'B'E'}$ be the state of Alice and Bob's outputs together with Eve's information after the protocol execution. As Eve might get a transcript C of the messages sent over the classical channel, the CPM that maps ρ_{ABE} to $\rho_{A'B'E'}$ is given by

$$\mathcal{E}^{\mathcal{P}}_{A'B'E'\leftarrow ABE} := \mathcal{E}^{\mathcal{P}}_{A'B'C\leftarrow AB} \otimes \mathrm{id}_E ,$$

where $\mathcal{H}_{E'} := \mathcal{H}_C \otimes \mathcal{H}_E$.

6.1.2 Robustness of protocols

Depending on its input, a protocol might be unable to produce the desired output. For example, if a key distillation protocol starts with uncorrelated randomness, it cannot generate a pair of secret keys. In this case, the best we can hope for is that the protocol recognizes this situation and aborts[2] (instead of generating an insecure result).

Clearly, one is interested in designing protocols that are successful on certain inputs. This requirement is captured by the notion of *robustness*.

Definition 6.1.1. Let $\mathcal{P}$ be a two-party protocol and let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$. We say that $\mathcal{P}$ is *ε-robust on ρ_{AB}* if, for inputs defined by ρ_{AB}, the probability that the protocol aborts is at most ε.

Mathematically, we represent the state that describes the situation after an abortion of the protocol as a zero operator. The CPM $\mathcal{E}^{\mathcal{P}}_{A'B'C\leftarrow AB}$ (as defined in Section 6.1.1) is then a projection onto the space that represents the outputs of successful protocol executions (i.e., where it did not abort). The probability that the protocol is successful when starting with an initial state ρ_{AB} is thus equal to the trace $\mathrm{tr}(\rho_{A'B'E})$ of the operator $\rho_{A'B'E} = \mathcal{E}^{\mathcal{P}}_{A'B'C\leftarrow AB}(\rho_{AB})$. In particular, if $\mathcal{P}$ is ε-robust on a density operator ρ_{AB} then $\mathrm{tr}(\rho_{A'B'E}) \geq 1-\varepsilon$.

6.1.3 Security definition for key distillation

A *(quantum) key distillation protocol* KD is a two-party protocol with classical communication where Alice and Bob take inputs from $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, and either output classical keys $s_A, s_B \in \mathcal{S}$, where $\mathcal{S}$ is called the *key space* of KD, or abort the protocol.

[2]Technically, the protocol might output a certain predefined symbol which indicates that it is unable to accomplish the task.

Definition 6.1.2. Let KD be a key distillation protocol and let $\rho_{ABE} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$. We say that KD is *ε-secure on ρ_{ABE}* if $\rho_{S_A S_B E'} := \mathcal{E}^{\mathsf{KD}}_{S_A S_B E' \leftarrow ABE}(\rho_{ABE})$ satisfies

$$\frac{1}{2}\left\|\rho_{S_A S_B E'} - \rho_{UU} \otimes \rho_{E'}\right\|_1 \leq \varepsilon ,$$

where $\rho_{UU} := \sum_{s \in \mathcal{S}} \frac{1}{|\mathcal{S}|} |s\rangle\langle s| \otimes |s\rangle\langle s|$, for some family $\{|s\rangle\}_{s\in\mathcal{S}}$ of orthonormal vectors representing the values of the key space $\mathcal{S}$.

Moreover, we say that KD is *ε-fully secure* if it is ε-secure on all density operators $\rho_{ABE} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$.

According to the discussion on universal security in Section 2.2.2,[3] this definition has a very intuitive interpretation: If the protocol is ε-fully secure then, for any arbitrary input, the probability of the event that Alice and Bob do not abort *and* the adversary gets information on the key pair[4] is at most ε.[5] In other words, except with probability ε, Alice and Bob either abort or generate a pair of keys which are identical to a perfect key.

Remark 6.1.3. The above security definition for key distillation protocols KD can be subdivided into two parts:

- *ε'-correctness:* $\Pr[s_A \neq s_B] \leq \varepsilon'$,[6] for s_A and s_B chosen according to the distribution defined by $\rho_{S_A S_B}$.
- *ε''-secrecy of Alice's key:* $\frac{1}{2} d(\rho_{S_A CE}|CE) \leq \varepsilon''$.[7]

In particular, if KD is ε'-correct and ε''-secret on ρ_{XYE} then it is $(\varepsilon' + \varepsilon'')$-secure on ρ_{XYE}.

6.2 Parameter estimation

The purpose of a parameter estimation is to decide whether the input given to the protocol can be used for a certain task, e.g. to distill a secret key. Technically, a *parameter estimation protocol* PE is simply a two-party protocol where Alice and Bob take inputs from $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, and either output "accept" or abort the protocol.

[3] If a key S is ε-secure, one could define a perfectly secure (independent and uniformly distributed) key U such that $\Pr[s \neq u] \leq \varepsilon$ (see also Proposition 2.1.1).

[4] According to Footnote 3, one could say that the adversary *gets information on a key* S whenever the value of S is not equal to the value of a perfect key U.

[5] Note that the adversary's information on the key, *conditioned* on the event that Alice and Bob generate a key, is not necessarily small. In fact, if, for a certain input, the probability that Alice and Bob generate a key is very small (e.g., smaller than ε) then—conditioned on this rare event—the key might be insecure (see also the discussion in [BBB+05]).

[6] $\Pr[s_A \neq s_B]$ is the probability of the event that Alice and Bob do not abort *and* the generated keys s_A and s_B are different.

[7] See Definition 5.2.1.

Fig. 6.1 Parameter estimation protocol $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$.

Parameters:

$\mathcal{M}$: bipartite POVM $\{M_w\}_{w\in\mathcal{W}}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$
$\mathcal{Q}$: set of frequency distributions on $\mathcal{W}$

Alice		**Bob**
input space: $\mathcal{H}_A^{\otimes n}$		input space: $\mathcal{H}_B^{\otimes n}$
$\mathcal{H}_A^{\otimes n}$	meas. $\mathcal{M}^{\otimes n}$ $\longleftrightarrow$	$\mathcal{H}_A^{\otimes n}$ $\rightarrow \mathbf{w} = (w_1, \ldots, w_n)$ if $\lambda_{\mathbf{w}} \notin \mathcal{Q}$ then abort else output "acc."

Definition 6.2.1. Let PE be a parameter estimation protocol and let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$. We say that PE *ε-securely filters* ρ_{AB} if, on input ρ_{AB}, the protocol aborts except with probability ε.

A typical and generic example for parameter estimation is the protocol $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ depicted in Fig. 6.1. Alice and Bob take inputs from an n-fold product space. Then they measure each of the n subspaces according to a POVM $\mathcal{M} = \{M_w\}_{w\in\mathcal{W}}$.[8] Finally, they output "accept" if the frequency distribution $\lambda_{\mathbf{w}}$ of the measurement outcomes $\mathbf{w} = (w_1, \ldots, w_n)$ is contained in a certain set $\mathcal{Q}$.

For the analysis of this protocol, it is convenient to consider the set $\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}$ of density operators σ_{AB} for which the measurement $\mathcal{M}$ leads to a distribution which has distance *at most* μ to the set $\mathcal{Q}$. Formally,

$$\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}} := \bigl\{\sigma_{AB} : \min_{Q\in\mathcal{Q}} \|P_W^{\sigma_{AB}} - Q\|_1 \leq \mu\bigr\} , \tag{6.1}$$

where $P_W^{\sigma_{AB}}$ denotes the probability distribution of the outcomes when measuring σ_{AB} according to $\mathcal{M}$, i.e., $P_W(w) = \mathrm{tr}(M_w \sigma_{AB})$, for any $w \in \mathcal{W}$.

Assume that the protocol $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ takes as input a product state $\rho_{A^nB^n} = \sigma_{AB}^{\otimes n}$. Then, by the law of large numbers, the measurement statistics $\lambda_{\mathbf{w}}$ must be close to $\mathcal{M}(\sigma_{AB})$. In particular, if the protocol accepts with non-negligible probability (i.e., $\lambda_{\mathbf{w}}$ is contained in $\mathcal{Q}$) then σ_{AB} is likely to be contained in $\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}$, for some small $\mu > 0$. In other words, the protocol

[8] $\mathcal{M}$ might be an arbitrary measurement that can be performed by two distant parties connected by a classical channel.

aborts with high probability if σ_{AB} is not an element of the set $\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}$. The following lemma generalizes this statement to permutation-invariant inputs.

Lemma 6.2.2. *Let $\mathcal{M} := \{M_w\}_{w\in\mathcal{W}}$ be a POVM on $\mathcal{H}_A\otimes\mathcal{H}_B$, let $\mathcal{Q}$ be a set of frequency distributions on $\mathcal{W}$, let $0 \leq r \leq \frac{1}{2}n$, and let $\varepsilon \geq 0$. Moreover, let $|\theta\rangle \in \mathcal{H}_{ABE} := \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ and let $\rho_{A^nB^nE^n}$ be a density operator on $\mathrm{Sym}(\mathcal{H}_{ABE}^{\otimes n}, |\theta\rangle^{\otimes n-r})$. If $\mathrm{tr}_E(|\theta\rangle\langle\theta|)$ is not contained in the set $\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}$ defined by* (6.1)*, for*

$$\mu := 2\sqrt{\frac{\log(1/\varepsilon)}{n} + h(r/n) + \frac{|\mathcal{W}|}{n}\log(\frac{n}{2}+1)}\ ,$$

then the protocol $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ defined by Fig. 6.1 ε-securely filters $\rho_{A^nB^n}$.

Proof. The assertion follows directly from Theorem 4.5.2. □

Similarly to (6.1), we can define a set $\overline{\Gamma}^{\geq\mu}_{\mathcal{M},\mathcal{Q}}$ containing all density operators σ_{AB} for which the measurement $\mathcal{M}$ leads to a distribution which has distance *at least* μ to the *complement* of $\mathcal{Q}$. Formally,

$$\overline{\Gamma}^{\geq\mu}_{\mathcal{M},\mathcal{Q}} := \bigl\{\sigma_{AB} : \min_{Q\notin\mathcal{Q}} \|P_W^{\sigma_{AB}} - Q\|_1 \geq \mu\bigr\}\ . \tag{6.2}$$

Analogously to the above argument, one can show that the protocol $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ defined by Fig. 6.1 is ε-robust on product operators $\sigma_{AB}^{\otimes n}$, for any $\sigma_{AB} \in \overline{\Gamma}^{\geq\mu}_{\mathcal{M},\mathcal{Q}}$.

6.3 Information reconciliation

Assume that Alice and Bob hold weakly correlated classical values x and y, respectively. The purpose of an information reconciliation protocol is to transform x and y into a pair of fully correlated strings, while leaking only a minimum amount of information (on the final strings) to an eavesdropper (see, e.g., [BS94]).

6.3.1 Definition

We focus on information reconciliation schemes where Alice keeps her input value x and where Bob outputs a guess $\hat{x}$ for x. Hence, technically, an *information reconciliation protocol* IR is a two-party protocol where Alice and Bob take classical inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively, and where Bob outputs a classical value $\hat{x} \in \mathcal{X}$ or aborts.

Definition 6.3.1. Let $P_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ and let $\varepsilon \geq 0$. We say that an information reconciliation protocol IR is *ε-secure on P_{XY}* if, for inputs x

and y chosen according to P_{XY}, the probability that Bob's output $\hat{x}$ differs from Alice's input x is at most ε, i.e., $\Pr[\hat{x} \neq x] \leq \varepsilon$.[9]

Moreover, we say that IR is *ε-fully secure* if it is ε-secure on all probability distributions $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$.

The communication transcript of an information reconciliation scheme IR generally contains useful information on Alice and Bob's values. If the communication channel is insecure, this information might be leaked to Eve. Clearly, in the context of key agreement, one is interested in information reconciliation schemes for which this leakage is minimal.

Definition 6.3.2. Let IR be an information reconciliation protocol where Alice and Bob take inputs from $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\mathcal{C}$ be the set of all possible communication transcripts c and let $P_{C|X=x,Y=y}$ be the distribution of the transcripts $c \in \mathcal{C}$ conditioned on inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Then the *leakage of* IR is

$$\mathrm{leak}_{\mathsf{IR}} := \log |\mathcal{C}| - \inf_{x,y} H_{\min}(P_{C|X=x,Y=y}) \ ,$$

where the infimum ranges over all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Note that the leakage is independent of the actual distribution P_{XY} of Alice and Bob's values.

6.3.2 Information reconciliation with minimum leakage

A typical information reconciliation protocol is the protocol $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$ defined by Fig. 6.2. It is a so-called *one-way* protocol where only Alice sends messages to Bob. We show that the leakage of this protocol, for appropriately chosen parameters, is roughly bounded by the max-entropy of X given Y (Lemma 6.3.3). This statement can be extended to smooth max-entropy (Lemma 6.3.4), which turns out to be optimal, i.e., the minimum leakage of an information reconciliation protocol for P_{XY} is exactly characterized by $H^{\varepsilon}_{\max}(X|Y)$. In particular, for the special case where the input is chosen according to a product distribution, we get an asymptotic expression in terms of Shannon entropy (Corollary 6.3.5), which corresponds to the Shannon coding theorem.

Lemma 6.3.3. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and let $\varepsilon > 0$. Then the information reconciliation protocol $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$ defined by Fig. 6.2, for an appropriate choice of the parameters $\hat{\mathcal{X}}$ and $\mathcal{F}$, is 0-robust on P_{XY}, ε-fully secure, and has leakage*

$$\mathrm{leak}_{\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}} \leq H_{\max}(P_{XY}|Y) + \log(2/\varepsilon) \ .$$

[9] We denote by $\Pr[\hat{x} \neq x]$ the probability of the event that the protocol does not abort *and* $\hat{x}$ is different from x.

Fig. 6.2 Information reconciliation protocol $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$.

Parameters:
$\hat{\mathcal{X}}$: family of sets $\hat{\mathcal{X}}_y \subseteq \mathcal{X}$ parameterized by $y \in \mathcal{Y}$.
$\mathcal{F}$: family of hash functions from $\mathcal{X}$ to $\mathcal{Z}$.

Alice		**Bob**
input: $x \in \mathcal{X}$		input: $y \in \mathcal{Y}$
$f \in_R \mathcal{F}$ $z := f(x)$	$\xrightarrow{f,z}$	$\hat{\mathcal{D}} :=$ $\{\hat{x} \in \hat{\mathcal{X}}_y : f(\hat{x}) = z\}$
		if $\hat{\mathcal{D}} \neq \emptyset$ then $\hat{x} \in_R \hat{\mathcal{D}}$ else abort
		output $\hat{x}$

Proof. Let $k := \lceil H_{\max}(P_{XY}|Y) + \log(1/\varepsilon) \rceil$ and let $\mathcal{F}$ be a two-universal family of hash functions from $\mathcal{X}$ to $\mathcal{Z} := \{0,1\}^k$ (which exists according to Lemma 5.4.2). Furthermore, let $\hat{\mathcal{X}} = \{\hat{\mathcal{X}}_y\}_{y \in \mathcal{Y}}$ be the family of sets defined by $\hat{\mathcal{X}}_y := \mathrm{supp}(P_X^y)$, where $\mathrm{supp}(P_X^y)$ denotes the support of the function $P_X^y : x \mapsto P_{XY}(x,y)$.

For any pair of inputs x and y and for any communication $(f,z) = (f, f(x))$ computed by Alice, Bob can only output a wrong value if the set $\hat{\mathcal{X}}_y = \mathrm{supp}(P_X^y)$ contains an element $\hat{x} \neq x$ such that $f(\hat{x}) = z$. Because f is chosen uniformly at random from the family of two-universal hash functions $\mathcal{F}$, we have $\Pr_f\bigl[f(\hat{x}) = f(x)\bigr] \leq \frac{1}{|\mathcal{Z}|} = 2^{-k}$, for any $\hat{x} \neq x$. Hence, by the union bound, for any fixed $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} \Pr[\hat{x} \neq x] &\leq \Pr_f\bigl[\exists \hat{x} \in \mathrm{supp}(P_X^y) : \hat{x} \neq x \wedge f(\hat{x}) = f(x)\bigr] \\ &\leq \bigl|\mathrm{supp}(P_X^y)\bigr| \cdot 2^{-k} . \end{aligned}$$

Because, by Remark 3.1.4, $\max_{y'} |\mathrm{supp}(P_X^{y'})| = 2^{H_{\max}(P_{XY}|Y)}$, we conclude

$$\Pr[\hat{x} \neq x] \leq 2^{H_{\max}(P_{XY}|Y) - \lceil H_{\max}(P_{XY}|Y) + \log(1/\varepsilon) \rceil} \leq \varepsilon ,$$

that is, $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$ is ε-secure on any probability distribution.

Moreover, if (x,y) is chosen according to the distribution P_{XY}, then, clearly, x is always contained in $\hat{\mathcal{X}}_y = \mathrm{supp}(P_X^y)$, that is, Bob never aborts. This proves that the protocol is 0-robust.

Since f is chosen uniformly at random and independently of x from the family of hash-functions $\mathcal{F}$, all nonzero probabilities of the distribution $P_{C|X=x}$ are equal to $\frac{1}{|\mathcal{F}|}$. Hence, using the fact that $\mathcal{C} = \mathcal{F} \times \mathcal{Z}$,

$$\begin{aligned} \mathrm{leak}_{\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}} &= \log|\mathcal{C}| - \inf_{x \in \mathcal{X}} H_{\min}(P_{C|X=x}) \\ &= \log|\mathcal{F} \times \mathcal{Z}| - \log|\mathcal{F}| \leq \log|\mathcal{Z}| = k \ . \end{aligned}$$

The claimed bound on the leakage then follows by the definition of k. □

Lemma 6.3.4. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and let $\varepsilon, \varepsilon' \geq 0$. Then the information reconciliation protocol $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$ defined by Fig. 6.2, for an appropriate choice of the parameters $\hat{\mathcal{X}}$ and $\mathcal{F}$, is ε'-robust on P_{XY}, ε-fully secure, and has leakage*

$$\mathrm{leak}_{\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}} \leq H_{\max}^{\varepsilon'}(P_{XY}|Y) + \log(2/\varepsilon) \ .$$

Proof. For any $\nu > 0$ there exists $\bar{P}_{XY} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_Y)$ such that

$$\left\| P_{XY} - \bar{P}_{XY} \right\|_1 \leq \varepsilon' \tag{6.3}$$

and

$$H_{\max}(\bar{P}_{XY}|Y) \leq H_{\max}^{\varepsilon'}(P_{XY}|Y) + \nu \ . \tag{6.4}$$

According to Lemma 6.3.3, there exists $\hat{\mathcal{X}}$ and $\mathcal{F}$ such that $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$ is ε-fully secure, 0-robust on $\bar{P}_{XY}$, and has leakage

$$\mathrm{leak}_{\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}} \leq H_{\max}(\bar{P}_{XY}|Y) + \log(2/\varepsilon) \ .$$

The stated bound on the leakage follows immediately from this inequality and (6.4). Moreover, the bound on the robustness is a direct consequence of the bound (6.3) and the fact that the protocol is 0-robust on $\bar{P}_{XY}$. □

Corollary 6.3.5. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a probability distribution, let $n \geq 0$, and let $\varepsilon \geq 0$. Then there exists an information reconciliation protocol IR which is ε-fully secure, ε-robust on the product distribution $P_{X^nY^n} := (P_{XY})^n$, and has leakage*

$$\frac{1}{n}\mathrm{leak}_{\mathsf{IR}} \leq H(X|Y) + \sqrt{\frac{3\log(2/\varepsilon)}{n}} \log(|\mathcal{X}| + 3) \ .$$

Proof. Using Lemma 6.3.4 (with $\varepsilon = \varepsilon'$) and Theorem 3.3.4, we find

$$\frac{1}{n}\mathrm{leak}_{\mathsf{IR}} \leq H(X|Y) + \frac{\log(2/\varepsilon)}{n} + \sqrt{\frac{2\log(1/\varepsilon)}{n}} \log(|\mathcal{X}| + 3) \ .$$

Let $a := \frac{\log(2/\varepsilon)}{n}$ and $b := \log(|\mathcal{X}|+3)$. The last two terms on the right hand side of this inequality are then upper bounded by $a + \sqrt{2ab} \leq (\frac{a}{2} + \sqrt{2a})b$, which holds because $b \geq 2$. We can assume without loss of generality that $3a \leq 1$ (otherwise, the statement is trivial). Then $\frac{a}{2} + \sqrt{2a} \leq \sqrt{3a}$. The last two terms in the above inequality are thus bounded by $\sqrt{3a}b$, which concludes the proof. □

Fig. 6.3 Classical post-processing protocol $\mathsf{PP}_{\mathsf{IR},\mathcal{F}}$.

Parameters:

IR: information reconciliation protocol.

$\mathcal{F}$: family of hash functions from $\mathcal{X}$ to $\{0,1\}^\ell$.

Alice		**Bob**
input: $x \in \mathcal{X}$		input: $y \in \mathcal{Y}$
x	$\xleftrightarrow{\mathsf{IR}}$	$y \to \hat{x}$
$f \in_R \mathcal{F}$	$\xrightarrow{f}$	
output $s_A := f(x)$		output $s_B := f(\hat{x})$

For practical applications, we are interested in protocols where Alice and Bob's computations can be done *efficiently* (e.g., in time that only depends polynomially on the length of their inputs). This is, however, not necessarily the case for the information reconciliation protocol $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$ described above. While Alice's task, i.e., the evaluation of the hash function, can be done in polynomial time,[10] no efficient algorithm is known for the decoding operation of Bob. Nevertheless, based on a specific encoding scheme, one can show that there exist information reconciliation protocols which only require polynomial-time computations and for which the statement of Corollary 6.3.5 (asymptotically) still holds (see Appendix C).

6.4 Classical post-processing

Classical post-processing is used to transform an only partially secure[11] pair of raw keys x and y held by Alice and Bob, respectively, into a fully secure key pair. A classical post-processing protocol is thus actually a key distillation protocol that starts with classical randomness.

In this section, we analyze the security of the generic post-processing protocol depicted in Fig. 6.3. It consists of an information reconciliation subprotocol (see Section 6.3) followed by privacy amplification (see Chapter 5).

[10] Recall that Alice only has to evaluate a function which is randomly chosen from a two-universal family of functions. For most known constructions of such families (see, e.g., [CW79, WC81]), this can be done efficiently.

[11] That is, x and y are only weakly correlated and partially secret strings.

Lemma 6.4.1. *Let* IR *be an information reconciliation protocol and let* $\mathcal{F}$ *be a two-universal family of hash functions from* $\mathcal{X}$ *to* $\{0,1\}^\ell$*. Additionally, let* $\rho_{XYE} \in \mathcal{P}(\mathcal{H}_X \otimes \mathcal{H}_Y \otimes \mathcal{H}_E)$ *be a density operator which is classical on* $\mathcal{H}_X \otimes \mathcal{H}_Y$ *and let* $\varepsilon', \varepsilon'' \geq 0$*. If* IR *is* ε'*-secure on the distribution defined by* ρ_{XY} *and if*

$$\ell \leq H_{\min}^{\varepsilon}(\rho_{XE}|E) - \mathrm{leak}_{\mathsf{IR}} - 2\log(1/\varepsilon) \ ,$$

for $\varepsilon := \frac{2}{3}\varepsilon''$*, then the key distillation protocol* $\mathsf{PP}_{\mathsf{IR},\mathcal{F}}$ *defined by Fig. 6.3 is* $(\varepsilon' + \varepsilon'')$*-secure on* ρ_{XYE}*.*

Proof. For simplicity, we assume in the following that the protocol IR is one-way. It is straightforward to generalize this argument to arbitrary protocols.

Note first that the keys s_A and s_B generated by Alice and Bob can only differ if $\hat{x} \neq x$. Hence, because the information reconciliation protocol IR is ε'-secure on the distribution defined by ρ_{XY}, the classical post-processing protocol $\mathsf{PP}_{\mathsf{IR},\mathcal{F}}$ is ε'-correct on ρ_{XYE}. According to Remark 6.1.3, it thus remains to show that Alice's key is ε''-secret.

For this, we use the result on the security of privacy amplification by two-universal hashing presented in Chapter 5. Because f is chosen from a two-universal family of hash functions, Corollary 5.6.1 implies that the key computed by Alice is ε''-secret if

$$H_{\min}^{\varepsilon}(\rho_{XC'E}|C'E) \geq 2\log(1/\varepsilon) + \ell \ , \tag{6.5}$$

where $\rho_{X\hat{X}C'E} := (\mathcal{E}^{\mathsf{IR}} \otimes \mathrm{id}_E)(\rho_{XYE})$ is the operator describing the situation after the execution of the information reconciliation protocol IR (where C' is the transcript of IR). It thus suffices to verify that the bound on the entropy (6.5) holds.

Using the chain rule (cf. (3.21) of Theorem 3.2.12), the left hand side of (6.5) can be bounded by

$$H_{\min}^{\varepsilon}(\rho_{XC'E}|C'E) \geq H_{\min}^{\varepsilon}(\rho_{XC'E}|E) - H_{\max}(\rho_{C'}) \ .$$

Moreover, because the communication c' is computed only from x, the conditional operators $\rho_{C'E}^x$ have product form and thus (cf. (3.22) of Theorem 3.2.12)

$$H_{\min}^{\varepsilon}(\rho_{XC'E}|C'E) \geq H_{\min}^{\varepsilon}(\rho_{XE}|E) + H_{\min}(\rho_{C'X}|\rho_X) - H_{\max}(\rho_{C'}) \ . \tag{6.6}$$

Using the fact that $H_{\max}(\rho_{C'}) = \log \mathrm{rank}(\rho_{C'})$ and Lemma 3.1.8, the last two terms in the above expression can be bounded by

$$H_{\max}(\rho_{C'}) - H_{\min}(\rho_{C'X}|\rho_X) \leq \log \mathrm{rank}(\rho_{C'}) - \inf_{x \in \mathcal{X}} H_{\min}(\bar{\rho}_{C'}^x) \ ,$$

where, for any $x \in \mathcal{X}$, $\bar{\rho}_{C'}^x$ is the normalized conditional operator defined by $\rho_{C'X}$. Hence, by the definition of leakage,

$$H_{\max}(\rho_{C'}) - H_{\min}(\rho_{C'X}|\rho_X) \leq \mathrm{leak}_{\mathsf{IR}} \ .$$

Combining this with (6.6), we find

$$H^{\varepsilon}_{\min}(\rho_{XC'E}|C'E) \geq H^{\varepsilon}_{\min}(\rho_{XE}|E) - \mathrm{leak}_{\mathsf{IR}} ,$$

which, by the assumption on the length of the final key ℓ, implies (6.5) and thus concludes the proof. □

6.5 Quantum key distillation

We are now ready to describe and analyze a general quantum key distillation protocol, which uses parameter estimation and classical post-processing as discussed above. (For a high-level description of the content of this section, we refer to Section 1.6.)

6.5.1 Description of the protocol

Consider the quantum key distillation protocol $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ depicted in Fig. 6.4. Alice and Bob take inputs from product spaces $\mathcal{H}_A^{\otimes N}$ and $\mathcal{H}_B^{\otimes N}$, respectively. Then, they subsequently run the following subprotocols (see also Table 6.1):

- *Random permutation of the subsystems:* Alice and Bob reorder their subsystems according to a commonly chosen random permutation π.

- *Parameter estimation* (PE): Alice and Bob sacrifice m subsystems to perform some statistical checks. We assume that they do this using a protocol of the form $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ (see Fig. 6.1), which is characterized by a POVM $\mathcal{M} = \{M_w\}_{w\in\mathcal{W}}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ and a set $\mathcal{Q}$ of valid frequency distributions on $\mathcal{W}$.

- *Block-wise measurement and processing* ($\mathsf{BI}^{\otimes n}$): In order to obtain classical data, Alice and Bob apply a measurement to the remaining $b \cdot n$ subsystems, possibly followed by some further processing (e.g., advantage distillation). We assume here that Alice and Bob group their $b \cdot n$ subsystems in n blocks of size b and then process each of these blocks independently, according to some subprotocol, denoted BI. Each application of BI to a block $\mathcal{H}_A^{\otimes b} \otimes \mathcal{H}_B^{\otimes b}$ results in a pair of classical outputs x_i and y_i.

- *Classical post-processing* (PP): Alice and Bob transform their classical strings $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ into a pair of secret keys. For this, they invoke a post-processing subprotocol of the form $\mathsf{PP}_{\mathsf{IR},\mathcal{F}}$ (see Fig. 6.3), for some (arbitrary) information reconciliation scheme IR and a two-universal family of hash functions $\mathcal{F}$ for privacy amplification.

Fig. 6.4 Quantum key distillation protocol $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$.

Parameters:

PE:	parameter estimation protocol on $\mathcal{H}_A^{\otimes m} \otimes \mathcal{H}_B^{\otimes m}$.
BI:	subprotocol on $\mathcal{H}_A^{\otimes b} \otimes \mathcal{H}_B^{\otimes b}$ with classical output in $\mathcal{X} \times \mathcal{Y}$.
PP:	classical post-processing protocol on $\mathcal{X}^n \times \mathcal{Y}^n$.
N:	Number of input systems ($N \geq bn + m$)

Alice		**Bob**
input space: $\mathcal{H}_A^{\otimes N}$		input space: $\mathcal{H}_B^{\otimes N}$
$\pi \in_R \mathcal{S}_N$	$\xrightarrow{\pi}$	
permute subsyst.		permute subsyst.
$\mathcal{H}_A^{\otimes m}$	$\overset{\mathsf{PE}}{\rightleftarrows}$	$\mathcal{H}_B^{\otimes m}$ $\to$ acc./abort
$(\mathcal{H}_A^{\otimes b})^{\otimes n}$ $\to (x_1, \ldots, x_n)$	$\overset{\mathsf{BI}^{\otimes n}}{\rightleftarrows}$	$(\mathcal{H}_B^{\otimes b})^{\otimes n}$ $\to (y_1, \ldots, y_n)$
$(x_1, \ldots, x_n) \to s_A$	$\overset{\mathsf{PP}}{\rightleftarrows}$	$(y_1, \ldots, y_n) \to s_B$
output s_A		output s_B

Table 6.1 Subprotocols used for $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ (cf. Fig. 6.4).

$\mathsf{PE} := \mathsf{PE}_{\mathcal{M},\mathcal{Q}}$	prot. on $\mathcal{H}_A^{\otimes m} \otimes \mathcal{H}_B^{\otimes m}$ defined by Fig. 6.1
$\mathcal{M} = \{M_w\}_{w \in \mathcal{W}}$	POVM on $\mathcal{H}_A \otimes \mathcal{H}_B$
$\mathcal{Q}$	set of freq. dist. on $\mathcal{W}$
BI	prot. on $\mathcal{H}_A^{\otimes b} \otimes \mathcal{H}_B^{\otimes b}$ with cl. output in $\mathcal{X} \times \mathcal{Y}$
$\mathsf{PP} := \mathsf{PP}_{\mathsf{IR},\mathcal{F}}$	prot. on $\mathcal{X}^n \times \mathcal{Y}^n$ defined by Fig. 6.3
IR	inf. rec. prot. on $\mathcal{X}^n \times \mathcal{Y}^n$
$\mathcal{F}$	two-univ. fam. of hash func. from $\mathcal{X}^n$ to $\{0,1\}^\ell$

Table 6.2 Security parameters for $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ (cf. Fig. 6.4).

N	$bn + m + k$
r	$\frac{N}{k}\bigl(2\log(9/\varepsilon) + \dim(\mathcal{H}_A \otimes \mathcal{H}_B)^2 \ln k\bigr)$
δ'	$(\frac{5}{2}\log\lvert\mathcal{X}\rvert + 4)\sqrt{h(r/n) + \frac{2}{n}\log(18/\varepsilon)}$
μ	$2\sqrt{h(r/m) + \frac{1}{m}\bigl(\log(9/2\varepsilon) + \lvert\mathcal{W}\rvert\log(\frac{m}{2}+1)\bigr)}$
δ	$\delta' + \frac{2(m+k)}{n}\log\dim(\mathcal{H}_A \otimes \mathcal{H}_B) + \frac{2}{n}\log(3/2\varepsilon)$

6.5.2 Robustness

The usefulness of a key distillation protocol depends on the set of inputs for which it is robust, i.e., from which it can successfully distill secret keys. Obviously, the described protocol $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ is robust on all inputs for which none of its subprotocols PE, BI, or PP aborts. Note that the post-processing $\mathsf{PP} = \mathsf{PP}_{\mathsf{IR},\mathcal{F}}$ only aborts if the underlying information reconciliation scheme IR aborts.

Typically, the subprotocols BI and IR are chosen in such a way that they are robust on any of the input states accepted by PE. In this case, the key distillation protocol $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ is successful whenever it starts with an input for which PE is robust. According to the discussion in Section 6.2, the protocol $\mathsf{PE} = \mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ is robust on product states $\sigma_{AB}^{\otimes m}$ if σ_{AB} is contained in the set $\overline{\Gamma}_{\mathcal{M},\mathcal{Q}}^{\geq\mu}$ defined by (6.2). Consequently, $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ is robust on all inputs of the form $\sigma_{AB}^{\otimes N}$, for $\sigma_{AB} \in \overline{\Gamma}_{\mathcal{M},\mathcal{Q}}^{\geq\mu}$.

6.5.3 Security

The following is a generic criterion for the security of QKD.

Theorem 6.5.1. *Let $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ be the quantum key distillation protocol defined by Fig. 6.4 and Table 6.1, let $\varepsilon, \varepsilon' \geq 0$, let δ, μ be defined by Table 6.2, and let $\Gamma_{\mathcal{M},\mathcal{Q}}^{\leq\mu}$ be defined by* (6.1)*. Then $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ is $(\varepsilon + \varepsilon')$-fully secure if the underlying information reconciliation protocol IR is ε'-fully secure and if*

$$\ell \leq n \min_{\sigma_{AB} \in \Gamma_{\mathcal{M},\mathcal{Q}}^{\leq\mu}} H(X|\bar{E}) - \mathrm{leak}_{\mathsf{IR}} - n\delta \ ,$$

where the entropy in the minimum is evaluated on

$$\sigma_{XY\bar{E}} = \mathcal{E}^{\mathsf{BI}}_{XY\bar{E}\leftarrow A^b B^b E^b}(\sigma_{ABE}^{\otimes b}) \ ,$$

for a purification σ_{ABE} of σ_{AB}.

Proof. Let $\rho_{A^N B^N}$ be any state held by Alice and Bob after they have applied the random permutation π (averaged over all possible choices of π). Because, obviously, $\rho_{A^N B^N}$ is permutation-invariant, Lemma 4.2.2 implies that there exists a purification $\rho_{A^N B^N E^N}$ of $\rho_{A^N B^N}$ on the symmetric subspace of $(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)^{\otimes N}$. We show that the remaining part of the protocol is secure on $\rho_{A^N B^N E^N}$. This is sufficient because any density operator $\rho_{A^N B^N \tilde{E}}$ which has the property that taking the partial trace over $\mathcal{H}_{\tilde{E}}$ gives $\rho_{A^N B^N}$ can be obtained from the pure state $\rho_{A^N B^N E^N}$ by a trace-preserving CPM which only acts on Eve's space.

Let $\rho_{A^{bn+m} B^{bn+m} E^{bn+m}}$ be the operator obtained by taking the partial trace (over k subsystems $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$) of $\rho_{A^N B^N E^N}$. It describes the joint state on the m subsystems used for parameter estimation and the $b \cdot n$ subsystems which are given as input to $\mathsf{BI}^{\otimes n}$. According to the de Finetti representation theorem (Theorem 4.3.2) this density operator is approximated by a convex combination of density operators, where each of them is on the symmetric subspace along vectors $|\theta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$. More precisely, with $\bar{\varepsilon} := \frac{2\varepsilon}{9}$,

$$\Big\| \rho_{A^{bn+m} B^{bn+m} E^{bn+m}} - \int_{\mathcal{S}_1} \rho^{|\theta\rangle}_{A^{bn+m} B^{bn+m} E^{bn+m}} \nu(|\theta\rangle) \Big\|_1 \leq \bar{\varepsilon} \ , \qquad (6.7)$$

where the integral runs over the set $\mathcal{S}_1 := \mathcal{S}_1(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$ of normalized vectors in $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ and where, for any $|\theta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$,

$$\rho^{|\theta\rangle}_{A^{bn+m} B^{bn+m} E^{bn+m}} \in \mathcal{P}\big(\mathrm{Sym}((\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)^{\otimes bn+m}, |\theta\rangle^{\otimes bn+m-r})\big) \ . \quad (6.8)$$

We first analyze the situation after the parameter estimation is completed. Let $\mathcal{E}^{\mathsf{PE}}_{A^m B^m}$ be the CPM which maps all density operators on $(\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m}$ either to the scalar 0 or 1, depending on whether the parameter estimation protocol $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ accepts or aborts. Moreover, define

$$\begin{aligned} \rho^{\mathsf{PE}}_{A^{bn} B^{bn} E^N} &:= (\mathrm{id}_{A^{bn} B^{bn}} \otimes \mathcal{E}^{\mathsf{PE}}_{A^m B^m} \otimes \mathrm{id}_{E^N})(\rho_{A^{bn+m} B^{bn+m} E^N}) \\ \rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn} B^{bn} E^{bn}} &:= (\mathrm{id}_{A^{bn} B^{bn}} \otimes \mathcal{E}^{\mathsf{PE}}_{A^m B^m} \otimes \mathrm{id}_{E^{bn}})(\rho^{|\theta\rangle}_{A^{bn+m} B^{bn+m} E^{bn}}) \ . \end{aligned}$$

Because of (6.8), we have

$$\rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn} B^{bn} E^{bn}} \in \mathcal{P}\big(\mathrm{Sym}((\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)^{\otimes bn}, |\theta\rangle^{\otimes bn-r})\big) \ , \qquad (6.9)$$

for any $|\theta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$. Moreover, from (6.7) and the fact that the L_1-distance cannot increase when applying a quantum operation (Lemma A.2.1) we have

$$\Big\| \rho^{\mathsf{PE}}_{A^{bn} B^{bn} E^{bn}} - \int_{\mathcal{S}_1} \rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn} B^{bn} E^{bn}} \nu(|\theta\rangle) \Big\|_1 \leq \bar{\varepsilon} \ . \qquad (6.10)$$

According to Lemma 6.2.2, the parameter estimation $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ $\bar{\varepsilon}$-securely filters all states $\rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn} B^{bn} E^{bn}}$ for which $|\theta\rangle$ is not contained in the set

$$\mathcal{V}^{\mu} := \big\{ |\theta\rangle \in \mathcal{S}_1 : \ \mathrm{tr}_E(|\theta\rangle\langle\theta|) \in \Gamma^{\leq \mu}_{\mathcal{M},\mathcal{Q}} \big\} \ .$$

We can thus restrict the integral in (6.10) to the set $\mathcal{V}^\mu$, thereby only losing terms with total weight at most $\bar{\varepsilon}$, i.e.,

$$\left\| \rho^{\mathsf{PE}}_{A^{bn}B^{bn}E^{bn}} - \int_{\mathcal{V}^\mu} \rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn}B^{bn}E^{bn}} \nu(|\theta\rangle) \right\|_1 \leq 2\bar{\varepsilon} \ . \tag{6.11}$$

To describe the situation after the measurement and blockwise processing $\mathsf{Bl}^{\otimes n}$, we define

$$\begin{aligned} \rho_{X^nY^n\bar{E}^nE^{m+k}} &:= \big((\mathcal{E}^{\mathsf{Bl}}_{XY\bar{E}\leftarrow A^bB^bE^b})^{\otimes n} \otimes \mathrm{id}_{E^{m+k}}\big)(\rho^{\mathsf{PE}}_{A^{bn}B^{bn}E^{bn+m+k}}) \\ \rho^{|\theta\rangle}_{X^nY^n\bar{E}^n} &:= (\mathcal{E}^{\mathsf{Bl}}_{XY\bar{E}\leftarrow A^bB^bE^b})^{\otimes n}(\rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn}B^{bn}E^{bn}}) \ . \end{aligned}$$

Using once again the fact that the L_1-distance cannot decrease under quantum operations (Lemma A.2.1), we conclude from (6.11) that

$$\left\| \rho_{X^nY^n\bar{E}^n} - \int_{\mathcal{V}^\mu} \rho^{|\theta\rangle}_{X^nY^n\bar{E}^n} \nu(|\theta\rangle) \right\|_1 \leq 2\bar{\varepsilon} \ . \tag{6.12}$$

According to (6.9), the density operator $\rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn}B^{bn}E^{bn}}$ lies in the symmetric subspace of the $(b\cdot n)$-fold product space $(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)^{\otimes bn}$ along $|\theta\rangle^{bn-r}$, i.e., it has product form except on r subsystems. Equivalently, we can view $\rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn}B^{bn}E^{bn}}$ as a density operator on the n-fold product of subsystems $\mathcal{H}_{A^bB^bE^b} := \mathcal{H}_A^{\otimes b} \otimes \mathcal{H}_B^{\otimes b} \otimes \mathcal{H}_E^{\otimes b}$. It then has product form an all but (at most) r of these subsystems. That is, $\rho^{|\theta\rangle,\mathsf{PE}}_{A^{bn}B^{bn}E^{bn}}$ is contained in the symmetric subspace of $\mathcal{H}^{\otimes n}_{A^bB^bE^b}$ along $|\theta^b\rangle^{\otimes n-r}$, where $|\theta^b\rangle := |\theta\rangle^{\otimes b} \in \mathcal{H}_{A^bB^bE^b}$. This allows us to apply Theorem 4.4.1 in order to bound the entropy of the symmetric states $\rho^{|\theta\rangle}_{X^nY^n\bar{E}^n}$. With the definition

$$\sigma^{|\theta\rangle}_{XY\bar{E}} := \mathcal{E}^{\mathsf{Bl}}_{XY\bar{E}\leftarrow A^bB^bE^b}(\sigma^{\otimes b}_{ABE}) \ ,$$

where $\sigma_{ABE} := |\theta\rangle\langle\theta|$, we obtain

$$H^{\bar{\varepsilon}}_{\min}(\rho^{|\theta\rangle}_{X^n\bar{E}^n}|\bar{E}^n) \geq n\big(H(\sigma^{|\theta\rangle}_{X\bar{E}}) - H(\sigma^{|\theta\rangle}_{\bar{E}}) - \delta'\big) \ .$$

Consequently, using (6.12) together with the inequalities (3.19) and (3.20) of Theorem 3.2.12,

$$H^{3\bar{\varepsilon}}_{\min}(\rho_{X^n\bar{E}^n}|\bar{E}^n) \geq n \min_{|\theta\rangle\in\mathcal{V}^\mu} \big(H(\sigma^{|\theta\rangle}_{X\bar{E}}) - H(\sigma^{|\theta\rangle}_{\bar{E}}) - \delta'\big) \ .$$

Moreover, by the chain rule for smooth min-entropy (cf. (3.21) of Theorem 3.2.12)

$$\begin{aligned} &H^{3\bar{\varepsilon}}_{\min}(\rho_{X^n\bar{E}^nE^{m+k}}|\bar{E}^nE^{m+k}) \\ &\qquad \geq n \min_{|\theta\rangle\in\mathcal{V}^\mu} \big(H(\sigma^{|\theta\rangle}_{X\bar{E}}) - H(\sigma^{|\theta\rangle}_{\bar{E}}) - \delta'\big) - 2H_{\max}(\rho_{E^{m+k}}) \ . \end{aligned}$$

Finally, we use Lemma 6.4.1 which provides a criterion on the maximum length ℓ such that the secret key computed by the post-processing subprotocol PP is $(\varepsilon + \varepsilon')$-secure,

$$\ell \leq n \min_{|\theta\rangle \in \mathcal{V}^\mu} \bigl(H(\sigma^{|\theta\rangle}_{X\bar{E}}) - H(\sigma^{|\theta\rangle}_{\bar{E}}) - \delta'\bigr) - 2H_{\max}(\rho_{E^{m+k}}) - \mathrm{leak}_{\mathsf{IR}} - 2\log(3/2\varepsilon) \ .$$

The assertion then follows from

$$H_{\max}(\rho_{E^{m+k}}) \leq (m+k) \log \dim(\mathcal{H}_A \otimes \mathcal{H}_B) \ ,$$

the fact that $|\theta\rangle \in \mathcal{V}^\mu$ if and only if the trace σ_{AB} of $\sigma_{ABE} := |\theta\rangle\langle\theta|$ is contained in the set $\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}$, and the definition of δ (cf. Table 6.2). □

Note that the protocol $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ takes as input N subsystems and generates a key of a certain fixed length ℓ. In order to make asymptotic statements, we need to consider a family $\{\mathsf{QKD}^N_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}\}_{N\in\mathbb{N}}$ of such protocols, where, for any $N \in \mathbb{N}$, the corresponding protocol takes N input systems and generates a key of length $\ell(N)$. The *rate* of the protocol family is then defined by

$$\mathrm{rate} := \lim_{N\to\infty} \frac{\ell(N)}{N} \ .$$

Corollary 6.5.2. *Let $\delta, \mu > 0$, a protocol BI acting on blocks of length b, a POVM $\mathcal{M} = \{M_w\}_{w\in\mathcal{W}}$, and a set $\mathcal{Q}$ of probability distributions on $\mathcal{W}$ be fixed, and let $\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}$ be the set defined by (6.1). Then there exist $\gamma > 0$ and parameters $n = n(N), m = m(N), \ell = \ell(N)$ such that the class of protocols $\mathsf{QKD}^N_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ (parameterized by $N \in \mathbb{N}$) defined by Fig. 6.4 and Table 6.1 has rate*

$$\mathrm{rate} = \frac{1}{b} \min_{\sigma_{AB}\in\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}} H(X|\bar{E}) - H(X|Y) - \delta \ ,$$

where the entropies in the minimum are evaluated on

$$\sigma_{XY\bar{E}} = \mathcal{E}^{\mathsf{BI}}_{XY\bar{E}\leftarrow A^bB^bE^b}(\sigma^{\otimes b}_{ABE}) \ ,$$

for a purification σ_{ABE} of σ_{AB}. Moreover, for any $N \geq 0$, the protocol $\mathsf{QKD}^N_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ is $e^{-\gamma N}$-fully secure.

Proof. The statement follows directly from Theorem 6.5.1 combined with Corollary 6.3.5. □

6.6 Quantum key distribution

As described in Section 1.2, one can think of a quantum key *distribution* (QKD) protocol as a two-step process where Alice and Bob first use the quantum channel to distribute entanglement and then apply a quantum key *distillation* scheme to generate the final key pair. To prove security of a QKD protocol, it thus suffices to verify that the underlying key distillation protocol is secure on any input. Hence, the security results for key distillation protocols derived in the previous section (Theorem 6.5.1 and Corollary 6.5.2) directly apply to QKD protocols.

We can, however, further improve these results by taking into account that the way Alice and Bob use the quantum channel in the first step imposes some additional restrictions on the possible inputs to the distillation protocol. For example, if Alice locally prepares entangled states and then sends parts of them to Bob (note that this is actually the case for most QKD protocols, viewed as entanglement-based schemes), it is impossible for the adversary to tamper with the part belonging to Alice. Formally, this means that the partial state on Alice's subsystem is independent of Eve's attack.

Using this observation, we can restrict the set $\Gamma^{\leq\mu}_{\mathcal{M},\mathcal{Q}}$ of states σ_{AB} (as defined by (6.1)) over which the minimum is taken in the criterion of Theorem 6.5.1 and Corollary 6.5.2. In fact, it follows directly from Remark 4.3.3 that it suffices to consider states σ_{AB} such that $\sigma_A = \mathrm{tr}_B(\sigma_{AB})$ is fixed.

Chapter 7

Examples

To illustrate the general results of the previous chapter, we analyze certain concrete QKD protocols. We first specialize the formula for the rate (cf. Corollary 6.5.2) to protocols based on *two-level* quantum systems (Section 7.1). Then, as an example, we analyze different variants of the six-state protocol and compute explicit values for their rates (Section 7.2).

7.1 Protocols based on two-level systems

A large class of QKD protocols, including the well-known BB84 protocol or the six-state protocol, are based on an encoding of binary classical values using the state of a *two-level* quantum system, such as the the spin of a photon. For the corresponding key distillation protocol (see Fig. 6.4), this means that Alice and Bob take inputs from (products of) two-dimensional Hilbert spaces on which they apply binary measurements. In the following, we analyze different variants of such protocols.

7.1.1 One-way protocols

We start with a basic key distillation protocol which only uses information reconciliation and privacy amplification (as described in Section 6.4) to transform the raw key pair into a pair of secret keys. More precisely, after the measurement of their subsystems, Alice and Bob immediately invoke an information reconciliation protocol (e.g., the protocol $\mathsf{IR}_{\hat{\mathcal{X}},\mathcal{F}}$ depicted in Fig. 6.2) such that Bob can compute a guess of Alice's values; the final key is then obtained by two-universal hashing. Because this post-processing only requires communication from Alice to Bob, such protocols are also called *one-way key distillation protocols*.[1]

[1] Note, however, that bidirectional communication is always needed for the parameter estimation step.

Clearly, the one-way key distillation protocol described above is a special case of the general protocol $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ depicted in Fig 6.4, where $\mathsf{BI} := \mathsf{Meas}$ is the subprotocol describing the measurement operation of Alice and Bob. Additionally, assume that the parameter estimation subprotocol PE is the protocol $\mathsf{PE}_{\mathcal{M},\mathcal{Q}}$ depicted in Fig. 6.1, where $\mathcal{M}$ is a POVM and $\mathcal{Q}$ is the set of statistics for which the protocol does not abort. We can then use Corollary 6.5.2 to compute the rate of the protocol, that is,

$$\text{rate} = \min_{\sigma_{AB} \in \Gamma} H(X|E) - H(X|Y) \ . \tag{7.1}$$

Here, the minimum ranges over the set

$$\Gamma := \{\sigma_{AB} : P_W^{\sigma_{AB}} \in \mathcal{Q}\} \tag{7.2}$$

of all density operators σ_{AB} on the 2×2-dimensional Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ such that the measurement with respect to $\mathcal{M}$ gives a probability distribution $P_W^{\sigma_{AB}}$ which is contained in the set $\mathcal{Q}$. Moreover, the von Neumann (or Shannon) entropies $H(X|E)$ and $H(X|Y)$ are evaluated for the operators

$$\sigma_{XYE} := (\mathcal{E}^{\mathsf{Meas}}_{XY \leftarrow AB} \otimes \mathrm{id}_E)(\sigma_{ABE}) \ ,$$

where σ_{ABE} is a purification of σ_{AB}.

Let $\{|0\rangle_A, |1\rangle_A\}$ and $\{|0\rangle_B, |1\rangle_B\}$ be the bases that Alice and Bob use for the measurement Meas.[2] Lemma 7.1.1 below provides an explicit lower bound on the entropy difference on the right hand side of (7.1) as a function of σ_{ABE}. The bound only depends on the diagonal values of σ_{AB} with respect to the *Bell basis*, which is defined by the vectors

$$\begin{aligned}
|\Phi_0\rangle &:= \frac{1}{\sqrt{2}}|0,0\rangle + \frac{1}{\sqrt{2}}|1,1\rangle \\
|\Phi_1\rangle &:= \frac{1}{\sqrt{2}}|0,0\rangle - \frac{1}{\sqrt{2}}|1,1\rangle \\
|\Phi_2\rangle &:= \frac{1}{\sqrt{2}}|0,1\rangle + \frac{1}{\sqrt{2}}|1,0\rangle \\
|\Phi_3\rangle &:= \frac{1}{\sqrt{2}}|0,1\rangle - \frac{1}{\sqrt{2}}|1,0\rangle \ ,
\end{aligned}$$

where $|x, y\rangle := |x\rangle_A \otimes |y\rangle_B$.

Lemma 7.1.1. *Let both $\mathcal{H}_A$ and $\mathcal{H}_B$ be two-dimensional Hilbert spaces, let $\sigma_{ABE} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$ be a density operator, and let σ_{XYE} be obtained*

[2] Meas describes the measurement that generates the data used for the computation of the final key. It might be different from the measurement $\mathcal{M}$ which is used for parameter estimation.

from σ_{ABE} by applying orthonormal measurements on $\mathcal{H}_A$ and $\mathcal{H}_B$. Then

$$H(X|E) - H(X|Y) \\ \geq 1 - (\lambda_0 + \lambda_1) h\bigl(\frac{\lambda_0}{\lambda_0 + \lambda_1}\bigr) - (\lambda_2 + \lambda_3) h\bigl(\frac{\lambda_2}{\lambda_2 + \lambda_3}\bigr) - h(\lambda_0 + \lambda_1) ,$$

where $\lambda_i := \langle \Phi_i | \sigma_{AB} | \Phi_i \rangle$ are the diagonal values of σ_{AB} with respect to the Bell basis (defined relative to the measurement basis).

Proof. Let $\mathcal{D}$ be the CPM defined by

$$\mathcal{D}(\sigma_{AB}) := \frac{1}{4} \sum_{\tau \in \{\mathrm{id}, \sigma_x, \sigma_y, \sigma_z\}} \tau^{\otimes 2} \sigma_{AB} \tau^{\otimes 2} ,$$

where $\sigma_x, \sigma_y, \sigma_z$ are the Pauli operators

$$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{7.3}$$

and let $\tilde{\sigma}_{ABE}$ be a purification of $\tilde{\sigma}_{AB} := \mathcal{D}(\sigma_{AB})$. Moreover, let $\tilde{\sigma}_{ABE}$ be an arbitrary purification of $\tilde{\sigma}_{AB}$ with auxiliary system $\mathcal{H}_E$ and define

$$\tilde{\sigma}_{XYE} := (\mathcal{E}^{\mathsf{Meas}}_{XY \leftarrow AB} \otimes \mathrm{id}_E)(\tilde{\sigma}_{ABE}) .$$

A straightforward calculation shows that the operator $\tilde{\sigma}_{AB}$ has the form

$$\tilde{\sigma}_{AB} = \sum_{i=0}^{3} \lambda_i |\Phi_i\rangle\langle\Phi_i| ,$$

i.e., it is diagonal with respect to the Bell basis. Moreover, because $\mathcal{D}$ commutes with the measurement operation on $\mathcal{H}_A \otimes \mathcal{H}_B$, it is easy to verify that the entropy $H(X|Y)$ evaluated for σ_{XY} is upper bounded by the corresponding entropy for $\tilde{\sigma}_{XY}$. Similarly, because $\tilde{\sigma}_{ABE}$ is a purification of $\tilde{\sigma}_{AB}$, the entropy $H(X|E)$ evaluated for σ_{XE} is lower bounded by the entropy of $\tilde{\sigma}_{XE}$. It thus suffices to show that the inequality of the lemma holds for the operator $\tilde{\sigma}_{XYE}$, which is obtained from the diagonal operator $\tilde{\sigma}_{AB}$.

Let $|e_i\rangle_i$ be an orthonormal basis of a 4-dimensional Hilbert space $\mathcal{H}_E$. Then the operator $\tilde{\sigma}_{ABE} := |\Psi\rangle\langle\Psi| \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$ defined by

$$|\Psi\rangle := \sum_i \sqrt{\lambda_i} |\Phi_i\rangle_{AB} \otimes |e_i\rangle_E$$

is a purification of $\tilde{\sigma}_{AB}$. With the definition

$$\begin{aligned}
|f_{0,0}\rangle &:= \sqrt{\frac{\lambda_0}{2}}|e_0\rangle + \sqrt{\frac{\lambda_1}{2}}|e_1\rangle \\
|f_{1,1}\rangle &:= \sqrt{\frac{\lambda_0}{2}}|e_0\rangle - \sqrt{\frac{\lambda_1}{2}}|e_1\rangle \\
|f_{0,1}\rangle &:= \sqrt{\frac{\lambda_2}{2}}|e_2\rangle + \sqrt{\frac{\lambda_3}{2}}|e_3\rangle \\
|f_{1,0}\rangle &:= \sqrt{\frac{\lambda_2}{2}}|e_2\rangle - \sqrt{\frac{\lambda_3}{2}}|e_3\rangle \ ,
\end{aligned}$$

the state $|\Psi\rangle$ can be rewritten as

$$|\Psi\rangle = \sum_{x,y} |x,y\rangle \otimes |f_{x,y}\rangle \ .$$

Because the operator $\tilde{\sigma}_{XYE}$ is obtained from $\tilde{\sigma}_{ABE}$ by orthonormal measurements on $\mathcal{H}_A$ and $\mathcal{H}_B$, we conclude

$$\tilde{\sigma}_{XYE} = \sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \tilde{\sigma}_E^{x,y}$$

where $\tilde{\sigma}_E^{x,y} := |f_{x,y}\rangle\langle f_{x,y}|$.

Using this representation of the operator $\tilde{\sigma}_{XYE}$, it is is easy to see that

$$\begin{aligned}
H(\tilde{\sigma}_{XE}) &= 1 + h(\lambda_0 + \lambda_1) \\
H(\tilde{\sigma}_E) &= h(\lambda_0 + \lambda_1) + (\lambda_0 + \lambda_1) h\big(\frac{\lambda_0}{\lambda_0 + \lambda_1}\big) + (\lambda_2 + \lambda_3) h\big(\frac{\lambda_2}{\lambda_2 + \lambda_3}\big) \\
H(X|Y) &= h(\lambda_0 + \lambda_1) \ ,
\end{aligned}$$

from which the assertion follows. □

Using Lemma 7.1.1, we conclude that the above described one-way protocol can generate secret-key bits at rate

$$\begin{aligned}
\text{rate} \geq \min_{(\lambda_0,\ldots,\lambda_3)\in \text{diag}(\Gamma)} & 1 - (\lambda_0 + \lambda_1) h\big(\frac{\lambda_0}{\lambda_0 + \lambda_1}\big) \\
& - (\lambda_2 + \lambda_3) h\big(\frac{\lambda_2}{\lambda_2 + \lambda_3}\big) - h(\lambda_0 + \lambda_1) \ , \quad (7.4)
\end{aligned}$$

where $\text{diag}(\Gamma)$ denotes the 4-tuples of diagonal entries (relative to the Bell basis) of the operators $\sigma_{AB} \in \Gamma$, for Γ defined by (7.2).

7.1.2 One-way protocols with noisy preprocessing

The efficiency of the basic QKD protocol described in Section 7.1.1 can be increased in different ways. We consider an extension of the protocol where, before starting with information reconciliation, Alice applies some local *preprocessing* operation to her raw key. A very simple—but surprisingly useful—variant of preprocessing is to add noise, i.e., Alice flips each of her bits independently with some probability q. In the following, we call this *noisy preprocessing*.

To compute the rate of the one-way protocol enhanced with this type of preprocessing, we need a generalization of Lemma 7.1.1.

Lemma 7.1.2. *Let both $\mathcal{H}_A$ and $\mathcal{H}_B$ be two-dimensional Hilbert spaces, let $\sigma_{ABE} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$ be a density operator, and let σ_{XYE} be obtained from σ_{ABE} by applying orthonormal measurements on $\mathcal{H}_A$ and $\mathcal{H}_B$ where, additionally, the outcome of the measurement on $\mathcal{H}_A$ is flipped with probability $q \in [0,1]$. Then*

$$\begin{aligned} H(X|E) - H(X|Y) & \\ \geq 1 - (\lambda_0 + \lambda_1)\bigl(h(\alpha) - \bar{h}(\alpha,q)\bigr) &- (\lambda_2 + \lambda_3)\bigl(h(\beta) - \bar{h}(\beta,q)\bigr) \\ &- h\bigl((\lambda_0+\lambda_1)q + (\lambda_2+\lambda_3)(1-q)\bigr) \ , \end{aligned}$$

where $\lambda_i := \langle \Phi_i | \sigma_{AB} | \Phi_i \rangle$, $\alpha := \frac{\lambda_0}{\lambda_0+\lambda_1}$, $\beta := \frac{\lambda_2}{\lambda_2+\lambda_3}$, and

$$\bar{h}(p,q) := h\bigl(\frac{1}{2} \pm \frac{1}{2}\sqrt{1 - 16p(1-p)q(1-q)}\bigr) \ .$$

Proof. The statement follows by a straightforward extension of the proof of Lemma 7.1.1. □

Similarly to formula (7.4), the rate of the one-way protocol with noisy preprocessing—where Alice additionally flips her bits with probability q—is given by the expression provided by Lemma 7.1.2, minimized over all 4-tuples $(\lambda_0, \ldots, \lambda_3) \in \mathrm{diag}(\Gamma)$. It turns out that this rate is generally larger than the rate of the corresponding one-way protocol without preprocessing (see Section 7.2 below).

7.1.3 Protocols with advantage distillation

To further increase the efficiency of the key distillation protocol described above, one might insert an additional *advantage distillation* step after the measurement Meas, i.e., before the classical one-way post-processing.[3] Its purpose is to identify subsets of highly correlated bit pairs such as to separate these from only weakly correlated information.

[3] The concept of advantage distillation has first been introduced in a purely classical context [Mau93], where a secret key is generated from some predistributed correlated data.

Fig. 7.1 Advantage distillation protocol AD_b.

Parameters:

b: block length

Alice		**Bob**
input: $(x_1, \ldots, x_b)$		input: $(y_1, \ldots, y_b)$
$r \in_R \{0, 1\}$		
$(c_1, \ldots, c_b) :=$ $(x_1 \oplus r, \ldots, x_b \oplus r)$	$\xrightarrow{(c_1, \ldots, c_b)}$	if $(y_1 \oplus c_1, \ldots, y_b \oplus c_b)$ $\in \{\mathbf{0}, \mathbf{1}\}$ then acc := true
	$\xleftarrow{\text{acc}}$	
if acc then output x_1 else output Δ		if acc then output y_1 else output Δ

A typical advantage distillation protocol is depicted in Fig. 7.1: Alice and Bob split their bitstrings into blocks $(x_1, \ldots, x_b)$ and $(y_1, \ldots, y_b)$ of size b. Then, depending on a randomly chosen binary value r, Alice announces to Bob either $(x_1, \ldots, x_b)$ or $(x_1 \oplus 1, \ldots, x_b \oplus 1)$ (where $\oplus$ denotes the bitwise xor). Bob compares this information with his block $(y_1, \ldots, y_b)$ and accepts if it either differs in none or in all positions, i.e., if the difference equals either $\mathbf{0} := (0, \ldots, 0)$ or $\mathbf{1} := (1, \ldots, 1)$. In this case, Alice and Bob both keep the first bit of their initial string. Otherwise, they output some dummy symbol Δ.[4] Obviously, if the error probability per bit (i.e., the error rate of the channel) is e then the probability p_{succ} that advantage distillation on a block of length b is successful (i.e., Alice and Bob keep their bit) is $p_{\mathrm{succ}} = e^b + (1-e)^b$.

Let us now consider the general protocol $\mathsf{QKD}_{\mathsf{PE},\mathsf{BI},\mathsf{PP}}$ where the subprotocol BI consists of b binary measurements Meas of Alice and Bob followed by the advantage distillation protocol AD_b described in Fig. 7.1, i.e.,

$$\mathcal{E}^{\mathsf{BI}}_{XY\bar{E}\leftarrow A^bB^bE^b} = \mathcal{E}^{\mathsf{AD}}_{XY\bar{E}\leftarrow X^bY^bE^b} \circ (\mathcal{E}^{\mathsf{Meas}}_{XY\leftarrow AB} \otimes \mathrm{id}_E)^{\otimes b} \tag{7.5}$$

[4] As suggested in [Mau93], the efficiency of this advantage distillation protocol is further increased if Alice and Bob, instead of acting on large blocks at once, iteratively repeat the described protocol step on very small blocks (consisting of only 2 or 3 bits).

It is easy to see that the subprotocol AD_b commutes with the measurement Meas, that is, (7.5) can be rewritten as

$$\mathcal{E}^{\mathsf{BI}}_{XY\bar{E}\leftarrow A^b B^b E^b} = ((\mathcal{E}^{\mathsf{Meas}}_{XY\leftarrow AB})^{\otimes b} \otimes \mathrm{id}_{\bar{E}}) \circ \mathcal{E}^{\mathsf{AD}}_{AB\bar{E}\leftarrow A^b B^b E^b} \ .$$

Moreover, a straightforward computation[5] shows that, if σ_{AB} has diagonal entries $\lambda_0, \ldots, \lambda_3$ with respect to the Bell basis then, with probability

$$p_{\mathrm{succ}} := (\lambda_0 + \lambda_1)^b + (\lambda_2 + \lambda_3)^b \ ,$$

the advantage distillation AD_b is successful and the operation $\mathcal{E}^{\mathsf{AD}}_{AB\leftarrow A^b B^b}$ induced by AD_b (conditioned on the event that it is successful) maps $\sigma_{AB}^{\otimes b}$ to an operator $\tilde{\sigma}_{AB}$ with diagonal entries

$$\begin{aligned}
\tilde{\lambda}_0 &= \frac{(\lambda_0 + \lambda_1)^b + (\lambda_0 - \lambda_1)^b}{2 p_{\mathrm{succ}}} \\
\tilde{\lambda}_1 &= \frac{(\lambda_0 + \lambda_1)^b - (\lambda_0 - \lambda_1)^b}{2 p_{\mathrm{succ}}} \\
\tilde{\lambda}_2 &= \frac{(\lambda_2 + \lambda_3)^b + (\lambda_2 - \lambda_3)^b}{2 p_{\mathrm{succ}}} \\
\tilde{\lambda}_3 &= \frac{(\lambda_2 + \lambda_3)^b - (\lambda_2 - \lambda_3)^b}{2 p_{\mathrm{succ}}} \ .
\end{aligned}$$

Inserting these coefficients into the expressions provided by Lemma 7.1.1 gives a bound on the entropy difference which can be inserted into the formula for the rate (7.1).[6] We conclude that the key distillation protocol enhanced with advantage distillation on blocks of length b can generate key bits at rate

$$\begin{aligned}
\text{rate} \geq \frac{1}{b} \min_{(\lambda_0,\ldots,\lambda_3)\in \mathrm{diag}(\Gamma)} p_{\mathrm{succ}} \cdot \bigl(1 - (\tilde{\lambda}_0 + \tilde{\lambda}_1) h\bigl(\frac{\tilde{\lambda}_0}{\tilde{\lambda}_0 + \tilde{\lambda}_1}\bigr) \\
- (\tilde{\lambda}_2 + \tilde{\lambda}_3) h\bigl(\frac{\tilde{\lambda}_2}{\tilde{\lambda}_2 + \tilde{\lambda}_3}\bigr) - h(\tilde{\lambda}_0 + \tilde{\lambda}_1)\bigr) \ , \quad (7.6)
\end{aligned}$$

where Γ is the set defined by (7.2). Note that, in the special case where the block size b equals 1, the advantage distillation is trivial, that is, $\tilde{\lambda}_i = \lambda_i$, and (7.6) reduces to (7.4).

Similarly to the discussion in Section 7.1.2, one might enhance the protocol with noisy preprocessing on Alice's side, i.e., Alice flips her bits with some probability q after the advantage distillation step. The rate is then given by a formula similar to (7.6), where the expression in the minimum is

[5]For this computation, it is convenient to use the mapping $\mathcal{D}$ defined above, which allows to restrict the argument to the special case where σ_{AB} is Bell diagonal.

[6]Note that, conditioned on the event that AD_b is not successful (i.e., Alice and Bob's outputs are Δ), the entropy difference is zero.

replaced by the bound on the entropy difference provided by Lemma 7.1.2, evaluated for the coefficients $\tilde{\lambda}_i$.

Note that, as the block size b increases, the coefficients $\tilde{\lambda}_2$ and $\tilde{\lambda}_3$ approach zero, while $\tilde{\lambda}_0$ and $\tilde{\lambda}_1$ both tend to $\frac{1}{2}$. To get an approximation, it is thus sufficient to evaluate the expression of Lemma 7.1.2 up to small orders in λ_2 and λ_3.

Lemma 7.1.3. *Let $\lambda_0, \cdots \lambda_3$ and $\sigma_{XY\bar{E}}$ be defined as in Lemma 7.1.2, where $\lambda_0 = (1-\delta)\frac{1+\varepsilon}{2}$, $\lambda_1 = (1-\delta)\frac{1-\varepsilon}{2}$, $\lambda_2 = \lambda_3 = \frac{\delta}{2}$ for some $\delta, \varepsilon \geq 0$. Then*

$$H(X|\bar{E}) - H(X|Y) \geq \frac{4}{\ln 8}(1-\delta)\bigl(\varepsilon^2 - 6\delta\bigr)\bigl(\frac{1}{2} - q\bigr)^2 + O\bigl(\delta^3 + \varepsilon^3 + (\frac{1}{2} - q)^3\bigr) \ .$$

In particular, this quantity is positive if $\varepsilon^2 \geq 6\delta$.

Proof. The assertion follows immediately from a series expansion of the bound provided by Lemma 7.1.2 about $\varepsilon = 0$ and $\delta = 0$. □

Lemma 7.1.3 can be used to compute a bound on the rate of the protocol described above (advantage distillation followed by noisy preprocessing). Under the assumption that the coefficients $\tilde{\lambda}_0, \ldots, \tilde{\lambda}_3$ are of the form

$$\tilde{\lambda}_0 = (1-\delta)\frac{1+\varepsilon}{2}$$
$$\tilde{\lambda}_1 = (1-\delta)\frac{1-\varepsilon}{2}$$
$$\tilde{\lambda}_2 = \tilde{\lambda}_3 = \frac{\delta}{2} \ ,$$

for some small $\delta, \varepsilon \geq 0$, we get, analogously to (7.6),

$$\text{rate} \geq \frac{1}{b} \min_{(\lambda_0,\ldots,\lambda_3)\in\text{diag}(\Gamma)} p_{\text{succ}} \cdot \Bigl(\frac{4}{\ln 8}(1-\delta)\bigl(\varepsilon^2 - 6\delta\bigr)\bigl(\frac{1}{2} - q\bigr)^2 + O\bigl(\delta^3 + \varepsilon^3 + (\frac{1}{2} - q)^3\bigr)\Bigr) \ . \quad (7.7)$$

7.2 The six-state protocol

To illustrate the results of Section 7.1, we apply them to different variants of the six-state QKD protocol, for which we explicitly compute the rate and the maximum tolerated channel noise. The six-state protocol is one of the most efficient QKD schemes based on two-level systems, that is, the rate at which secret key bits can be generated per channel use is relatively close to the theoretical maximum. On the other hand, it is not very suitable for practical implementations, as it requires devices for preparing and measuring two-level quantum systems with respect to six different states.

7.2.1 Description

Instead of describing the actual six-state QKD protocol, we specify the underlying key distillation scheme: Alice and Bob take as input entangled two-level systems and measure each of them using at random one of three mutually unbiased bases, which results in a pair of raw keys.[7] Usually, these are the *rectilinear* or *z-basis* $\{|0\rangle_z, |1\rangle_z\}$, the *diagonal* or *x-basis* $\{|0\rangle_x, |1\rangle_x\}$, and the *circular* or *y-basis* $\{|0\rangle_y, |1\rangle_y\}$), which are related by

$$\begin{array}{rclrcl} |0\rangle_x &=& \frac{1}{\sqrt{2}}(|0\rangle_z + |1\rangle_z) & |0\rangle_y &=& \frac{1}{\sqrt{2}}(|0\rangle_z + i|1\rangle_z) \\ |1\rangle_x &=& \frac{1}{\sqrt{2}}(|0\rangle_z - |1\rangle_z) & |1\rangle_y &=& \frac{1}{\sqrt{2}}(|0\rangle_z - i|1\rangle_z) \end{array} .$$

Next, in a *sifting* step, Alice and Bob compare their choices of bases and discard all outcomes for which these do not agree. Note that, if Alice and Bob choose one of the bases with probability almost one, they only have to discard a small fraction of their raw keys (see discussion in Section 1.2).

In the parameter estimation step, Alice and Bob compare the bit values of their raw keys for a small fraction of randomly chosen positions. They abort if the *error rate* e—i.e., the fraction of positions for which their bits differ—is larger than some threshold. For the following analysis, we assume that Alice and Bob additionally check whether the error e is equally distributed among the different choices of the measurement bases and symmetric under bitflips.

Finally, Alice and Bob use the remaining part of their raw key to generate a pair of secret keys. For this, they might invoke different variants of advantage distillation and one-way post-processing subprotocols, as described in Section 7.1.

7.2.2 Analysis

To compute the rate of the six-state protocol (for different variants of the post-processing) we use the formulas derived in Section 7.1. The set Γ, as defined by (7.2), depends on the error rate e. For any fixed e, we get six conditions on the operators σ_{AB} contained in Γ, namely

$$(\langle b|_u \otimes \langle b'|_u)\sigma_{AB}(|b\rangle_u \otimes |b'\rangle_u) = \frac{e}{2} , \tag{7.8}$$

for any $u \in \{x, y, z\}$ and $b, b' \in \{0, 1\}$ with $b \neq b'$. It is easy to verify that the only density operator that satisfies these equalities is Bell-diagonal and has eigenvalues $\lambda_0 = 1 - \frac{3e}{2}$, $\lambda_1 = \lambda_2 = \lambda_3 = \frac{e}{2}$. Γ is thus the set of all

[7] Because each of the three bases consists of two orthonormal vectors, the information is encoded into six different states, which explains the name of the protocol.

density operators of the form (with respect to the Bell basis)

$$\sigma_{AB} = \begin{pmatrix} 1-\frac{3e}{2} & 0 & 0 & 0 \\ 0 & \frac{e}{2} & 0 & 0 \\ 0 & 0 & \frac{e}{2} & 0 \\ 0 & 0 & 0 & \frac{e}{2} \end{pmatrix} ,$$

for any $e \geq 0$ below some threshold.

One-way six-state protocol

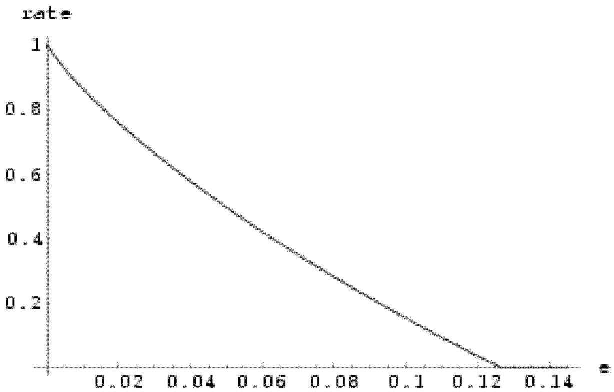
In a basic version of the six-state QKD protocol, Alice and Bob apply post-processing (i.e., information reconciliation followed by privacy amplification) directly to their measured data, as described in Section 7.1.1. The rate of this protocol can be computed using (7.4) where, according to the above discussion, $\lambda_0 = 1 - \frac{3e}{2}$ and $\lambda_1 = \lambda_2 = \lambda_3 = \frac{e}{2}$. Plot 7.1 shows the result of a numerical evaluation of this formula. In particular, the maximum tolerated channel noise for which the key rate is nonzero is 12.6%.

Next, we consider the one-way six-state protocol enhanced with additional noisy preprocessing as described in Section 7.1.2. That is, before the information reconciliation step, Alice applies random bitflips with probability q to her measurement outcomes. The rate of this protocol can be computed with Lemma 7.1.2. A little bit surprisingly, it turns out that noisy preprocessing increases its performance (see Plot 7.2). As shown in Plot 7.3, the optimal value of the bit-flip probability q depends on the error rate of the channel e. The protocol can tolerate errors up to 14.1% and thus beats the basic version (without noisy preprocessing) described above. Note that this result also improves on the previously best known lower bound for the maximum error tolerance of the six-state protocol with one-way processing, which was 12.7% [Lo00]. (Similarly, the same preprocessing can be applied to the BB84 protocol, in which case we get an error tolerance of 12.4%, compared to the best known value of 11.0% [SP00].)

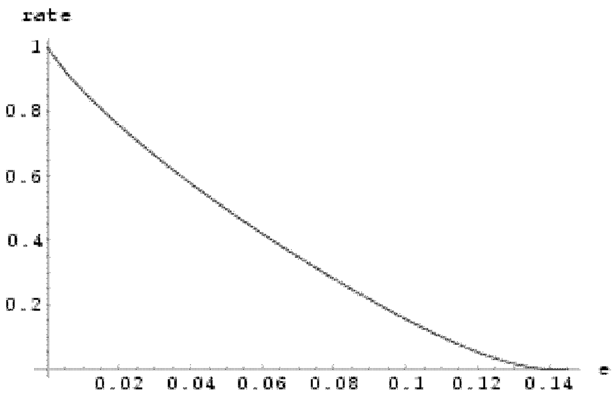
Six-state protocol with advantage distillation

The performance of the six-state protocol is increased if Alice and Bob additionally use advantage distillation as described in Section 7.1.3. For example, Alice and Bob might invoke the protocol AD_b depicted in Fig. 7.1 to process their measurement outcomes before the information reconciliation and privacy amplification step. The rate of the protocol is then given by (7.6). Because $\lambda_0 = 1 - \frac{3e}{2}$ and $\lambda_1 = \lambda_2 = \lambda_3 = \frac{e}{2}$, the coefficients $\tilde{\lambda}_i$ occurring in

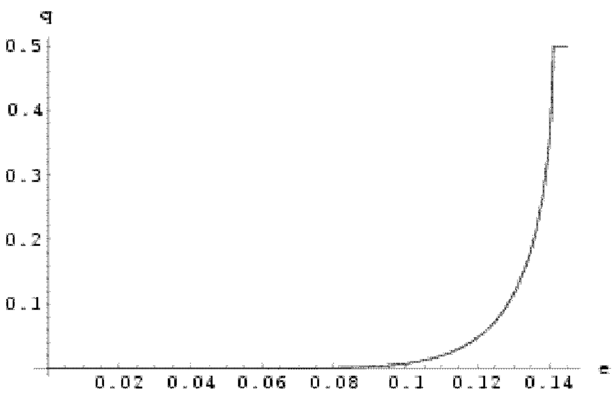
Plot 7.1 Rate of the basic one-way six-state protocol (without noisy preprocessing) as a function of the error rate e.



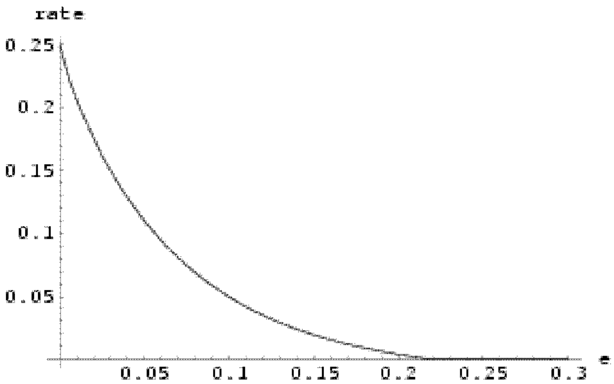
Plot 7.2 Rate of the one-way six-state protocol with noisy preprocessing (where Alice flips her bits with probability q as depicted in Plot 7.3).



Plot 7.3 Optimal value of the bit-flip probability q for the noisy preprocessing used in the one-way six-state protocol.



Plot 7.4 Rate of the six-state protocol with advantage distillation on blocks of length 4.



this formula are

$$\tilde{\lambda}_0 = \frac{(1-e)^b + (1-2e)^b}{2p_{\mathrm{succ}}}$$
$$\tilde{\lambda}_1 = \frac{(1-e)^b - (1-2e)^b}{2p_{\mathrm{succ}}}$$
$$\tilde{\lambda}_2 = \frac{e^b}{2p_{\mathrm{succ}}}$$
$$\tilde{\lambda}_3 = \frac{e^b}{2p_{\mathrm{succ}}} \ ,$$

where $p_{\mathrm{succ}} = (1-e)^b + e^b$. Plot 7.4 shows the result of this computation for a block size of $b = 4$.

Finally, we have a look at an extended protocol which combines advantage distillation and noisy preprocessing. That is, after the advantage distillation AD_b, Alice flips her bits with probability q (see Plot 7.5). For large block sizes b, the rate of the protocol is given by (7.7), for

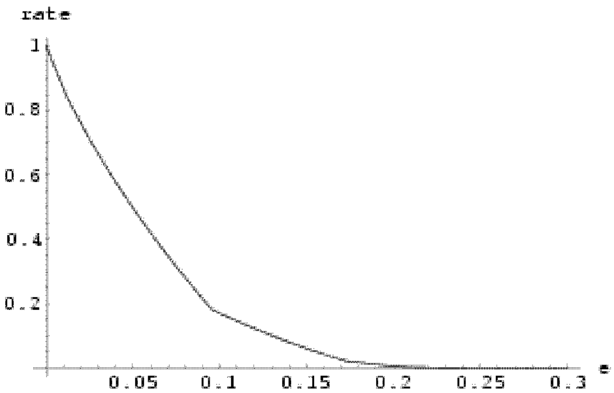
$$\delta = \frac{e^b}{(1-e)^b + e^b}$$
$$\varepsilon = \Bigl(\frac{1-2e}{1-e}\Bigr)^b \ .$$

In particular, for b approaching infinity, the secret-key rate is positive if (see Lemma 7.1.3)

$$\Bigl(\frac{1-2e}{1-e}\Bigr)^{2b} \geq 6 \frac{e^b}{(1-e)^b + e^b} \ .$$

Some simple analysis shows that this inequality is satisfied (for large b) if $e \leq \frac{1}{2} - \frac{\sqrt{5}}{10} \approx 0.276$. We conclude that the protocol can tolerate errors up to 27.6%.

Plot 7.5 Rate of the six-state protocol with advantage distillation (on blocks of optimal length) followed by (optimal) random bit-flips on Alice's side.



Note that this value coincides with the corresponding error tolerance of another variant of the six-state protocol due to Chau [Cha02] and is actually optimal for this class of protocols (cf. [ABB+04]). However, compared to Chau's protocol, the above described variant of the six-state protocol is simpler[8] and has a higher key rate.

[8]Instead of adding noise, Chau's protocol uses xor operations between different bits of the raw key.

Appendix A

Distance measures

A.1 Fidelity

The fidelity between two (not necessarily normalized) states $\rho, \rho' \in \mathcal{P}(\mathcal{H})$ is defined by

$$F(\rho, \rho') := \operatorname{tr}\sqrt{\rho^{1/2}\rho'\rho^{1/2}} \ .$$

In particular, if $\rho = |\psi\rangle\langle\psi|$ and $\rho' = |\psi'\rangle\langle\psi'|$ are pure states,

$$F(\rho, \rho') = |\langle\psi|\psi'\rangle| \ .$$

Remark A.1.1. For any $\alpha, \beta \in \mathbb{R}^+$,

$$F(\alpha\rho, \beta\rho') = \sqrt{\alpha\beta}F(\rho, \rho') \ .$$

Fidelity of purifications

Uhlmann's theorem states that the fidelity between two operators is equal to the maximum fidelity of their purifications.

Theorem A.1.2 (Uhlmann). *Let $\rho, \rho' \in \mathcal{P}(\mathcal{H})$ and let $|\psi\rangle\langle\psi|$ be a purification of ρ. Then*

$$F(\rho, \rho') = \max_{|\psi'\rangle\langle\psi'|} F(|\psi\rangle\langle\psi|, |\psi'\rangle\langle\psi'|)$$

where the maximum is taken over all purifications $|\psi'\rangle\langle\psi'|$ of ρ'.

Proof. The assertion follows directly from the corresponding statement for normalized density operators (see, e.g., Theorem 9.4 in [NC00]) and Remark A.1.1. □

Remark A.1.3. Because the fidelity $F(|\psi\rangle\langle\psi|, |\psi'\rangle\langle\psi'|)$ does not depend on the phase of the vectors, the vector $|\psi'\rangle$ which maximizes the expression of Theorem A.1.2 can always be chosen such that $\langle\psi|\psi'\rangle$ is real and nonnegative.

Fidelity and quantum operations

The fidelity between two density operators is equal to the minimum fidelity between the distributions of the outcomes resulting from a measurement.

Lemma A.1.4. *Let $\rho, \rho' \in \mathcal{P}(\mathcal{H})$. Then*

$$F(\rho, \rho') = \min_{\{M_z\}_z} F(P_Z, P'_Z)$$

where the minimum ranges over all POVMs $\{M_z\}_{z \in \mathcal{Z}}$ on $\mathcal{H}$ and where $P_Z, P'_Z \in \mathcal{P}(\mathcal{Z})$ are defined by $P_Z(z) = \mathrm{tr}(\rho M_z)$ and $P'_Z(z) = \mathrm{tr}(\rho' M_z)$, respectively.

Proof. The statement follows directly from the corresponding statement for normalized density operators (cf. formula (9.74) in [NC00]) and Remark A.1.1. □

The fidelity between two operators cannot decrease when applying the same quantum operation to both of them.

Lemma A.1.5. *Let $\rho, \rho' \in \mathcal{P}(\mathcal{H})$ and let $\mathcal{E}$ be a trace-preserving CPM on $\mathcal{H}$. Then*

$$F(\mathcal{E}(\rho), \mathcal{E}(\rho')) \geq F(\rho, \rho') \ .$$

Proof. See Theorem 9.6 of [NC00] and Remark A.1.1. □

A.2 L_1-distance

L_1-distance and quantum operations

The L_1-distance between two density operators cannot increase when applying the same (trace-preserving) quantum operation to both of them.

Lemma A.2.1. *Let $\rho, \rho' \in \mathcal{P}(\mathcal{H})$ and let $\mathcal{E}$ be a CPM such that $\mathrm{tr}(\mathcal{E}(\sigma)) \leq \mathrm{tr}(\sigma)$ for any $\sigma \in \mathcal{P}(\mathcal{H})$. Then*

$$\|\mathcal{E}(\rho) - \mathcal{E}(\rho')\|_1 \leq \|\rho - \rho'\|_1 \ .$$

Proof. It suffices to show that $\|\mathcal{E}(T)\|_1 \leq \|T\|_1$, for any hermitian operator T. The assertion then follows with $T := \rho - \rho'$ because $\mathcal{E}$ is linear.

For any hermitian operator S, let $\|S\|_\infty := \sup_{|\phi\rangle \in \mathcal{H} : \||\phi\rangle\| \leq 1} \|S|\phi\rangle\|$ be the *L_∞-operator norm*. Note that the L_∞-operator norm can equivalently be written as

$$\|S\|_\infty = \sup_{\sigma \in \mathcal{P}(\mathcal{H}) : \mathrm{tr}(\sigma) \leq 1} \mathrm{tr}(S\sigma) \ .$$

Moreover, it is easy to see that for any hermitian operator T

$$\|T\|_1 = \sup_{S:\, \|S\|_\infty \leq 1} |\mathrm{tr}(ST)| \ . \tag{A.1}$$

Let $\{E_k\}_k$ be the family of linear operators from $\mathcal{H}$ to $\mathcal{H}'$ defined by the CPM $\mathcal{E}$, i.e., $\mathcal{E}(\sigma) = E_k \sigma E_k^\dagger$, for any $\sigma \in \mathcal{P}(\mathcal{H})$. Moreover, let $\mathcal{E}^\dagger$ be the CPM defined by $\mathcal{E}^\dagger(S') := \sum_k E_k^\dagger S' E_k$, for any hermitian operator S' on $\mathcal{H}'$. We then have the identity

$$\mathrm{tr}(\mathcal{E}^\dagger(S')\sigma) = \mathrm{tr}(S'\mathcal{E}(\sigma)) \ .$$

Hence

$$\begin{aligned} \|\mathcal{E}^\dagger(S')\|_\infty &= \sup_{\sigma \in \mathcal{P}(\mathcal{H}):\, \mathrm{tr}(\sigma) \leq 1} \mathrm{tr}(\mathcal{E}^\dagger(S')\sigma) \\ &= \sup_{\sigma \in \mathcal{P}(\mathcal{H}):\, \mathrm{tr}(\sigma) \leq 1} \mathrm{tr}(S'\mathcal{E}(\sigma)) \\ &\leq \|S'\|_\infty \ , \end{aligned} \tag{A.2}$$

where the inequality holds because $\mathcal{E}(\sigma) \in \mathcal{P}(\mathcal{H}')$ and $\mathrm{tr}(\mathcal{E}(\sigma)) \leq \mathrm{tr}(\sigma) = 1$, for any $\sigma \in \mathcal{P}(\mathcal{H})$. Using (A.1), this implies that

$$\begin{aligned} \|\mathcal{E}(T)\|_1 &= \sup_{S':\, \|S'\|_\infty \leq 1} |\mathrm{tr}(\mathcal{E}(T)S')| \\ &= \sup_{S':\, \|S'\|_\infty \leq 1} |\mathrm{tr}(T\mathcal{E}^\dagger(S'))| \\ &\leq \sup_{S:\, \|S\|_\infty \leq 1} \mathrm{tr}(TS) \\ &= \|T\|_1 \ , \end{aligned}$$

where the inequality follows from (A.2).

□

L_1-distance of mixtures

Lemma A.2.2. *Let ρ_{AZ} and $\bar{\rho}_{AZ}$ be classical with respect to an orthonormal basis $\{|z\rangle\}_{z\in\mathcal{Z}}$ of $\mathcal{H}_Z$ and let $\{\rho_A^z\}_{z\in\mathcal{Z}}$ and $\{\bar{\rho}_A^z\}_{z\in\mathcal{Z}}$ be the corresponding conditional operators. Then*

$$\left\|\rho_{AZ} - \bar{\rho}_{AZ}\right\|_1 = \sum_{z\in\mathcal{Z}} \left\|\rho_A^z - \bar{\rho}_A^z\right\|_1 \ .$$

Proof. For any $z \in \mathcal{Z}$, let $\{|\phi_x^z\rangle\}_{x\in\mathcal{X}}$ be an eigenbasis of $\rho_A^z - \bar{\rho}_A^z$. Then, the family $\{|\phi_x^z\rangle \otimes |z\rangle\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ is an eigenbasis of $\rho_{AZ} - \bar{\rho}_{AZ}$. Hence,

$$\begin{aligned}\big\|\rho_{AZ} - \bar{\rho}_{AZ}\big\|_1 &= \sum_{z'\in\mathcal{Z}}\sum_{x\in\mathcal{X}}\big|(\langle\phi_x^{z'}| \otimes \langle z'|)\big(\sum_{z\in\mathcal{Z}}(\rho_A^z - \bar{\rho}_A^z) \otimes |z\rangle\langle z|\big)(|\phi_x^{z'}\rangle \otimes |z'\rangle)\big| \\ &= \sum_{z\in\mathcal{Z}}\sum_{x\in\mathcal{X}}\big|\langle\phi_x^z|\rho_A^z - \bar{\rho}_A^z|\phi_x^z\rangle\big| \\ &= \sum_{z\in\mathcal{Z}}\big\|\rho_A^z - \bar{\rho}_A^z\big\|_1 .\end{aligned}$$

□

L_1-distance of pure operators in terms of vector distance

The scalar product of a Hilbert space $\mathcal{H}$ induces a canonical norm, defined by $\||\phi\rangle\| := \sqrt{\langle\phi|\phi\rangle}$, for any $|\phi\rangle \in \mathcal{H}$. In particular, the norm of the difference between two vectors $|\psi\rangle$ and $|\psi'\rangle$, $\||\psi\rangle - |\psi'\rangle\|$, is a metric on $\mathcal{H}$.

The following lemma relates the L_1-distance between two pure states $|\psi\rangle\langle\psi|$ and $|\psi'\rangle\langle\psi'|$ to the vector distance $\||\psi\rangle - |\psi'\rangle\|$.

Lemma A.2.3. *Let $|\psi\rangle, |\psi'\rangle \in \mathcal{H}$ such that $\langle\psi|\psi'\rangle$ is real. Then*

$$\big\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\big\|_1 = \big\||\psi\rangle - |\psi'\rangle\big\| \cdot \big\||\psi\rangle + |\psi'\rangle\big\| .$$

Proof. Define $|\alpha\rangle := |\psi\rangle + |\psi'\rangle$, $|\beta\rangle := |\psi\rangle - |\psi'\rangle$ and let $a := \||\alpha\rangle\|$, $b := \||\beta\rangle\|$. We then have

$$\big\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\big\|_1 = \mathrm{tr}\big||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\big| = \frac{1}{2}\mathrm{tr}\big||\alpha\rangle\langle\beta| + |\beta\rangle\langle\alpha|\big| .$$

Moreover, because $\langle\psi|\psi'\rangle$ is real, the scalar product $\langle\alpha|\beta\rangle = \langle\psi|\psi\rangle - \langle\psi'|\psi'\rangle$ is real as well. Using this, it is easy to verify that $b|\alpha\rangle + a|\beta\rangle$ and $b|\alpha\rangle - a|\beta\rangle$ are eigenvectors of $|\alpha\rangle\langle\beta| + |\beta\rangle\langle\alpha|$ with eigenvalues $\langle\alpha|\beta\rangle + ab$ and $\langle\alpha|\beta\rangle - ab$, respectively. Hence,

$$\mathrm{tr}\big||\alpha\rangle\langle\beta| + |\beta\rangle\langle\alpha|\big| = \big|\langle\alpha|\beta\rangle + ab\big| + \big|\langle\alpha|\beta\rangle - ab\big| = 2ab ,$$

where the last equality holds because the Cauchy-Schwartz inequality implies $|\langle\alpha|\beta\rangle| \leq ab$. □

Upper bound on L_1-distance in terms of fidelity

Lemma A.2.4. *Let $\rho, \rho \in \mathcal{P}(\mathcal{H})$. Then*

$$\|\rho - \rho'\|_1 \leq \sqrt{\big(\mathrm{tr}(\rho) + \mathrm{tr}(\rho')\big)^2 - 4F(\rho, \rho')^2} .$$

Proof. It follows from Uhlmann's theorem (see Theorem A.1.2 and remark thereafter) that there exist purifications $|\psi\rangle\langle\psi|$ and $|\psi'\rangle\langle\psi'|$ of ρ and ρ', respectively, such that $\langle\psi|\psi'\rangle$ is nonnegative and $F(\rho,\rho') = F(|\psi\rangle\langle\psi|, |\psi'\rangle\langle\psi'|)$. Using Lemma A.2.3, a simple calculation leads to

$$\big\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\big\| = \sqrt{(\langle\psi|\psi\rangle + \langle\psi'|\psi'\rangle)^2 - 4\langle\psi|\psi'\rangle^2} \ .$$

Since $\langle\psi|\psi\rangle = \mathrm{tr}(|\psi\rangle\langle\psi|) = \mathrm{tr}(\rho)$, $\langle\psi'|\psi'\rangle = \mathrm{tr}(\rho')$, and $F(|\psi\rangle\langle\psi|, |\psi'\rangle\langle\psi'|) = \langle\psi|\psi'\rangle$, this identity can be rewritten as

$$\big\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\big\| = \sqrt{\big(\mathrm{tr}(\rho) + \mathrm{tr}(\rho')\big)^2 - 4F(\rho,\rho')^2} \ .$$

The assertion then follows from the fact that the L_1-distance can only decrease when taking the partial trace (cf. Lemma A.2.1). □

Upper bound on L_1-distance in terms of vector distance

The following lemma is a generalization of one direction of Lemma A.2.3 to mixed states.

Lemma A.2.5. *Let $\rho, \rho' \in \mathcal{P}(\mathcal{H})$ and let $|\psi\rangle\langle\psi|$ and $|\psi'\rangle\langle\psi'|$ be purifications of ρ and ρ', respectively. Then*

$$\|\rho - \rho'\|_1 \leq \big(\sqrt{\mathrm{tr}(\rho)} + \sqrt{\mathrm{tr}(\rho')}\big) \cdot \big\||\psi\rangle - |\psi'\rangle\big\| \ .$$

Proof. Let $\nu \in [0, 2\pi]$ such that $e^{i\nu}\langle\psi|\psi'\rangle$ is nonnegative and define $|\tilde{\psi}'\rangle := e^{i\nu}|\psi'\rangle$. Then, from Lemma A.2.3,

$$\begin{aligned}\big\||\psi\rangle\langle\psi| - |\tilde{\psi}'\rangle\langle\tilde{\psi}'|\big\|_1 &= \big\||\psi\rangle - |\tilde{\psi}'\rangle\big\| \cdot \big\||\psi\rangle + |\tilde{\psi}'\rangle\big\| \\ &\leq \big\||\psi\rangle - |\tilde{\psi}'\rangle\big\| \cdot \big(\big\||\psi\rangle\big\| + \big\||\psi'\rangle\big\|\big)\end{aligned} \tag{A.3}$$

where the inequality follows from the triangle inequality for the norm $\|\cdot\|$ and $\big\||\tilde{\psi}'\rangle\big\| = \big\||\psi'\rangle\big\|$. Moreover, since $\langle\psi|\tilde{\psi}'\rangle$ is nonnegative, it cannot be smaller than the real value of the scalar product $\langle\psi|\psi'\rangle$, that is, $\Re(\langle\psi|\tilde{\psi}'\rangle) = |\langle\psi|\tilde{\psi}'\rangle| = |\langle\psi|\psi'\rangle| \geq \Re(\langle\psi|\psi'\rangle)$, and thus

$$\begin{aligned}\big\||\psi\rangle - |\tilde{\psi}'\rangle\big\| &= \sqrt{\langle\psi|\psi\rangle + \langle\tilde{\psi}'|\tilde{\psi}'\rangle - 2\Re(\langle\psi|\tilde{\psi}'\rangle)} \\ &\leq \sqrt{\langle\psi|\psi\rangle + \langle\psi'|\psi'\rangle - 2\Re(\langle\psi|\psi'\rangle)} \\ &= \big\||\psi\rangle - |\psi'\rangle\big\| \ .\end{aligned}$$

Combining this with (A.3) gives

$$\begin{aligned}\big\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\big\| &= \big\||\psi\rangle\langle\psi| - |\tilde{\psi}'\rangle\langle\tilde{\psi}'|\big\| \\ &\leq \big(\big\||\psi\rangle\big\| + \big\||\psi'\rangle\big\|\big) \cdot \big\||\psi\rangle - |\psi'\rangle\big\| \ .\end{aligned}$$

The assertion follows from the fact that the L_1-distance cannot increase when taking the partial trace (cf. Lemma A.2.1). □

Lower bound on L_1-distance in terms of fidelity

The following statement is the converse of Lemma A.2.4.

Lemma A.2.6. *Let $\rho, \rho' \in \mathcal{P}(\mathcal{H})$. Then*

$$\mathrm{tr}(\rho) + \mathrm{tr}(\rho') - 2F(\rho, \rho') \leq \|\rho - \rho'\|_1 \ .$$

The proof is a direct generalization of an argument given in [NC00] (see formula (9.109) of [NC00]).

Proof. According to Lemma A.1.4, there exists a POVM $\mathcal{M} = \{M_z\}_{z \in \mathcal{Z}}$ such that

$$F(\rho, \rho') = F(P_Z, P'_Z) \ ,$$

for P_Z and P'_Z defined by $P_Z(z) = \mathrm{tr}(\rho M_z)$ and $P'_Z(z) = \mathrm{tr}(\rho' M_z)$. Using the abbreviation $p_z := P_Z(z)$ and $p'_z := P_{Z'}$, we observe that

$$\begin{aligned} \sum_{z \in \mathcal{Z}} \bigl(\sqrt{p_z} - \sqrt{p'_z}\bigr)^2 &= \sum_{z \in \mathcal{Z}} \bigl(p_z + p'_z - 2\sqrt{p_z p'_z}\bigr) \\ &= \mathrm{tr}(\rho) + \mathrm{tr}(\rho') - 2F(\rho, \rho') \ . \end{aligned} \tag{A.4}$$

Moreover, because $\bigl|\sqrt{p_z} - \sqrt{p'_z}\bigr| \leq \sqrt{p_z} + \sqrt{p'_z}$,

$$\begin{aligned} \sum_{z \in \mathcal{Z}} \bigl(\sqrt{p_z} - \sqrt{p'_z}\bigr)^2 &\leq \sum_{z \in \mathcal{Z}} \bigl|\sqrt{p_z} - \sqrt{p'_z}\bigr| \cdot \bigl(\sqrt{p_z} + \sqrt{p'_z}\bigr) \\ &= \sum_{z \in \mathcal{Z}} \bigl|p_z - p'_z\bigr| \\ &\leq \|\rho - \rho'\|_1 \ , \end{aligned}$$

where the last inequality follows from the fact that the trace distance cannot increase when applying a POVM (cf. Lemma A.2.1). The assertion then follows by combining this with (A.4). □

Lower bound on L_1-distance in terms of vector distance

The following statement can be seen as the converse of Lemma A.2.5.

Lemma A.2.7. *Let $\rho, \rho' \in \mathcal{P}(\mathcal{H})$ and let $|\psi\rangle\langle\psi|$ be a purification of ρ. Then there exists a purification $|\psi'\rangle\langle\psi'|$ of ρ' such that*

$$\bigl\||\psi\rangle - |\psi'\rangle\bigr\| \leq \sqrt{\|\rho - \rho'\|_1} \ .$$

Proof. Uhlmann's theorem (see Theorem A.1.2 and remark thereafter) implies that there exists a purification $|\psi'\rangle\langle\psi'|$ of ρ' such that $F(\rho, \rho') = \langle\psi|\psi'\rangle$. Hence,

$$\begin{aligned} \bigl\||\psi\rangle - |\psi'\rangle\bigr\| &= \sqrt{\langle\psi|\psi\rangle + \langle\psi'|\psi'\rangle - \langle\psi|\psi'\rangle - \langle\psi'|\psi\rangle} \\ &= \sqrt{\mathrm{tr}(\rho) + \mathrm{tr}(\rho') - 2F(\rho, \rho')} \ . \end{aligned}$$

The assertion then follows from Lemma A.2.6. □

L_1-distance and trace

A slightly different variant of the following statement is known as the *Gentle Measurement Lemma* [Win99].

Lemma A.2.8. *Let $\rho, \bar{\rho} \in \mathcal{P}(\mathcal{H})$ such that $\bar{\rho} = P\rho P$ for some projector P on $\mathcal{H}$. Then,*

$$\|\rho - \bar{\rho}\|_1 \leq 2\sqrt{\operatorname{tr}(\rho)\big(\operatorname{tr}(\rho) - \operatorname{tr}(\bar{\rho})\big)} \ .$$

Proof. We first show that the assertion holds if ρ is normalized (i.e., $\operatorname{tr}(\rho) = 1$) and pure, that is, $\rho = |\phi\rangle\langle\phi|$ for some normalized vector $|\phi\rangle$. Since P is a projector, the vector $|\phi\rangle$ can be written as a weighted sum of two orthonormal vectors $|a\rangle$ and $|b\rangle$, $|\phi\rangle = \alpha|a\rangle + \beta|b\rangle$, for $\alpha, \beta \geq 0$, such that $P|a\rangle = |a\rangle$ and $P|b\rangle = 0$. In particular, $\bar{\rho} = \alpha^2|a\rangle\langle a|$. A straightforward calculation then shows that

$$\begin{aligned} \|\rho - \bar{\rho}\|_1 &= \big\|(\alpha|a\rangle + \beta|b\rangle)(\alpha\langle a| + \beta\langle b|) - \alpha^2|a\rangle\langle a|\big\|_1 \\ &\leq 2\beta = 2\sqrt{1 - \operatorname{tr}(\bar{\rho})} \end{aligned}$$

which concludes the proof for normalized pure states ρ.

To show that the assertion holds for general operators $\rho \in \mathcal{P}(\mathcal{H})$, let $\rho = \sum_{x \in \mathcal{X}} p_x |x\rangle\langle x|$ be a spectral decomposition of ρ. In particular, $\sum_{x \in \mathcal{X}} p_x = \operatorname{tr}(\rho)$. Define $\rho_x := |x\rangle\langle x|$ and $\bar{\rho}_x := P\rho_x P$. By linearity, we have

$$\bar{\rho} = P\rho P = \sum_{x \in \mathcal{X}} p_x \bar{\rho}_x \ .$$

Hence, using the triangle inequality and the fact that the assertion holds for the normalized pure states ρ_x, we find

$$\|\rho - \bar{\rho}\|_1 \leq \sum_{x \in \mathcal{X}} p_x \|\rho_x - \bar{\rho}_x\|_1 \leq 2 \sum_{x \in \mathcal{X}} p_x \sqrt{1 - \operatorname{tr}(\bar{\rho}_x)} \ .$$

Moreover, with Jensen's inequality we find

$$\begin{aligned} \sum_{x \in \mathcal{X}} p_x \sqrt{1 - \operatorname{tr}(\bar{\rho}_x)} &= \operatorname{tr}(\rho) \sum_{x \in \mathcal{X}} \frac{p_x}{\operatorname{tr}(\rho)} \sqrt{1 - \operatorname{tr}(\bar{\rho}_x)} \\ &\leq \operatorname{tr}(\rho) \sqrt{\sum_{x \in \mathcal{X}} \frac{p_x}{\operatorname{tr}(\rho)} \big(1 - \operatorname{tr}(\bar{\rho}_x)\big)} \\ &= \sqrt{\operatorname{tr}(\rho)\big(\operatorname{tr}(\rho) - \operatorname{tr}(\bar{\rho})\big)} \ , \end{aligned}$$

which concludes the proof. □

Appendix B

Various Technical Results

B.1 Combinatorics

For proofs of the following statements, we refer to the standard literature on combinatorics.

Lemma B.1.1. *The set $\mathcal{Q}_n^{\mathcal{X}}$ of types with denominator n on a set $\mathcal{X}$ has cardinality*

$$|\mathcal{Q}_n^{\mathcal{X}}| = \binom{n + |\mathcal{X}| - 1}{n} .$$

Lemma B.1.2. *Let $Q \in \mathcal{Q}_n^{\mathcal{X}}$ be a type with denominator n on a set $\mathcal{X}$. Then the type class Λ_n^Q has cardinality*

$$|\Lambda_n^Q| = \frac{n!}{\prod_{x \in \mathcal{X}} (nQ(x))!} .$$

Lemma B.1.3. *A set of cardinality n has at most $2^{nh(r/n)}$ subsets of cardinality r.*

Proof. A set of cardinality n has exactly $\binom{n}{r}$ subsets of cardinality r. The assertion thus follows from the inequality[1] $\binom{n}{r} \leq 2^{nh(r/n)}$. □

B.2 Birkhoff's Theorem

Definition B.2.1. A matrix $(a_{x,y})_{x \in \mathcal{X}, y \in \mathcal{Y}}$ is *bistochastic* if $a_{x,y} \geq 0$, for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $\sum_{y \in \mathcal{Y}} a_{x,y} = \sum_{x \in \mathcal{X}} a_{x,y} = 1$.

It is easy to see that a matrix $(a_{x,y})_{x \in \mathcal{X}, y \in \mathcal{Y}}$ can only be bistochastic if $|\mathcal{X}| = |\mathcal{Y}|$. The following theorem due to Birkhoff [Bir46] states that any bistochastic matrix can be written as a mixture of permutation matrices. (See, e.g., [HJ85] for a proof.)

[1] See, e.g., [CT91], Formula (12.40).

Theorem B.2.2 (Birkhoff's theorem). *Let $(a_{x,y})_{x\in\mathcal{X},y\in\mathcal{Y}}$ be a bistochastic matrix. Then there exist nonnegative coefficients μ_π, parameterized by the bijections π from $\mathcal{Y}$ to $\mathcal{X}$, such that $\sum_\pi \mu_\pi = 1$ and, for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$,*[2]

$$a_{x,y} = \sum_\pi \mu_\pi \delta_{x,\pi(y)} \ .$$

It follows immediately from Birkhoff's theorem that any sum of the form

$$S = \sum_{x,y} a_{x,y} S_{x,y}$$

can be rewritten as

$$S = \sum_{x,y} \sum_\pi \mu_\pi \delta_{x,\pi(y)} S_{x,y} = \sum_\pi \mu_\pi \sum_y S_{\pi(y),y} \ .$$

B.3 Typical sequences

Let $\mathbf{x}$ be an n-tuple chosen according to an n-fold product distribution $(P_X)^n$. Then, with probability almost one, $\mathbf{x}$ is a *typical sequence*, i.e., its frequency distribution $\lambda_\mathbf{x}$ is close to the distribution P_X.

Theorem B.3.1. *Let P_X be a probability distribution on $\mathcal{X}$ and let $\mathbf{x}$ be chosen according to the n-fold product distribution $(P_X)^n$. Then, for any $\delta \geq 0$,*

$$\Pr_\mathbf{x}\big[D(\lambda_\mathbf{x} \| P_X) > \delta\big] \leq 2^{-n(\delta - |\mathcal{X}|\frac{\log(n+1)}{n})} \ .$$

Proof. See Theorem 12.2.1 of [CT91]. □

Theorem B.3.1 quantifies the distance between $\lambda_\mathbf{x}$ and P_X with respect to the relative entropy. To obtain a statement in terms of the L_1-distance, we need the following lemma.

Lemma B.3.2. *Let P and Q be probability distributions. Then*

$$\|P - Q\|_1 \leq \sqrt{2(\ln 2) D(P\|Q)} \ .$$

Proof. See Lemma 12.6.1 of [CT91]. □

Corollary B.3.3. *Let P_X be a probability distribution on $\mathcal{X}$ and let $\mathbf{x}$ be chosen according to the n-fold product distribution $(P_X)^n$. Then, for any $\delta \geq 0$,*

$$\Pr_\mathbf{x}\big[\|\lambda_\mathbf{x} - P_X\|_1 > \delta\big] \leq 2^{-n(\frac{\delta^2}{2\ln 2} - |\mathcal{X}|\frac{\log(n+1)}{n})} \ .$$

Proof. The assertion follows directly from Theorem B.3.1 combined with Lemma B.3.2. □

---

[2] $\delta_{x,\pi(y)}$ denotes the Kronecker symbol which equals one if $x = \pi(y)$ and zero otherwise.

B.4 Product spaces

Lemma B.4.1. *Let* $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$*. Then*

$$\mathrm{supp}(\rho_{AB}) \subseteq \mathrm{supp}(\rho_A) \otimes \mathrm{supp}(\rho_B) \ .$$

Proof. Assume first that ρ_{AB} is pure, i.e., $\rho_{AB} = |\Psi\rangle\langle\Psi|$. Let $|\Psi\rangle = \sum_{z\in\mathcal{Z}} \alpha_z |\phi^z\rangle \otimes |\psi^z\rangle$ be a Schmidt decomposition of $|\Psi\rangle$, i.e., $\{|\phi^z\rangle\}_{z\in\mathcal{Z}}$ and $\{|\psi^z\rangle\}_{z\in\mathcal{Z}}$ are families of orthonormal vectors in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Then

$$\mathrm{supp}(\rho_{AB}) = \{|\Psi\rangle\} \subseteq \mathrm{span}\{|\phi^z\rangle\}_{z\in\mathcal{Z}} \otimes \mathrm{span}\{|\psi^z\rangle\}_{z\in\mathcal{Z}} \ .$$

Because $\mathrm{span}\{|\phi^z\rangle\}_{z\in\mathcal{Z}} = \mathrm{supp}(\rho_A)$ and $\mathrm{span}\{|\psi^z\rangle\}_{z\in\mathcal{Z}} = \mathrm{supp}(\rho_B)$ the assertion follows.

To show that the statement also holds for mixed states, let $\rho_{AB} = \sum_{x\in\mathcal{X}} \rho_{AB}^x$ be a decomposition of ρ_{AB} into pure states ρ_{AB}^x, for $x \in \mathcal{X}$. Then, because the lemma holds for the states ρ_{AB}^x,

$$\begin{aligned}
\mathrm{supp}(\rho_{AB}) &= \mathrm{span} \bigcup_{x\in\mathcal{X}} \mathrm{supp}(\rho_{AB}^x) \\
&\subseteq \mathrm{span} \bigcup_{x\in\mathcal{X}} \mathrm{supp}(\rho_A^x) \otimes \mathrm{supp}(\rho_B^x) \\
&\subseteq \Bigl(\mathrm{span} \bigcup_{x\in\mathcal{X}} \mathrm{supp}(\rho_A^x)\Bigr) \otimes \Bigl(\mathrm{span} \bigcup_{x\in\mathcal{X}} \mathrm{supp}(\rho_B^x)\Bigr) \\
&= \mathrm{supp}(\rho_A) \otimes \mathrm{supp}(\rho_B) \ .
\end{aligned}$$

□

Lemma B.4.2. *Let* $\rho_{AB}, \bar{\rho}_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *such that* $\mathrm{supp}(\bar{\rho}_{AB}) \subseteq \mathrm{supp}(\rho_{AB})$*. Then* $\mathrm{supp}(\bar{\rho}_A) \subseteq \mathrm{supp}(\rho_A)$*.*

Proof. Assume first that $\bar{\rho}_{AB}$ is pure, i.e., $\bar{\rho}_{AB} = |\Psi\rangle\langle\Psi|$. Let $|\Psi\rangle = \sum_{z\in\mathcal{Z}} \alpha_z |\phi^z\rangle \otimes |\psi^z\rangle$ be a Schmidt decomposition of $|\Psi\rangle$, i.e., $\{|\phi^z\rangle\}_{z\in\mathcal{Z}}$ and $\{|\psi^z\rangle\}_{z\in\mathcal{Z}}$ are families of orthonormal vectors in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Then $\mathrm{supp}(\bar{\rho}_{AB}) = \{|\Psi\rangle\}$. Moreover, by Lemma B.4.1,

$$\mathrm{supp}(\bar{\rho}_{AB}) \subseteq \mathrm{supp}(\rho_{AB}) \subseteq \mathrm{supp}(\rho_A) \otimes \mathrm{supp}(\rho_B) \ ,$$

i.e., $|\Psi\rangle \in \mathrm{supp}(\rho_A) \otimes \mathrm{supp}(\rho_B)$. This implies $|\phi^z\rangle \in \mathrm{supp}(\rho_A)$, for any $z \in \mathcal{Z}$, and thus $\mathrm{span}\{|\phi^z\rangle\}_{z\in\mathcal{Z}} \subseteq \mathrm{supp}(\rho_A)$. The assertion then follows because $\mathrm{span}\{|\phi^z\rangle\}_{z\in\mathcal{Z}} = \mathrm{supp}(\bar{\rho}_A)$.

To show that the statement holds for mixed states, let $\bar{\rho}_{AB} = \sum_{x\in\mathcal{X}} \bar{\rho}_{AB}^x$ be a decomposition of $\bar{\rho}_{AB}$ into pure states $\bar{\rho}_{AB}^x$, for $x \in \mathcal{X}$. We then have $\mathrm{supp}(\bar{\rho}_{AB}^x) \subseteq \mathrm{supp}(\rho_{AB})$, for any $x \in \mathcal{X}$, and thus, because the lemma holds for pure states, $\mathrm{supp}(\bar{\rho}_A^x) \subseteq \mathrm{supp}(\rho_A)$. Consequently,

$$\mathrm{supp}(\bar{\rho}_A) = \mathrm{span} \bigcup_{x\in\mathcal{X}} \mathrm{supp}(\bar{\rho}_A^x) \subseteq \mathrm{supp}(\rho_A) \ .$$

□

B.5 Nonnegative operators

Lemma B.5.1. *Let $\rho \in \mathcal{P}(\mathcal{H})$ and let S be a hermitian operator on $\mathcal{H}$. Then $S\rho S$ is nonnegative.*

Proof. Let $\rho = \sum_{x\in\mathcal{X}} p_x |x\rangle\langle x|$ be a spectral decomposition of ρ. Then, for any vector $|\theta\rangle \in \mathcal{H}$,

$$\langle\theta|S\rho S|\theta\rangle = \sum_{x\in\mathcal{X}} p_x \langle\theta|S|x\rangle\langle x|S|\theta\rangle = \sum_{x\in\mathcal{X}} p_x |\langle\theta|S|x\rangle|^2 \geq 0 .$$

The assertion then follows because $S\rho S$ is hermitian. □

Lemma B.5.2. *Let $\rho, \sigma \in \mathcal{P}(\mathcal{H})$. Then* $\mathrm{tr}(\rho\sigma) \geq 0$.

Proof. The assertion is an immediate consequence of the fact that $\mathrm{tr}(\rho\sigma) = \mathrm{tr}(\sigma^{1/2}\rho\sigma^{1/2})$ and Lemma B.5.1. □

Lemma B.5.3. *Let $\rho, \sigma \in \mathcal{P}(\mathcal{H})$ such that σ is invertible. Then the operator $\lambda \cdot \sigma - \rho$ is nonnegative if and only if*

$$\lambda_{\max}(\sigma^{-1/2}\rho\sigma^{-1/2}) \leq \lambda .$$

Proof. With $D := \lambda \cdot \mathrm{id} - \sigma^{-1/2}\rho\sigma^{-1/2}$, we have $\lambda \cdot \sigma - \rho = \sigma^{1/2} D \sigma^{1/2}$. Because of Lemma B.5.1, this operator is nonnegative if and only if D is nonnegative, which is equivalent to say that all eigenvalues of $\sigma^{-1/2}\rho\sigma^{-1/2}$ are upper bounded by λ. □

Lemma B.5.4. *Let $\rho, \sigma \in \mathcal{P}(\mathcal{H})$ such that $\lambda \cdot \sigma - \rho$ is nonnegative and σ is invertible. Then*

$$\lambda_{\max}(\rho^{1/2}\sigma^{-1}\rho^{1/2}) \leq \lambda .$$

Proof. Assume without loss of generality that ρ is invertible (otherwise, the statement follows by continuity). Because the operator $\lambda \cdot \sigma - \rho$ is nonnegative, the same holds for $\rho^{-1/2}(\lambda \cdot \sigma - \rho)\rho^{-1/2} = \lambda \cdot \rho^{-1/2}\sigma\rho^{-1/2} - \mathrm{id}$ (cf. Lemma B.5.1). Hence, all eigenvalues of $\rho^{-1/2}\sigma\rho^{-1/2}$ are at least λ^{-1}. Consequently, the eigenvalues of the inverse $\rho^{1/2}\sigma^{-1}\rho^{1/2}$ cannot be larger than λ. □

B.6 Properties of the function r_t

The class of functions $r_t : z \mapsto z^t - t \ln z - 1$, for $t \in \mathbb{R}$, is used in Section 3.3 for the proof of a Chernoff style bound. In the following, we list some of its properties.

Lemma B.6.1. *For any $t \in \mathbb{R}$, the function r_t is monotonically increasing on the interval $[1, \infty)$.*

Proof. The first derivative of r_t is given by

$$\frac{d}{dz} r_t(z) = t z^{t-1} - \frac{t}{z} = \frac{t}{z}(z^t - 1) \ .$$

The assertion follows because the term on the right hand side is nonnegative for any $z \in [1, \infty)$. □

Lemma B.6.2. *For any $t \in \mathbb{R}$ and $z \in (0, \infty)$,*

$$r_t(z) \leq r_{|t|}(z + \frac{1}{z}) \ .$$

Proof. Observe first that $r_t(z) = r_{-t}(\frac{1}{z})$. It thus suffices to show that the statement holds for $t \geq 0$. If $z \geq 1$, the assertion follows directly from Lemma B.6.1. For the case where $t \geq 0$ and $z < 1$, let $v := -t \ln z$. Then $r_t(\frac{1}{z}) = e^v - v - 1$ and $r_t(z) = e^{-v} + v - 1$. Because $v \geq 0$, we have $e^v - e^{-v} \geq 2v$, which implies $r_t(z) \leq r_t(\frac{1}{z})$. The assertion then follows again from Lemma B.6.1. □

Lemma B.6.3. *For any $t \in [-\frac{1}{2}, \frac{1}{2}]$, the function r_t is concave on the interval* $[4, \infty]$.

Proof. We show that $\frac{d^2}{dz^2} r_t(z) \leq 0$ for any $z \geq 4$. Because $\frac{d^2}{dz^2} r_t(z) = t(t-1)z^{t-2} + \frac{t}{z^2}$, this is equivalent to $t(1-t)z^t \geq t$. It thus suffices to verify that

$$z \geq \left(\frac{1}{1-t}\right)^{\frac{1}{t}} \ ,$$

for any $z \geq 4$. Using some simple analysis, it is easy to see that the term on the right hand side is monotonically increasing in t on the interval $[-\frac{1}{2}, \frac{1}{2}]$ and thus takes its maximum at $t = \frac{1}{2}$, in which case it equals 4. □

Lemma B.6.4. *For any $z \in [1, \infty)$ and $t \in [-\frac{1}{\log z}, \frac{1}{\log z}]$,*

$$r_t(z) \leq \bigl(1 - \ln 2\bigr)(\log z)^2 t^2 \ .$$

Proof. Let $v := t \ln z$. Then

$$\frac{r_t(z)}{t^2} = \frac{e^{t \ln z} - t \ln z - 1}{t^2} = \frac{e^v - v - 1}{v^2}(\ln z)^2 \ . \tag{B.1}$$

We first show that the term on the right hand side of (B.1) is monotonically increasing in v, that is,

$$\frac{d}{dv}\frac{e^v - v - 1}{v^2} = \frac{e^v - 1}{v^2} - 2\frac{e^v - v - 1}{v^3} \geq 0 \ .$$

A simple calculation shows that this inequality can be rewritten as

$$1 \geq \frac{2}{v}\frac{e^{v/2} - e^{-v/2}}{e^{v/2} + e^{-v/2}} \ ,$$

which holds because, for any $v \in \mathbb{R}$,

$$\left|\frac{e^{v/2} - e^{-v/2}}{e^{v/2} + e^{-v/2}}\right| = \left|\tanh \frac{v}{2}\right| \leq \frac{|v|}{2} \ .$$

Hence, in order to find an upper bound on (B.1), it is sufficient to evaluate the right hand side of (B.1) for the maximum value of v. By assumption, we have $v \leq \ln 2$, i.e.,

$$\frac{e^v - v - 1}{v^2}(\ln z)^2 \leq (1 - \ln 2)(\log z)^2 \ ,$$

which concludes the proof. □

Appendix C

Computationally Efficient Information Reconciliation

In Section 6.3, we have proposed a general one-way information reconciliation scheme which is optimal with respect to its information leakage. The scheme, however, requires the receiver of the error-correcting information to perform some decoding operation for which no efficient algorithm is known. In the following, we propose an alternative information reconciliation scheme based on error-correcting codes where all computations can be done efficiently.

C.1 Preliminaries

To describe and analyze the protocol, we need some terminology and basic results from the theory of channel coding. Let $\mathfrak{C}$ be a discrete memoryless channel which takes inputs from a set $\mathcal{U}$ and gives outputs from a set $\mathcal{V}$.[1] An *encoding scheme for* $\mathfrak{C}$ is a family of pairs $(\mathcal{C}_n, \mathrm{dec}_n)$ parameterized by $n \in \mathbb{N}$ where $\mathcal{C}_n$ is a *code on* $\mathcal{U}$ *of length* n, i.e., a set of n-tuples $\mathbf{u} \in \mathcal{U}^n$, called *codewords*, and dec_n is a *decoding function*, i.e., a mapping from $\mathcal{V}^n$ to $\mathcal{C}_n$. The *rate* of the code $\mathcal{C}_n$ is defined by $\mathrm{rate}(\mathcal{C}_n) := \frac{1}{n} \log |\mathcal{C}_n|$. Moreover, the *maximum error probability* of $(\mathcal{C}_n, \mathrm{dec}_n)$ is defined by

$$\varepsilon_{\max}(C_n, \mathrm{dec}_n) := \max_{\mathbf{u} \in \mathcal{C}_n} \Pr_{\mathbf{v}}[\mathbf{u} \neq \mathrm{dec}(\mathbf{v})] ,$$

where, for any $\mathbf{u} = (u_1, \ldots, u_n) \in \mathcal{C}_n$, the probability is over all outputs $\mathbf{v} = (v_1, \ldots, v_n)$ of n parallel invocations of $\mathfrak{C}$ on input $\mathbf{u}$.

We will use the following fundamental theorem for channel coding (cf., e.g., [CT91], Section 8.7).

[1] A *discrete memoryless channel* $\mathfrak{C}$ from $\mathcal{U}$ to $\mathcal{V}$ is defined by the conditional probability distributions $P_{V|U=u}$ on $\mathcal{V}$, for any $u \in \mathcal{U}$.

Proposition C.1.1. *Let $\mathfrak{C}$ be a discrete memoryless channel from $\mathcal{U}$ to $\mathcal{V}$ and let $\delta > 0$. Then there exists an encoding scheme $\{(\mathcal{C}_n, \mathrm{dec}_n)\}_{n\in\mathbb{N}}$ for $\mathfrak{C}$ such that the following holds:*

- $\mathrm{rate}(\mathcal{C}_n) \geq \max_{P_U} H(U) - H(U|V) - \delta$, *for any $n \in \mathbb{N}$. (The entropies in the maximum are computed for the distribution P_{UV} of an input/output pair (u, v) of $\mathfrak{C}$, where u is chosen according to P_U.)*
- $\lim_{n\to\infty} \varepsilon_{\max}(\mathcal{C}_n, \mathrm{dec}_n) = 0$.

C.2 Information reconciliation based on codes

Let us now consider an information reconciliation protocol based on channel coding. For this, we assume that Alice's and Bob's inputs are strings $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, respectively. Our protocol shall be secure if the inputs $\mathbf{x}, \mathbf{y}$ are distributed according to a product distribution $P_{X^nY^n} = (P_{XY})^n$.

Let $\mathfrak{C}$ be the channel which maps any $u \in \mathcal{X}$ to $v := (x \oplus u, y)$, where the pair (x, y) is chosen according to the probability distribution P_{XY} and where $\oplus$ is a group operation on $\mathcal{X}$. For any $n \in \mathbb{N}$, let $\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}$ be the information reconciliation protocol specified by Fig. C.1, where $\mathcal{C}_n$ is the code and dec_n the decoding function defined by Proposition C.1.1.

It is easy to see that $\hat{\mathbf{x}} = \mathbf{x}$ holds whenever dec_n decodes to the correct value $\hat{\mathbf{u}} = \mathbf{u}$. Hence, the information reconciliation protocol $\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}$ is ε_n-secure, for $\varepsilon_n := \varepsilon_{\max}(\mathcal{C}_n, \mathrm{dec}_n)$. Because, by Proposition C.1.1, the maximum error probability $\varepsilon_{\max}(\mathcal{C}_n, \mathrm{dec}_n)$ of $(\mathcal{C}_n, \mathrm{dec}_n)$ goes to zero, for n approaching infinity, the protocol $\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}$ is asymptotically secure.

Moreover, by Proposition C.1.1,

$$\mathrm{rate}(\mathcal{C}_n) \geq \max_{P_U} H(U) - H(U|X \oplus U, Y) - \delta \ .$$

Using the fact that the input u is chosen independently of the randomness of the channel (x, y), a simple information-theoretic computation shows that the entropy difference in the maximum can be rewritten as $H(X \oplus U|Y) - H(X|Y)$. Hence, because $\max_{P_U} H(X \oplus U|Y) = H_{\max}(P_U) = \log|\mathcal{X}|$, we find

$$\frac{1}{n}\log|\mathcal{C}_n| = \mathrm{rate}(\mathcal{C}_n) \geq \log|\mathcal{X}| - H(X|Y) - \delta \ . \tag{C.1}$$

The communication $\mathbf{c}$ of the protocol is contained in the set $\mathcal{X}^n$. Furthermore, because $\mathbf{u}$ is chosen uniformly at random from $\mathcal{C}_n$, the distribution $P_{\mathbf{C}|X^n=\mathbf{x}}$ of the communication $\mathbf{c}$, conditioned on any input $\mathbf{x} \in \mathcal{X}^n$, is uniform over a set of size $|\mathcal{C}_n|$. The leakage of $\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}$ is thus given by

$$\mathrm{leak}_{\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}} = \log|\mathcal{X}^n| - \min_{\mathbf{x}} H_{\min}(P_{\mathbf{C}|X^n=\mathbf{x}}) = n\log|\mathcal{X}| - \log|\mathcal{C}_n| \ .$$

Fig. C.1 Information reconciliation protocol $\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}$.

Parameters:

$\mathcal{C}_n$:	set of codewords from $\mathcal{X}^n$
dec_n:	decoding function from $\mathcal{X}^n \times \mathcal{Y}^n$ to $\mathcal{C}_n$
$\oplus$:	group operation on $\mathcal{X}$ (with inverse $\ominus$).

Alice		**Bob**
input: $\mathbf{x} \in \mathcal{X}^n$		input: $\mathbf{y} \in \mathcal{Y}^n$
$\mathbf{u} \in_R \mathcal{C}_n$ $\mathbf{c} := \mathbf{x} \oplus \mathbf{u}$	$\xrightarrow{\mathbf{c}}$	$\hat{\mathbf{u}} := \mathrm{dec}_n(\mathbf{c}, \mathbf{y})$
		if decoding not succ. then abort
		output $\hat{\mathbf{x}} := \mathbf{c} \ominus \hat{\mathbf{u}}$

Combining this with (C.1) we conclude

$$\frac{1}{n} \mathrm{leak}_{\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}} \leq H(X|Y) + \delta \ .$$

Because Proposition C.1.1 also holds for efficient[2] encoding schemes (see, e.g., [Dum98]), Corollary 6.3.5 is asymptotically still true if we restrict to computationally efficient protocols (see also [HR05]). More precisely, this result can be formulated as follows.

Proposition C.2.1. *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a probability distribution and let $\delta > 0$. Then there exists a family of computationally efficient information reconciliation protocols $\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}$ (parameterized by $n \in \mathbb{N}$) which are ε_n-fully secure, ε_n-robust on the product distribution $(P_{XY})^n$, and have leakage $\frac{1}{n}\mathrm{leak}_{\mathsf{IR}_{\mathcal{C}_n,\mathrm{dec}_n}} \leq H(X|Y) + \delta$, for any $n \in \mathbb{N}$, where $\lim_{n\to\infty} \varepsilon_n = 0$.*

[2] An encoding scheme $\{(\mathcal{C}_n, \mathrm{dec}_n)\}_{n\in\mathbb{N}}$ is said to be *efficient* if there exist polynomial-time algorithms (in n) for sampling a codeword from the set $\mathcal{C}_n$ and for evaluating the decoding function dec_n.

Appendix D

Notation

General

log	binary logarithm
ln	natural logarithm
$\delta_{x,y}$	Kronecker symbol: $\delta_{x,y} \in \{0,1\}$, $\delta_{x,y} = 1$ iff $x = y$
$\overline{c}$	complex conjugate of c
$\Re(c)$	real value of c
$\mathcal{P}(\mathcal{X})$	set of nonnegative functions on the set $\mathcal{X}$
$\mathcal{S}_n$	set of permutations on the set $\{1, \ldots, n\}$
$\mathbb{E}_x[f(x)]$	expectation of $f(x)$ over random choices of x
$\mathrm{supp}(f)$	support of the function f
$[a, b]$	set of real numbers r such that $a \leq r \leq b$
$[a, b)$	set of real numbers r such that $a \leq r < b$

Frequency distributions and types

$\lambda_{\mathbf{x}}$	frequency distribution of the n-tuple $\mathbf{x}$
$\mathcal{Q}_n^{\mathcal{X}}$	set of types with denominator n on the set $\mathcal{X}$
Λ_n^Q	type class of the type Q with denominator n

Vectors

$\mathrm{span}\,\mathcal{V}$	space spanned by the set of vectors $\mathcal{V}$
$\langle\phi\|\psi\rangle$	scalar product of the vectors $\|\phi\rangle$ and $\|\psi\rangle$
$\\|\|\phi\rangle\\|$	norm of the vector $\|\phi\rangle$
$\|\phi\rangle\langle\phi\|$	projector onto the vector $\|\phi\rangle$
$\mathcal{S}_1(\mathcal{H})$	set of normalized vectors on $\mathcal{H}$

Operators

$\mathcal{P}(\mathcal{H})$	set of nonnegative operators on $\mathcal{H}$
id	identity
$\mathrm{tr}(S)$	trace of the hermitian operator S
$\mathrm{supp}(S)$	support of the hermitian operator S
$\mathrm{rank}(S)$	rank of the hermitian operator S
$\lambda_{\max}(S)$	maximum eigenvalue of the hermitian operator S
$\|S\|_1$	trace norm of the hermitian operator S

Distance measures for operators

$\|\rho - \rho'\|_1$	L_1-distance between ρ and ρ'
$F(\rho, \rho')$	fidelity between ρ and ρ'.
$d(\rho_{AB}\|B)$	L_1-distance from uniform of ρ_{AB} given B
$d_2(\rho_{AB}\|\sigma_B)$	L_2-distance from uniform of ρ_{AB} relative to σ_B

Entropies

$H(P_X)$	Shannon entropy of the probability distribution P_X
$h(p)$	binary Shannon entropy with bias p
$H(\rho_A)$	von Neumann entropy of the density operator ρ_A
$H(A\|B)$	conditional entropy $H(\rho_{AB}) - H(\rho_B)$
$D(\rho\\|\sigma)$	relative entropy of ρ to σ
$H_{\min}(\rho_{AB}\|\sigma_B)$	min-entropy of ρ_{AB} relative to σ_B
$H_{\max}(\rho_{AB}\|\sigma_B)$	max-entropy of ρ_{AB} relative to σ_B
$H^\varepsilon_{\min}(\rho_{AB}\|\sigma_B)$	ε-smooth min-entropy of ρ_{AB} relative to σ_B
$H^\varepsilon_{\max}(\rho_{AB}\|\sigma_B)$	ε-smooth max-entropy of ρ_{AB} relative to σ_B
$H^\varepsilon_{\min}(\rho_{AB}\|B)$	ε-smooth min-entropy of ρ_{AB} given $\mathcal{H}_B$
$H^\varepsilon_{\max}(\rho_{AB}\|B)$	ε-smooth max-entropy of ρ_{AB} given $\mathcal{H}_B$
$H^\varepsilon_{\min}(A\|B)$	abbreviation for $H^\varepsilon_{\min}(\rho_{AB}\|B)$
$H^\varepsilon_{\max}(A\|B)$	abbreviation for $H^\varepsilon_{\max}(\rho_{AB}\|B)$
$H_2(\rho_{AB}\|\sigma_B)$	collision entropy of ρ_{AB} relative to σ_B

Symmetric spaces

$\mathrm{Sym}(\mathcal{H}^{\otimes n})$	Symmetric subspace of $\mathcal{H}^{\otimes n}$
$\mathrm{Sym}(\mathcal{H}^{\otimes n}, \|\theta\rangle^{\otimes m})$	Symmetric subspace of $\mathcal{H}^{\otimes n}$ along $\|\theta\rangle^{\otimes m}$

Bibliography

[ABB+04] A. Acín, J. Bae, E. Bagan, M. Baig, Ll. Masanes, and R. Muñoz-Tapia. Secrecy content of two-qubit states. `http://arxiv.org/abs/quant-ph/0411092`, November 2004.

[Bar05] A. Bariska. On a weakness of common security definitions for QKD protocols. Master's thesis, ETH Zürich, March 2005. See also `http://arxiv.org/abs/quant-ph/0512021`.

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: Public-key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, 1984.

[BBB+02] E. Biham, M. Boyer, G. Brassard, J. van de Graaf, and T. Mor. Security of quantum key distribution against all collective attacks. *Algorithmica*, 34:372–388, 2002.

[BBB+05] E. Biham, M. Boyer, P. O. Boykin, T. Mor, and V. Roychowdhury. A proof of the security of quantum key distribution. *Journal of Cryptology*, 2005. to appear.

[BBCM95] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer. Generalized privacy amplification. *IEEE Transaction on Information Theory*, 41(6):1915–1923, 1995.

[BBM92] C. H. Bennett, G. Brassard, and N. D. Mermin. Quantum cryptography without Bell's theorem. *Phys. Rev. Lett.*, 68:557–559, 1992.

[BBP+96] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacherand J. Smolin, and W. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–726, 1996.

[BBR88] C. H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, 1988.

[Ben92] C. H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68(21):3121–3124, 1992.

[Bir46] G. Birkhoff. Three observations on linear algebra. *Univ. Nac. Tucumán. Rev. Ser. A*, 5:147–151, 1946.

[BM97a] E. Biham and T. Mor. Bounds on information and the security of quantum cryptography,. *Phys. Rev. Lett.*, 79:4034–4037, 1997.

[BM97b] E. Biham and T. Mor. Security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 78(11):2256–2259, 1997.

[BMS96] C. H. Bennett, T. Mor, and J. A. Smolin. The parity bit in quantum cryptography. *Phys. Rev. A*, 54:2675–2684, 1996.

[BOHL+05] M. Ben-Or, M. Horodecki, D. W. Leung, D. Mayers, and J. Oppenheim. The universal composable security of quantum key distribution. In *Second Theory of Cryptography Conference TCC*, volume 3378 of *Lecture Notes in Computer Science*, pages 386–406. Springer, 2005. Also available at `http://arxiv.org/abs/quant-ph/0409078`.

[BOM04] M. Ben-Or and D. Mayers. General security definition and composability for quantum and classical protocols. `http://arxiv.org/abs/quant-ph/0409062`, September 2004.

[BPG99] H. Bechmann-Pasquinucci and N. Gisin. Incoherent and coherent eavesdropping in the six-state protocol of quantum cryptography. *Phys. Rev. A,*, 59:4238, 1999.

[Bru98] D. Bruss. Optimal eavesdropping in quantum cryptography with six states. *Phys. Rev. Lett.*, 81:3018, 1998.

[BS94] G. Brassard and L. Salvail. Secret-key reconciliation by public discussion. In *Advances of Cryptology — EUROCRYPT '93*, Lecture Notes in Computer Science, pages 410–423. Springer, 1994.

[Can01] R. Canetti. Universally composable security: A new paradigm for cryptographic protocols. In *Proc. 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 136–145, 2001.

[CFS02] C. M. Caves, C. A. Fuchs, and R. Schack. Unknown quantum states: The quantum de Finetti representation. *Journal of Mathematical Physics*, page 4537, 2002.

[Cha02] H. F. Chau. Practical scheme to share a secret key through a quantum channel with a 27.6% bit error rate. *Phys. Rev. A*, 66:060302, 2002.

[CRE04] M. Christandl, R. Renner, and A. Ekert. A generic security proof for quantum key distribution. `http://arxiv.org/abs/quant-ph/0402131`, February 2004.

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. Wiley, New York, 1991.

[CW79] J. L. Carter and M. N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18:143–154, 1979.

[DFSS05] I. Damgaard, S. Fehr, L. Salvail, and C. Schaffner. Cryptography in the bounded quantum-storage model. In *46th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 449–458, 2005.

[DHL$^+$04] D. P. DiVincenzo, M. Horodecki, D. W. Leung, J. A. Smolin, and B. M. Terhal. Locking classical correlation in quantum states. *Phys. Rev. Lett.*, 92:067902, 2004.

[DM04] S. Dziembowski and U. Maurer. Optimal randomizer efficiency in the bounded-storage model. *Journal of Cryptology*, 17(1):5–26, 2004. Conference version appeared in Proc. of STOC '02.

[Dum98] Ilya I. Dumer. Concatenated codes and their multilevel generalizations. In Vera Saine Pless and W. C. Huffman, editors, *The Handbook of Coding Theory*, volume 2, chapter 23, pages 1191–1988. North-Holland, Elsevier, 1998.

[DW05] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. R. Soc. Lond. A*, 461:207–235, 2005.

[Eke91] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661, 1991.

[GL03] D. Gottesman and H.-K. Lo. Proof of security of quantum key distribution with two-way classical communications. *IEEE Transactions on Information Theory*, 49(2):457–475, 2003.

[GN93] P. Gemmell and N. Naor. Codes for interactive authentication. In *Advances in Cryptology — CRYPTO '93*, volume 773 of *Lecture Notes in Computer Science*, pages 355–367. Springer, 1993.

[Gro05] F. Grosshans. Collective attacks and unconditional security in continuous variable quantum key distribution. *Phys. Rev. Lett*, 94:020504, 2005.

[HJ85] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, 1985.

[HLSW04] P. Hayden, D. Leung, P. W. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics*, 250(2):371–391, 2004.

[HM76] R. L. Hudson and G. R. Moody. Locally normal symmetric states and an analogue of de Finetti's theorem. *Z. Wahrschein. verw. Geb.*, 33:343–351, 1976.

[HR05] T. Holenstein and R. Renner. One-way secret-key agreement and applications to circuit polarization and immunization of public-key encryption. In *Advances in Cryptology — CRYPTO '05*, Lecture Notes in Computer Science, pages 478–493. Springer, 2005.

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions (extended abstract). In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, pages 12–24, 1989.

[ILM01] H. Inamori, N. Lütkenhaus, and D. Mayers. Unconditional security for practial quantum key distribution. `http://arxiv.org/abs/quant-ph/0107017`, July 2001.

[KGR05] B. Kraus, N. Gisin, and R. Renner. Lower and upper bounds on the secret key rate for QKD protocols using one-way classical communication. *Phys. Rev. Lett.*, 95:080501, 2005.

[KMR05] R. König, U. Maurer, and R. Renner. On the power of quantum memory. *IEEE Transactions on Information Theory*, 51(7):2391–2401, 2005. A preliminary version appeared 2003 on the eprint archive `http://arxiv.org/abs/quant-ph/0305154`.

[KR05] R. König and R. Renner. A de Finetti representation for finite symmetric quantum states. *Journal of Mathematical Physics*, 46:122108, 2005.

[LC99] H.-K. Lo and H. F. Chau. Unconditional security of quantum key distribution over arbitrarily long distances. *Science*, 283:2050–2056, 1999.

[LCA05] H.-K. Lo, H. F. Chau, and M. Ardehali. Efficient quantum key distribution scheme and a proof of its unconditional security. *Journal of Cryptology*, 18(2):133–165, 2005.

[Lo00] H.-K. Lo. Proof of unconditional security of six-state quantum key distribution scheme. *Quantum Information and Computation*, 1(2):81, 2000.

[Lo05] H.-K. Lo. Getting something out of nothing. `http://arxiv.org/abs/quant-ph/0503004`, March 2005.

[Mau93] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993.

[May96] D. Mayers. Quantum key distribution and string oblivious transfer in noisy channels. In *Advances in Cryptology — CRYPTO '96*, volume 1109 of *Lecture Notes in Computer Science*, pages 343–357. Springer, 1996.

[May01] D. Mayers. Unconditional security in quantum cryptography. *Journal of the ACM*, 48(3):351–406, 2001.

[MC93] P. Monari and D. Cocchi, editors. *Introduction to Bruno de Finetti's "Probabiliá e Induzione"*. Cooperativa Libraria Universitaria Editrice, Bologna, 1993.

[MRH04] U. Maurer, R. Renner, and C. Holenstein. Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology. In *First Theory of Cryptography Conference TCC*, volume 2951 of *Lecture Notes in Computer Science*, pages 21–39. Springer, 2004.

[NA05] M. Navascués and A. Acín. Security bounds for continuous variables quantum key distribution. *Phys. Rev. Lett.*, 94:020505, 2005.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

[PW00] B. Pfitzmann and M. Waidner. Composition and integrity preservation of secure reactive systems. In *7th ACM Conference on Computer and Communications Security*, pages 245–254. ACM Press, 2000.

[RGK05] R. Renner, N. Gisin, and B. Kraus. Information-theoretic security proof for quantum key distribution protocols. *Phys. Rev. A*, 72:012332, 2005.

[RK05] R. Renner and R. König. Universally composable privacy amplification against quantum adversaries. In *Second Theory of Cryptography Conference TCC*, volume 3378 of *Lecture Notes in Computer Science*, pages 407–425. Springer, 2005. Also available at `http://arxiv.org/abs/quant-ph/0403133`.

[RW03] R. Renner and S. Wolf. Unconditional authenticity and privacy from an arbitrarily weak secret. In *Advances in Cryptology — CRYPTO '03*, Lecture Notes in Computer Science, pages 78–95. Springer, 2003.

[RW04] R. Renner and S. Wolf. The exact price for unconditionally secure asymmetric cryptography. In *Advances in Cryptology — EUROCRYPT '04*, Lecture Notes in Computer Science, pages 109–125. Springer, 2004.

[SARG04] V. Scarani, A. Acin, G. Ribordy, and N. Gisin. Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations. *Phys. Rev. Lett.*, 92:057901, 2004.

[Sha49] Claude E. Shannon. Communication theory of secrecy systems. *Bell Systems Technical Journal*, 28:656–715, 1949.

[SP00] P. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85:441, 2000.

[Sti91] D. R. Stinson. Universal hashing and authentication codes. In *Advances in Cryptology — CRYPTO '91*, volume 576 of *Lecture Notes in Computer Science*, pages 74–85, 1991.

[TKI03] K. Tamaki, M. Koashi, and N. Imoto. Unconditionally secure key distribution based on two nonorthogonal states. *Phys. Rev. Lett.*, 90:167904, 2003.

[Unr04] D. Unruh. Simulatable security for quantum protocols. `http://arxiv.org/abs/quant-ph/0409125`, September 2004.

[Ver26] G. S. Vernam. Cipher printing telegraph systems for secret wire and radio telegraphic communications. *J. Am. Inst. Elec. Eng.*, 55:109–115, 1926.

[WC81] M. N. Wegman and J. L. Carter. New hash functions and their use in authentication and set equality. *Journal of Computer and System Sciences*, 22:265–279, 1981.

[WG00] N. R. Wallach and R. Goodman. *Representations and Invariants of the Classical Groups*. Cambridge University Press, 2000.

[Wie83] S. Wiesner. Conjugate coding. *Sigact News*, 15(1):78–88, 1983.

[Win99] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7), 1999.

Index

Curriculum Vitae

2001–2005	Teaching and research assistant, *Institute of Theoretical Computer Science, Eidgenössische Technische Hochschule (ETH) Zürich*, supervisor: Prof. U. Maurer.
1997–2000	Graduate in theoretical physics (Diplom in Physik), *ETH Zürich*.
1996	Military Service (Officers school of the Swiss Army Signal Corps).
1995	Undergraduate in physics (Premier Propédeutique), *École Polytechnique Fédérale (EPF) de Lausanne*.
1994	Realgymnasium (Matura Typus C), *Kantonsschule Alpenquai Luzern*.