Université de Montréal

An Exploration of Approximation Chains

par

Ashutosh Satyajit Marwah

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Informatique

November 3, 2024

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée

An Exploration of Approximation Chains

présentée par

Ashutosh Satyajit Marwah

a été évaluée par un jury composé des personnes suivantes :

Louis Salvail
(président-rapporteur)

Frédéric Dupont-Dupuis
(directeur de recherche)

Gilles Brassard
(membre du jury)

Graeme Smith
(examinateur externe)

Louis Salvail
(représentant du doyen de la FESP)

Résumé

La théorie de l'information à coup unique vise à étudier les tâches de communication et de traitement de l'information pour des états et des processus généraux avec une structure minimale. Une telle généralité est cruciale pour analyser les tâches de communication avec des ressources limitées et la sécurité des protocoles cryptographiques. Dans le régime asymptotique pour les tâches d'information avec une structure i.i.d. (indépendante et identiquement distribuée), les taux sont typiquement caractérisés par l'entropie de von Neumann et ses dérivées. Dans le régime à coup unique, une multitude d'entropies différentes sont nécessaires à cette fin. L'une des plus importantes est la min-entropie lisse, qui caractérise les taux des protocoles cryptographiques. Contrairement à l'entropie de von Neumann, le comportement de la min-entropie lisse est souvent contre-intuitif. Les outils de décomposition de la min-entropie lisse sont également assez restrictifs, rendant difficile l'analyse des structures qui émergent naturellement en théorie de l'information.

Une telle structure, que nous appelons une chaîne d'approximation, constitue le thème central de cette thèse. Pour un état $\rho_{A_1^n B}$, nous appelons une séquence d'états $(\sigma^{(k)}_{A_1^k B})_{k=1}^n$ une chaîne d'approximation de ρ si pour chaque k, $\rho_{A_1^k B} \approx_\epsilon \sigma^{(k)}_{A_1^k B}$. Ces structures émergent fréquemment lors de l'incorporation d'approximations dans les identités entropiques, l'étude des imperfections et le développement de preuves de sécurité. Alors que l'entropie de von Neumann de ρ peut être facilement exprimée en termes d'entropies des états de sa chaîne d'approximation, il n'est généralement pas possible de le faire avec la min-entropie lisse.

Dans cette thèse, nous développons des techniques pour établir des bornes entropiques avec des chaînes d'approximation et les appliquons à des scénarios cryptographiques. Notre travail commence par considérer l'un des cas les plus simples d'une telle chaîne, où les registres de ρ sont presque indépendants les uns des autres, et culmine avec l'établissement d'une règle de chaînage universelle pour la min-entropie lisse, qui permet de borner celle-ci en termes des entropies des états d'une chaîne d'approximation.

De plus, nous prouvons deux versions approximatives du théorème d'accumulation d'entropie (EAT), qui est un outil important pour borner la min-entropie lisse d'un état produit par un processus séquentiel. La première utilise des approximations des canaux utilisés dans le processus EAT, tandis que la seconde, appelée EAT approximatif non-structuré, relâche significativement la structure séquentielle requise sur l'état.

Nous mettons en valeur ces outils en les utilisant pour résoudre deux problèmes cryptographiques importants. Tout d'abord, nous prouvons la sécurité de la distribution quantique de clés (QKD) avec des corrélations à la source, qui sont des corrélations indésirables entre les rounds du protocole survenant en raison des imperfections de la source. Ces corrélations ont été un défi persistant pour la QKD. Nous fournissons une méthode simple et générale pour réduire la sécurité d'un protocole QKD avec ces corrélations à un protocole sans ces dernières.

Notre deuxième application majeure est la preuve de la sécurité de la distribution quantique de clés *device-independent* (DIQKD) parallèle. En adaptant les techniques de répétition parallèle des jeux non-locaux, nous construisons une chaîne d'approximation structurée pour la sortie du protocole. L'application du EAT approximatif non-structuré à cette chaîne fournit alors une preuve de sécurité pour le protocole.

Mots-clés : Information quantique, théorie de l'information, cryptographie, distribution quantique de clés.

Abstract

One-shot information theory aims to study communication and information processing tasks for general states and processes under minimal structure. Such generality is crucial for analysing communication tasks with limited resources and the security of cryptographic protocols. In the asymptotic regime for information tasks with an i.i.d. (independent and identically distributed) structure, the rates are typically characterised by the von Neumann entropy and its derivatives. In the one-shot regime, a multitude of different entropies are required for this purpose. One of the most important among these is the smooth min-entropy, which characterises the rates of cryptographic protocols. In contrast to the von Neumann entropy, the behaviour of the smooth min-entropy is often unintuitive. Tools for decomposing the smooth min-entropy are also quite restrictive, making it challenging to analyse structures that naturally arise in information theory.

One such structure, which we call an *approximation chain*, forms the central theme of this thesis. For a state $\rho_{A_1^n B}$, we term a sequence of states $(\sigma^{(k)}_{A_1^k B})_{k=1}^n$ an approximation chain of ρ if for each k, $\rho_{A_1^k B} \approx_\epsilon \sigma^{(k)}_{A_1^k B}$. These structures frequently emerge when incorporating approximations in entropic identities, studying imperfections, and developing security proofs. While the von Neumann entropy of ρ can be readily expressed in terms of the entropies of its approximation chain states, it is generally not possible to do so with the smooth min-entropy.

In this thesis, we develop techniques for establishing entropic bounds with approximation chains and apply these to cryptographic scenarios. Our work begins by considering one of the simplest cases of such a chain, whereby the registers of ρ are almost independent of one another, and culminates in establishing a *universal chain rule* for the smooth min-entropy, which enables one to bound the smooth min-entropy for a state in terms of entropies of its approximation chain.

Furthermore, we prove two approximate versions of the entropy accumulation theorem (EAT), which is an important tool for bounding the smooth min-entropy of a state produced

by a sequential process. The first enables approximations to the channels used in the EAT process, while the second, termed the *unstructured* approximate EAT, significantly relaxes the sequential structure required on the state.

We showcase these tools by using them to solve two significant cryptographic problems. First, we prove the security of quantum key distribution (QKD) under *source correlations*, which are undesired correlations among states across independent QKD rounds arising due to source imperfections. These correlations have been a persistent challenge for QKD. We provide a simple and general method to reduce the security for a QKD protocol with these correlations to one without.

Our second major application is proving the security of parallel device-independent quantum key distribution (DIQKD). By adapting techniques from the parallel repetition of non-local games, we construct a structured approximation chain for the protocol output. Applying the unstructured approximate EAT to this chain then yields a proof of security for the protocol.

Keywords: Quantum information, information theory, cryptography, quantum key distribution.

Contents

List of tables

List of figures

List of Acronyms and Abbreviations

i.i.d.	independent and identically distributed
POVM	Positive operator valued measure
QKD	Quantum key distribution
DIQKD	Device-independent quantum key distribution
IR	Information reconciliation
AEP	Asymptotic equipartition property
EAT	Entropy accumulation theorem
GEAT	Generalised entropy accumulation theorem
GT inequality	Golden-Thompson inequality

Acknowledgements

I have been extremely lucky to spend my time the past few years thinking and working on challenging problems. I would like to begin by thanking my supervisor, Frédéric Dupuis, for his invaluable guidance. He presented me with a question that turned out to be far deeper than we initially anticipated, ultimately blossoming into the contents of this thesis. His insights were instrumental in solving and refining numerous challenges we encountered along the way. I could depend on him for help not just for research but also for administrative questions and French translations.

I am deeply grateful to Norbert Lütkenhaus and his group, particularly Ernest Tan, Amir Arqand, and Shlok Ashok Nahar, for hosting me and including me in their Slack channel. Discussions with them significantly enriched my research. Ernest and Amir's observations helped improve Theorem 3.12. Additionally, Ernest and Shlok introduced me to the source correlation problem and explained the requirements for a practical solution– all of which greatly inform this work.

I would also like to thank Omar Fawzi for hosting me and for multiple interesting discussions. In particular, Omar pointed out Lemma A.2 to me. I am also indebted to user:fedja on MathOverflow, who provided a proof for Lemma 5.13– a challenge that had eluded us for a long time– and graciously permitted its reproduction. I also thank Mohammed Barhoush and Louis Salvail for interesting discussions.

Finally, I extend my gratitude to Louis Salvail, Gilles Brassard and Graeme Smith for agreeing to be a part of my defence committee. Their time and effort is greatly appreciated.

This work was supported by scholarships and funding sponsored by Google, the J.A. DeSève Foundation, and the Natural Sciences and Engineering Research Council of Canada.

Chapter 1

Introduction

Information theory studies the transmission and processing of information. The field stemmed from Claude Shannon's seminal paper [Sha48], which aimed to identify the fundamental limits of compression and the rate of information transmission. The advent of quantum computation and information necessitated the development of quantum information theory, extending this field into the quantum realm. To establish a foundation in quantum information theory, researchers began by first understanding the rates of communication and compression, similar to Shannon before them. To determine the fundamental rates for these problems, it was natural to use the asymptotic limit, which involves studying the problem under the limit of a large number of channel uses or input samples. This approach allows one to accurately calculate the rates, without getting encumbered by the complexity which arises from the higher-order terms. In this asymptotic regime, these rates are characterised by the von Neumann entropy and its derivatives [Sch95, SW97, Hol98, BSST99, Wil13]. For a state ρ_A, the von Neumann entropy is defined as[(1)]

$$H(A)_\rho \coloneqq -\operatorname{tr}(\rho_A \log \rho_A) \tag{1.1}$$

and the von Neumann conditional entropy of register A with respect to register B for a state ρ_{AB} is defined as

$$H(A|B)_\rho \coloneqq H(AB)_\rho - H(B)_\rho. \tag{1.2}$$

While results in the asymptotic limit provide valuable insights, they assume an infinite number of resources, which is not physically realizable in real-world systems. For practical applications, one must move beyond asymptotic analysis and account for finite-size effects. Furthermore, these analyses rely on the problem having some sort of independent and identically distributed (i.i.d.) structure. For instance, in source coding, which studies the compression of information produced by a quantum source, the states are typically assumed

[(1)] We use base e for exp and log, and *nat* units for entropies throughout this thesis.

to be i.i.d. In channel coding, where one analyses the limits of information transmission using noisy quantum channels, the channels are considered to be independent and identical. This assumption greatly simplifies these problems. However, it is not universally applicable in practical scenarios. As an example, consider an eavesdropper trying to listen in during a secret distribution protocol, which requires multiple rounds. There is no reason to believe a priori that the eavesdropper's actions in the n^{th} round of such a protocol would be independent of the previous rounds.

Therefore, recent efforts have focused on developing a *one-shot* theory of quantum information [**Ren06**, **Tom12**], which accounts for finite-size effects and goes beyond the assumption of the i.i.d. setting. One-shot information theory, in its most general form, aims to characterise information-theoretic tasks without any assumptions on the structure of the processes or states involved. This approach allows for tight characterisations of general scenarios, which can then be refined by incorporating additional structure specific to the problem at hand, thus enabling one to incorporate finite-size effects. Moreover, it makes the study of communication and cryptographic protocols more modular and helps to reveal the minimal conditions necessary for an application to work.

This generalized approach is crucial for unconditional or information theoretically secure cryptography, where one cannot place any kind of assumption on the actions of the adversary (see, for example, [**TLGR12**,**KWW12**]). Quantum key distribution is one of the most prominent examples of such a cryptographic protocol. Key distribution is a primitive in cryptography, which allows two parties to establish a *secret key* for secure communication even in the presence of an adversary. While classical key distribution protocols cannot achieve unconditional security, it is possible to use properties of quantum particles like the no-cloning principle to create unconditionally secure *quantum key distribution (QKD)* protocols [**BB84**, **Ben92**, **SP00**].

The quest to prove security for QKD has been a significant driver of research in one-shot information theory. The primary strategy for establishing the security of QKD relies on an information-theoretic tool called *privacy amplification*, which allows for the extraction of a completely random key from a string that might be partially correlated with an adversary's information. This setting can be analysed using one-shot information theory. It can be shown that the length of the random key, which can be extracted in such a setting is determined by an entropy measure called the smooth min-entropy, $H_{\min}^{\epsilon}(A|B)_{\rho}$ [**TRSS10**]. By demonstrating that this entropy for the string held by the honest parties at the end of

the QKD protocol is sufficiently large, one can use privacy amplification to derive a key that is almost independent of the adversary. Privacy amplification is used similarly in other cryptographic protocols as well. For this reason, the smooth min-entropy is one of the most important entropies in one-shot information theory.

Formally for a state ρ_{AB}, the min-entropy of register A conditioned on register B is defined as

$$H_{\min}(A|B)_\rho := \sup\left\{\lambda \in \mathbb{R} : \text{there exists a state } \sigma_B \text{ such that } \rho_{AB} \leq e^{-\lambda}\, \mathbb{1}_A \otimes \sigma_B\right\} \tag{1.3}$$

and the smooth min-entropy is defined as

$$H^\epsilon_{\min}(A|B)_\rho := \sup_{\tilde{\rho}} H_{\min}(A|B)_{\tilde{\rho}} \tag{1.4}$$

where the supremum is over all subnormalised states $\tilde{\rho}_{AB}$ which are ϵ-close to the state ρ in the purified distance. For a classical-quantum state, $\rho_{AB} = \sum_a p(a) \left|a\right\rangle\left\langle a\right| \otimes \rho_B^{(a)}$, the min-entropy simplifies to the negative log of the probability with which one can guess register A given access only to register B [**KRS09**].

The smooth min-entropy behaves very differently from the von Neumann conditional entropy, which characterises privacy amplification in the i.i.d. setting. In particular, the difference between these two can be very large. For example, consider the probability distribution $p_{A_1^n B}$ where $B \in_R \{0,1\}$ is sampled randomly and A_1^n[(2)] are sampled uniformly at random from n-bit strings if $B = 1$ and otherwise set to the all zero string if $B = 0$. For this probability distribution, we have

$$H^\epsilon_{\min}(A_1^n|B)_p \leq \log(2) \quad \text{and} \quad H(A_1^n|B)_p = \frac{n}{2}\log(2). \tag{1.5}$$

Roughly speaking, the smooth min-entropy places a much higher weight on the worst possible scenario of the conditioning register, whereas the von Neumann entropy places an equal weight on all possible scenarios.

Thus, the von Neumann entropy generally follows the behaviour one intuitively expects from a measure of uncertainty. Central to the work presented here, is the intuitive expectation that the total uncertainty of two registers A_1 and A_2 given a register B should be the sum of the uncertainty of register A_1 given B and the uncertainty of register A_2, conditioned on knowing both A_1 and B. The von Neumann entropy, respects this intuition nicely and

(2)The notation A_1^n denotes the set of registers $A_1, A_2, \cdots, A_n$.

satisfies the following chain rule:

$$H(A_1A_2|B)_\rho = H(A_1|B)_\rho + H(A_2|A_1B)_\rho. \tag{1.6}$$

More generally, we can decompose the von Neumann entropy of a large system A_1^n given B into a sum of the entropies of its parts as:

$$H(A_1^n|B)_\rho = \sum_{k=1}^{n} H(A_k|A_1^{k-1}B)_\rho. \tag{1.7}$$

The behaviour of the smooth min-entropy, on the other hand, deviates from these intuitive expectations. This is evident in the chain rule for smooth min-entropy [**DBWR14**, **VDTR13**], which is notably more complex than Eq. 1.6:

$$H_{\min}^{\epsilon_1+2\epsilon_2+\delta}(A_1A_2|B)_\rho \geq H_{\min}^{\epsilon_1}(A_1|B)_\rho + H_{\min}^{\epsilon_2}(A_2|A_1B)_\rho - k(\delta) \tag{1.8}$$

where $k(\delta) = O\left(\log \frac{1}{\delta}\right)$. Moreover, it is easy to demonstrate that a smooth min-entropy counterpart for the relation in Eq. 1.7 cannot hold true. Mathematically speaking, a relation of the following form cannot be valid for ϵ in a neighborhood of 0:

$$H_{\min}^{g_1(\epsilon)}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\epsilon}(A_k|A_1^{k-1}B)_\rho - ng_2(\epsilon) - k(\epsilon), \tag{1.9}$$

where the functions g_1, g_2, and k are dependent solely on ϵ and $|A|$ (the dimension of the A_k registers, assumed to be constant in n) and are independent of n. Furthermore, $g_1(\epsilon)$ and $g_2(\epsilon)$ are required to be *small* functions of ϵ, meaning they are continuous and approach 0 as ϵ tends to 0. To see that the smooth min-entropy cannot satisfy such a bound, simply consider the classical distribution $p_{A_1^nB}$ used above in Eq. 1.5. In this case, $H_{\min}^{g_1(\epsilon)}(A_1^n|B)_p = O(1)$ and for each k, $H_{\min}^{\epsilon}(A_k|A_1^{k-1}B)_p \geq \log 4/3$. For small ϵ, Eq. 1.9 would have a constant left-hand side and a linearly growing right-hand side, which results in a contradiction.

As a consequence of this impossibility, identities for the smooth min-entropy, like the chain rules [**VDTR13**], are much more restrictive. Similarly, tools like entropy accumulation [**DFR20**, **MFSR24**], which decompose the smooth min-entropy, are quite rigid, in the sense that they cannot be applied unless certain (Markov chain or non-signalling) conditions apply. It is also not clear how one could relax the conditions for such tools.

Eq. 1.9 brings us to the concept of *approximation chains*. For a state $\rho_{A_1^nB}$, we define a sequence of states $(\sigma^{(k)}_{A_1^kB})_{k=1}^n$ as an ϵ-approximation chain of ρ if for every $1 \leq k \leq n$, we have

$$\rho_{A_1^kB} \approx_\epsilon \sigma^{(k)}_{A_1^kB}. \tag{1.10}$$

It is noteworthy that the right-hand side of Eq. 1.9 is essentially a maximum over the sum of the min-entropies of all possible approximation chains of state $\rho_{A_1^n B}$.

As we will see in this thesis, one often encounters scenarios where direct bounds on the conditional entropies of the state ρ are unavailable; however, they can be obtained for the entropies of its approximation chain states. For instance, consider a state $\rho_{A_1^n B}$ where the A_k registers are produced sequentially. Ideally, each register A_k should be sampled independently from the rest. However, say imperfections in the production process introduce minor correlations between A_k and the earlier registers. In such a case, we might only be able to confirm that $\rho_{A_1^k B} \approx_\epsilon \rho_{A_k} \otimes \rho_{A_1^{k-1} B}$ for every k. Here the states $\sigma^{(k)}_{A_1^k B} := \rho_{A_k} \otimes \rho_{A_1^{k-1} B}$ form an ϵ-approximation chain for ρ. Despite these correlations, we expect the entropy of A_1^n given B to be large for ρ, as the A_k registers are almost independent of B. In such cases, a chain rule-like tool relating the entropy of ρ to the entropies of its approximation chain states would be helpful.

The impossibility of a bound of the form in Eq. 1.9 also implies that $H^\epsilon_{\min}(A_1^n|B)_\rho$ cannot be lower bounded meaningfully in terms of the min-entropies of the approximation chain states of ρ, since these approximation chain states can simply be states satisfying $H_{\min}(A_k|A_1^{k-1}B)_{\sigma^{(k)}} = H^\epsilon_{\min}(A_k|A_1^{k-1}B)_\rho$ for every k. On the other hand, the von Neumann entropy of a state can easily be bounded in terms of its approximation chain by using the continuity of the conditional von Neumann entropy [**AF04**, **Win16**] to modify Eq. 1.7 and deriving:

$$H(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H(A_k|A_1^{k-1}B)_{\sigma^{(k)}} - nf(\epsilon) \tag{1.11}$$

where $f(\epsilon) = O\left(\epsilon \log \frac{|A|}{\epsilon}\right)$.

Approximation chains are commonly encountered while studying imperfections and errors in protocols, approximations of entropic tools and cryptographic proofs. The absence of a comparable bound for the smooth min-entropy severely limits us. In this thesis, we initiate a study of approximation chains, with the aim of proving entropic bounds in terms of these objects and subsequently using them for cryptographic applications.

1.1. Outline and Results

In the following, we present a chapterwise overview of the main results in this thesis. Connections with existing literature and techniques used for proving these results will be

discussed in the chapters themselves. The main chapters of this thesis are based on the following papers and manuscripts:

(1) Smooth min-entropy lower bounds for approximation chains [MD24c].
(2) Proving security of BB84 under source correlations [MD24a].
(3) Universal chain rules from entropic triangle inequalities [MD24e].
(4) Security for parallel DIQKD [MD24b].

We have arranged the contents in the order in which they were developed, which is also roughly the increasing order of sophistication.

Chapter 2: Background

In this chapter, we review the notation and fundamental concepts used throughout this thesis. We present key tools in quantum information and information theory, and introduce cryptographic primitives used in our work. These include the entropy accumulation theorem (EAT), quantum key distribution (QKD), and device-independent quantum key distribution (DIQKD). Readers familiar with the subject may choose to skip this chapter and refer back to it as needed.

Chapter 3: Smooth min-entropy lower bounds for approximation chains

In Chapter 3, we begin our exploration of approximation chains by examining them under additional simplifying conditions. We first examine the simplest scenario: a state with *approximately independent registers*. This is a state $\rho_{A_1^n B}$ which, for every $1 \leq k \leq n$, satisfies

$$\left\| \rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1} B} \right\|_1 \leq \epsilon. \tag{1.12}$$

for some small $\epsilon > 0$ and a large n (in particular $n \gg \frac{1}{\epsilon}$). That is, for every k, the system A_k is almost independent of the system B and everything else which came before it. For simplicity, let us further assume that for all k the state $\rho_{A_k} = \rho_{A_1}$. Intuitively, one expects that the smooth min-entropy (with the smoothing parameter depending on ϵ and not on n) for such a state will be large and close to $\approx n(H(A_1) - g(\epsilon))$ (for some small function $g(\epsilon)$). However, it is not possible to prove this result using traditional techniques, which rely only on the triangle inequality and smoothing. The triangle inequality, in general, can only be used to bound the trace distance between $\rho_{A_1^n B}$ and $\otimes_{k=1}^n \rho_{A_k} \otimes \rho_B$ by $n\epsilon$, which will result in

a trivial bound when $n \gg \frac{1}{\epsilon}$.

To demonstrate such a smooth min-entropy lower bound, we introduce a key tool used throughout this thesis for analysing approximation chains: *the entropic triangle inequality*. For two general states ρ_{AB} and η_{AB}, such that $d := D^{\delta}_{\max}(\rho_{AB}\|\eta_{AB})$ (see Definition 2.19), we can easily bound the smooth min-entropy of ρ in terms of the min-entropy of η by using the fact that there exists a state $\tilde{\rho}_{AB}$ such that $\tilde{\rho}_{AB} \approx_{\delta} \rho_{AB}$ and

$$\tilde{\rho}_{AB} \leq e^{d}\eta_{AB}. \tag{1.13}$$

Suppose, the state σ_B satisfies $D_{\max}(\eta_{AB}\|\mathbb{1}_A \otimes \sigma_B) = -H_{\min}(A|B)_{\eta}$, then

$$\tilde{\rho}_{AB} \leq e^{-(H_{\min}(A|B)_{\eta} - d)}\,\mathbb{1}_A \otimes \sigma_B. \tag{1.14}$$

This implies that

$$H^{\delta}_{\min}(A|B)_{\rho} \geq H_{\min}(A|B)_{\eta} - D^{\delta}_{\max}(\rho_{AB}\|\eta_{AB}) \tag{1.15}$$

We call this an entropic triangle inequality, since it is based on the triangle inequality property of $D_{\max}$. We can further improve this smooth min-entropy triangle inequality to (Lemma 3.5)

$$H^{\epsilon+\delta}_{\min}(A|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\eta} - \frac{\alpha}{\alpha-1}D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) - \frac{g_1(\delta,\epsilon)}{\alpha-1} \tag{1.16}$$

for some function g_1, $\epsilon + \delta < 1$ and $1 < \alpha \leq 2$.

Our general strategy for bounding the smooth min-entropy of a state in terms of its approximation chain will be to first bound the "one-shot information theoretic" distance (the smooth max-relative entropy distance) between the real state ρ ($\rho_{A_1^n B}$ in the above scenario) and a virtual, but *nicer* state, η ($\otimes_{k=1}^{n}\rho_{A_k} \otimes \rho_B$ above) by $nf(\epsilon)$ for some small $f(\epsilon)$. Then, we use one of the entropic triangle inequalities above to reduce the problem of bounding the smooth min-entropy on state ρ to that of bounding an entropy on the state η. Using this strategy, we prove (Corollary 3.10) that for states satisfying the approximately independent registers assumption, we have that

$$H^{\tilde{O}(\epsilon^{1/4})}_{\min}(A_1^n|B)_{\rho} \geq n\left(H(A_1)_{\rho} - \tilde{O}(\epsilon^{1/4})\right) - \tilde{O}\left(\frac{1}{\epsilon^{3/4}}\right) \tag{1.17}$$

where $\tilde{O}$ hides logarithmic factors in $1/\epsilon$.

We also consider the scenario of approximate entropy accumulation in this chapter. In the setting for entropy accumulation, a sequence of channels $\mathcal{M}_k : R_{k-1} \to A_k B_k R_k$ for $1 \leq k \leq n$ sequentially act on a state $\rho^{(0)}_{R_0 E}$ to produce the state $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E})$. It is

assumed that the channels $\mathcal{M}_k$ are such that the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$ is satisfied for every k. The entropy accumulation theorem (EAT) [DFR20], then provides a tight lower bound for $H_{\min}^{\delta}(A_1^n|B_1^nE)_\rho$. We consider an approximate version of the above setting where the channels $\mathcal{M}_k$ themselves do not necessarily satisfy the Markov chain condition, but they can be ϵ-approximated by a sequence of channels $\mathcal{M}'_k$, which satisfies certain Markov chain conditions. Such relaxations are important to understand the behaviour of cryptographic protocols, like device-independent quantum key distribution [AFDF+18], implemented with imperfect devices [JK23, Tan23]. Once again we can model this scenario as an approximation chain: for every $1 \le k \le n$, the state produced in the k^{th} step satisfies

$$\begin{aligned}\rho_{A_1^k B_1^k E} &= \mathrm{tr}_{R_k} \circ \mathcal{M}_k \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \right) \\ &\approx_\epsilon \mathrm{tr}_{R_k} \circ \mathcal{M}'_k \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \right) := \sigma^{(k)}_{A_1^k B_1^k E}.\end{aligned}$$

Moreover, the assumptions on the channel $\mathcal{M}'_k$ guarantee that the state $\sigma^{(k)}_{A_1^k B_1^k E}$ satisfies the Markov chain condition $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$, and so one expects that the smooth min-entropy is large for the state ρ similar to the original setting.

Following the strategy described above, in Theorem 3.12, we show the following smooth min-entropy lower bound for the state $\rho_{A_1^n B_1^n E}$ for sufficiently small ϵ and an arbitrary $\delta > 0$

$$H_{\min}^{\delta}(A_1^n|B_1^nE)_\rho \ge \sum_{k=1}^{n} \inf_{\omega} H(A_k|B_k\tilde{R}_{k-1})_{\mathcal{M}'_k(\omega)} - nO(\epsilon^{1/24}) - O\left(\frac{1}{\epsilon^{1/24}}\right) \tag{1.18}$$

where the infimum is over all possible input states $\omega_{R_{k-1}\tilde{R}_{k-1}}$ for reference register $\tilde{R}_{k-1}$ isomorphic to R_{k-1}, and the dimensions $|A|$ and $|B|$ are assumed constant while using the asymptotic notation.

Chapter 4: Proving security of BB84 under source correlations

We use the techniques developed in Chapter 3 to address the source correlation problem in QKD [PCLN+22]. Briefly speaking, the security proofs of QKD require that one of the honest parties produce randomly and independently sampled quantum states in each round of the protocol. However, the states produced by a realistic quantum source are somewhat correlated across different rounds due to device memory and imperfections. These correlations are called *source correlations*. Proving security for QKD under such a correlated source has been a challenging problem, and no general satisfying solution was known before. In this chapter, we use the entropic triangle inequality to reduce the security of the BB84 QKD protocol with a correlated source to that of the QKD protocol with an almost perfect source, for which security can be proven using existing techniques. This allows us to

provide a general and simple proof of security for QKD in the presence of source correlations.

Chapter 5: Universal chain rules

In this chapter, we revisit the question of bounding the smooth min-entropy of a state in terms of the entropies of its approximation chains, focusing on general approximation chains unlike Chapter 3.

There are multiple alternative definitions of the smooth min-entropy, which are equal to the one we defined above up to a constant [Ren06, TRSS10, ABJT20]. One of these, is the $H_{\min}^{\downarrow}$ min-entropy and its smoothed variant $H_{\min}^{\downarrow,\epsilon}$, defined as:

$$H_{\min}^{\downarrow}(A|B)_{\rho} := \sup\left\{\lambda \in \mathbb{R} : \rho_{AB} \leq e^{-\lambda}\, \mathbb{1}_A \otimes \rho_B\right\} \tag{1.19}$$

$$H_{\min}^{\downarrow,\epsilon}(A|B)_{\rho} := \sup_{\tilde{\rho}} H_{\min}^{\downarrow}(A|B)_{\tilde{\rho}} \tag{1.20}$$

where the supremum is over all subnormalised states $\tilde{\rho}_{AB}$ which are ϵ-close to the state ρ in the purified distance. [TRSS10, Lemma 20] showed that this smooth min-entropy is equal to the conventional $H_{\min}^{\epsilon}$ up to a constant:

$$H_{\min}^{\epsilon/2}(A_1^n|B)_{\rho} - O\left(\log\frac{1}{\epsilon}\right) \leq H_{\min}^{\downarrow,\epsilon}(A_1^n|B)_{\rho} \leq H_{\min}^{\epsilon}(A_1^n|B)_{\rho}. \tag{1.21}$$

One can now ask whether $H_{\min}^{\downarrow,\epsilon}$ satisfies a chain rule like Eq. 1.9, that is, does

$$H_{\min}^{\downarrow,g_1(\epsilon)}(A_1^n|B)_{\rho} \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_{\rho} - ng_2(\epsilon) - k(\epsilon), \tag{1.22}$$

hold true for some g_1, g_2 and k as in Eq. 1.9? We prove that this is indeed the case. Establishing this in Theorem 5.7 is the first main result of this chapter. We term this a *universal chain rule* for the smooth min-entropy to emphasise the fact that it is true and meaningful for a constant $\epsilon \in (0,1)$ and an arbitrary $n \in \mathbb{N}$. It is worth noting that this chain rule allows us to break the smooth min-entropy of a system into a sum of the entropies of its parts, with the entropies being almost as strong as the smooth min-entropy itself.

The second major result in this chapter is an *unstructured* approximate entropy accumulation theorem (Theorem 5.8). Unlike the approximate EAT in Chapter 3, this theorem does not require the state ρ to be produced by a sequential process. Nor does it consider approximations at the level of channels. It shows that for any state $\rho_{A_1^n B_1^n E}$ with an approximation chain $(\sigma^{(k)}_{A_1^k B_1^k E})_{k=1}^n$ such that for every $1 \leq k \leq n$:

$$\sigma^{(k)}_{A_1^k B_1^k E} = \mathcal{N}_k\left(\tilde{\sigma}^{(k)}_{A_1^{k-1} B_1^{k-1} E R_k}\right) \tag{1.23}$$

for some state $\tilde{\sigma}^{(k)}_{A_1^{k-1}B_1^{k-1}ER_k}$ and a channel $\mathcal{N}_k : R_k \to A_kB_k$ which samples B_k independent of the previous registers, we have the bound

$$H_{\min}^{\tilde{O}(\epsilon^{1/6})}(A_1^n|B_1^nE)_\rho \geq \sum_{k=1}^{n} \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k(\omega)} - n\tilde{O}(\epsilon^{1/12}) - \tilde{O}\left(\frac{1}{\epsilon^{5/12}}\right) \tag{1.24}$$

where the infimum is over all states $\omega_{R_k\tilde{R}_k}$.

Chapter 6: Security for parallel DIQKD

Security proofs for QKD protocols assume fully characterized devices, which is unrealistic in practical implementations. Consequently, these implementations remain vulnerable to side-channel attacks, where an adversary exploits imperfections in preparation and detection devices to extract additional information [LWW+10, SRK+15]. Device-independent quantum key distribution (DIQKD) addresses this issue by implementing key distribution with untrusted or potentially malicious devices. The security of DIQKD protocols relies solely on the properties of quantum mechanics.

Typically, during a DIQKD protocol, multiple non-local games are played using quantum states shared between the participants. The ability of the participants to win these games with probabilities exceeding the limits of classical strategies ensures security of these protocols. These games may be played sequentially or parallelly. Sequential protocols are easier to analyse as they can be broken down into smaller steps, each depending only on the preceding steps. In contrast, during parallel DIQKD, the two parties input all the questions for the multiple games into their devices and receive all the answers at once. This simultaneous nature makes the analysis of these protocols significantly more challenging.

This chapter focuses on proving security for parallel DIQKD. We employ techniques developed for analysing parallel repetition of non-local games. Roughly speaking, we show that the state produced at the end of the protocol, $\rho_{A_1^nB_1^nX_1^nY_1^nE}$, where X_1^n, Y_1^n are the questions, A_1^n, B_1^n are the answers for the non-local games, and E is the adversary's register, has an approximation chain $\left(\sigma^{(k)}_{A_1^kB_1^kX_1^kY_1^kE}\right)_{k=1}^{n}$ satisfying:

$$\sigma^{(k)}_{A_1^kB_1^kX_1^kY_1^kE} = \mathcal{M}_k^{R_k\to X_kY_kA_kB_k}\left(\sigma^{(k,0)}_{A_1^{k-1}B_1^{k-1}X_1^{k-1}Y_1^{k-1}R_kE}\right) \tag{1.25}$$

where $\mathcal{M}_k$ is the channel applied by participants playing a single round of the non-local game. This result allows us to apply the unstructured approximate EAT proved in Chapter 5 to establish the smooth min-entropy lower bound required for the security proof. Our approach yields a more information-theoretic and general proof for parallel DIQKD

compared to previous proofs, although it also comes with certain limitations.

Chapter 7: Exploring further...

We conclude by briefly examining some problems in the broader field where approximation chains appear, and discussing how the techniques developed in this thesis might prove useful for them.

Chapter 2

Background

2.1. Notation

We use the notation $[n]$ to denote the set $\{1,2,\cdots,n\}$. For n quantum registers or variables $(X_1, X_2, \cdots, X_n)$, the notation X_i^j refers to the tuple $(X_i, X_{i+1}, \cdots, X_j)$. For a set $S \subseteq [n]$, the notation X_S represents the tuple $(X_i)_{i \in S}$.

For a classical probability distribution p_{AB}, the conditional probability distribution $p_{A|B}$ is defined as $p_{A|B}(a|b) := \frac{p_{AB}(a,b)}{p_B(b)}$ when $p_B(b) > 0$. For the case when $p_B(b) = 0$, we define $p_{A|B}$ to be the uniform distribution for our purposes. The probability distribution $q_B p_{A|B}$ for a distribution q on random variable B is defined as $q_B p_{A|B}(b,a) := q_B(b) p_{A|B}(a|b)$.

For a quantum register A, $|A|$ represents the dimension of the underlying Hilbert space. If X and Y are Hermitian operators, then the operator inequality $X \geq Y$ denotes the fact that $X - Y$ is a positive semidefinite operator and $X > Y$ denotes that $X - Y$ is a strictly positive operator. We write $\mathrm{supp}(X)$ to denote the support of the Hermitian operator X and use $X \ll Y$ to denote that $\mathrm{supp}(X) \subseteq \mathrm{supp}(Y)$.

A quantum state (or briefly just *state*) refers to a positive semidefinite operator with unit trace. At times, we will also need to consider positive semidefinite operators with trace less than equal to 1. We call these operators subnormalised states. We will denote the set of registers a quantum state describes (equivalently, its Hilbert space) using a subscript. For example, a quantum state on registers A and B, will be written as ρ_{AB} and its partial states on registers A and B, will be denoted as ρ_A and ρ_B. A classical-quantum state on registers X and B is given by $\rho_{XB} = \sum_x p(x) |x\rangle\langle x| \otimes \rho_{B|x}$, where $\rho_{B|x}$ are normalised quantum states on register B.

The identity operator on register A is denoted using $\mathbb{1}_A$.

The term "channel" is used for completely positive trace preserving (CPTP) linear maps between two spaces of Hermitian operators. A channel $\mathcal{N}$ mapping registers A to B will be denoted by $\mathcal{N}_{A\to B}$. We refer the reader to [Wat18] for more information about quantum states, channels and their properties.

A measurement (also called a positive operator valued measure or POVM) μ on register A with outcomes $\mathcal{Z}$ stored in register Z is a tuple of positive operators $(\mu(z))_{z\in\mathcal{Z}}$ such that $\mu(z) \geq 0$ and $\sum_{z\in\mathcal{Z}} \mu(z) = \mathbb{1}_Z$. For a quantum state, ρ_A, the probability of outcome z is given by $\mathrm{tr}(\mu(z)\rho_A)$.

Finally, we note that we use base e for both the exp and log functions in this thesis.

For quick reference, we summarise the key notation used throughout this thesis in Table 2.1.

Notation	**Brief Description**
$[n]$	Set $\{1,2,\cdots,n\}$
X_i^j	Tuple $(X_i, X_{i+1}, \cdots, X_j)$
X_S	Tuple $(X_i)_{i\in S}$ for $S \subseteq [n]$
$p_{A\|B}$	Conditional probability distribution of A given B for p_{AB}
$\|A\|$	Dimension of Hilbert space of register A
$X \geq Y$	$X - Y$ is positive semidefinite
$X > Y$	$X - Y$ is strictly positive
$\mathrm{supp}(X)$	Support of operator X
$X \ll Y$	$\mathrm{supp}(X) \subseteq \mathrm{supp}(Y)$
$\mathbb{1}_A$	Identity operator on register A
$\mathcal{N}_{A\to B}$	Channel from register A to B
$\exp(\cdot)$	Exponential function with base e
$\log(\cdot)$	Logarithm function with base e

Table 2.1. Summary of notation used in this thesis.

2.2. Distance measures

2.2.1. Schatten p-norms

Similar to how $\|\cdot\|_p$ is defined for vectors, one can define the Schatten p-norm for matrices, including quantum states, as well.

Definition 2.1. *For $1 \leq p \leq \infty$, the Schatten p-norm for a matrix A is defined as*

$$\|A\|_p := \left(\operatorname{tr}\left((A^\dagger A)^{\frac{p}{2}}\right)\right)^{\frac{1}{p}}. \tag{2.1}$$

The p-norm for a matrix A is in fact equal to the p-norm for the vector of its singular values, $s(A)$:

$$\|A\|_p = \|s(A)\|_p. \tag{2.2}$$

In quantum information, we are particularly interested in the Schatten 1-norm also called the *trace norm*, which is useful for comparing the distance between quantum states. The Holevo-Helstrom theorem gives us an operational interpretation for the trace norm.

Theorem 2.2 (Holevo-Helstrom theorem [Wat18, Theorem 3.4])**.** *Suppose a source produces quantum states ρ and σ each with probability $1/2$, then the probability of successfully determining which state was prepared is at most*

$$\frac{1}{2}\left(1 + \frac{1}{2}\|\rho - \sigma\|_1\right). \tag{2.3}$$

This probability is achievable for appropriately chosen measurements.

In particular, in cryptography, if a protocol produces a state ρ, which has a trace norm distance at most ϵ from the ideal state of the protocol, then the real state ρ acts like the ideal state under all operations up to an error probability of ϵ [PR14].

2.2.2. Diamond norm for channels

We will also use the diamond norm distance as a measure of the distance between two channels.

Definition 2.3. *For a linear transform $\mathcal{N}_{A\to B}$ from operators on register A to operators on register B, the diamond norm distance is defined as*

$$\|\mathcal{N}_{A\to B}\|_\diamond := \max_{X_{AR}:\|X_{AR}\|_1\leq 1} \|\mathcal{N}_{A\to B}(X_{AR})\|_1 \tag{2.4}$$

where the maximum is taken over all Hilbert spaces R (fixing $|R| = |A|$ is sufficient) and operators X_{AR} such that $\|X_{AR}\|_1 \leq 1$.

Similar to the trace norm for states, the diamond norm determines the probability with which one can operationally distinguish two channels [Wat18, Theorem 3.52].

2.2.3. Fidelity

The fidelity and the generalised fidelity are distance measures which are closely related to the trace norm distance. They are defined as follows.

Definition 2.4. *The fidelity between two positive operators P and Q is defined as*

$$F(P,Q) = \left\|\sqrt{P}\sqrt{Q}\right\|_1^2. \tag{2.5}$$

Definition 2.5. *The generalised fidelity between two subnormalised states ρ and σ is defined as*

$$F_*(\rho,\sigma) := \left(\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 + \sqrt{(1-\operatorname{tr}\rho)(1-\operatorname{tr}\sigma)}\right)^2. \tag{2.6}$$

2.2.4. Purified distance

The generalised fidelity turns out to be more natural than the trace norm for quantum states while *smoothing* entropic measures. The purified distance is a metric defined using the generalised fidelity. It will be used to quantify the distance while smoothing.

Definition 2.6. *The purified distance between two subnormalised states is defined as*

$$P(\rho,\sigma) := \sqrt{1 - F_*(\rho,\sigma)}. \tag{2.7}$$

Purified distance and the trace norm are related by the following inequalities [Tom16, Lemma 3.5], which are called the Fuchs-van de Graaf inequalities:

$$\frac{1}{2}\|\rho-\sigma\|_1 \le P(\rho,\sigma) \le \|\rho-\sigma\|_1^{1/2}. \tag{2.8}$$

2.3. Entropic quantities

We briefly define and discuss the commonly used entropic quantities in this thesis. We refer the reader to the books [Wil13] and [Tom16] for a more comprehensive treatment of these topics.

Throughout this thesis, we will use base e for both the functions log and exp. Therefore, all the entropies in this thesis are in *nat* units. We note that this does not result in a major change in well known information theoretic identities.

2.3.1. von Neumann entropy

Entropic quantities are used to quantify the amount of information or uncertainty in different scenarios. The most commonly used entropic measure is the *von Neumann entropy*, which is defined for a state ρ_A as

$$H(A)_\rho := -\operatorname{tr}(\rho_A \log \rho_A). \tag{2.9}$$

The von Neumann entropy operationally characterises the asymptotic rate of compression for an i.i.d. source (see, for example [Wil13, Theorem 18.2.1]). We can use this definition to create a conditional entropy, which measures the uncertainty of a register with respect to another register (see [Wil13, Corollary 22.4.2] for an operational interpretation of $H(A|B)_\rho$ in terms of the entanglement required for state transfer).

Definition 2.7. *For a state ρ_{AB}, the conditional von Neumann entropy of register A with respect to register B is defined*

$$H(A|B)_\rho := H(AB)_\rho - H(B)_\rho. \tag{2.10}$$

The conditional von Neumann entropy satisfies the following chain rule:

$$H(AB|C) = H(A|C) + H(B|AC). \tag{2.11}$$

It also satisfies the data processing inequality, which states that if one applies a channel $\Phi_{B\to B'}$ to the state ρ_{AB}, then the entropy increases, that is

$$H(A|B')_{\Phi(\rho)} \geq H(A|B)_\rho. \tag{2.12}$$

2.3.2. Quantum relative entropy

The quantum relative entreopy, also called the Umegaki relative entropy, can be viewed as an information theoretic measure of the distance between two states, although it is not a metric (it is not even symmetric). It plays an important role in information theory and this thesis specifically. Operationally, it can be interpreted through the task of hypothesis testing [HP91].

Definition 2.8. *The relative entropy between two nonzero positive semidefinite operators P and Q is given by*

$$D(P\|Q) := \begin{cases} \frac{\operatorname{tr}(P\log P - P\log Q)}{\operatorname{tr}(P)} & \text{if } (P \ll Q) \\ \infty & \text{else.} \end{cases} \tag{2.13}$$

The relative entropy can be viewed as the parent quantity for the von Neumann entropy. In particular, we have that for a state ρ_{AB}

$$H(A|B)_\rho = -D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) \tag{2.14}$$

$$= \sup_{\sigma_B} -D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{2.15}$$

where the supremum in the last line is over all states σ_B on the register B.

The relative entropy also satisfies a data processing inequality. For any channel Φ and positive operators P and Q, we have

$$D(P\|Q) \geq D(\Phi(P)\|\Phi(Q)). \tag{2.16}$$

In this thesis, we will also need the measured relative entropy D_m.

Definition 2.9. *The measured relative entropy between two nonzero positive semidefinite operators P and Q is given by*

$$D_m(P\|Q) := \sup_{\mathcal{M}} D(\mathcal{M}(P)\|\mathcal{M}(Q)) \tag{2.17}$$

where the supremum is taken over all measurement channels $\mathcal{M}$, that is, channels such that $\mathcal{M}(\rho) = \sum_{x\in\mathcal{X}} \operatorname{tr}(M_x\rho)\,|x\rangle\langle x|$ for some POVM elements $(M_x)_x$ and an alphabet $\mathcal{X}$.

Due to the data processing inequality, we have

$$D_m(P\|Q) \leq D(P\|Q) \tag{2.18}$$

for all positive operators P and Q.

2.3.3. Mutual information

The mutual information between two registers for a state provides a measure of the correlations between them.

Definition 2.10. *For a state ρ_{AB} the mutual information between registers A and B is defined as*

$$I(A:B)_\rho := D(\rho_{AB}\|\rho_A \otimes \rho_B). \tag{2.19}$$

Operationally, it characterises the entanglement-assisted capacity for communication over a channel (see, for example [**Wil13**, Theorem 21.3.1]).

2.3.4. Rényi relative entropies

We follow the notation in Tomamichel's book [Tom16] for Rényi entropies. These entropies are necessary for understanding information tasks in the one-shot regime.

Definition 2.11 ([MLDS+13, WWY14]). *The sandwiched α-Rényi relative entropy for $\alpha \in [\frac{1}{2},1) \cup (1,\infty]$ between the positive operator P and Q is defined as*

$$\tilde{D}_\alpha(P\|Q) = \begin{cases} \frac{1}{\alpha-1} \log \frac{\mathrm{tr}(Q^{-\alpha'/2} P Q^{-\alpha'/2})^\alpha}{\mathrm{tr}(P)} & \text{if } (\alpha < 1 \text{ and } P \pm Q) \text{ or } (P \ll Q) \\ \infty & \text{else.} \end{cases} \tag{2.20}$$

where $\alpha' = \frac{\alpha-1}{\alpha}$.

Definition 2.12 ([Pet86]). *For $\alpha \in [0,1) \cup (1,2]$, the Petz α-Rényi relative entropy between the positive operators P and Q is defined as*

$$\bar{D}_\alpha(P\|Q) = \begin{cases} \frac{1}{\alpha-1} \log \frac{\mathrm{tr}(P^\alpha Q^{1-\alpha})}{\mathrm{tr}(P)} & \text{if } (\alpha < 1 \text{ and } P \pm Q) \text{ or } (P \ll Q) \\ \infty & \text{else.} \end{cases} \tag{2.21}$$

In the limit $\alpha \to \infty$, the sandwiched relative entropy becomes equal to the max-relative entropy, $D_{\max}$, which can equivalently be defined as follows.

Definition 2.13. *The max-relative entropy between two positive operator P and Q is defined as*

$$D_{\max}(P\|Q) := \inf \left\{ \lambda \in \mathbb{R} : P \le e^\lambda Q \right\}. \tag{2.22}$$

Both $\tilde{D}_\alpha$ and $\bar{D}_\alpha$ are monotonically increasing in α. In the limit of $\alpha \to 1$, both the Petz and the sandwiched relative entropies equal the quantum relative entropy, $D(P\|Q)$.

For α in their range of definition, both the Petz and the sandwiched Rényi relative entropies satisfy the data processing inequality, that is, for all quantum channels Φ and positive operators P and Q, we have

$$\tilde{D}_\alpha(P\|Q) \ge \tilde{D}_\alpha(\Phi(P)\|\Phi(Q)) \tag{2.23}$$

$$\bar{D}_\alpha(P\|Q) \ge \bar{D}_\alpha(\Phi(P)\|\Phi(Q)). \tag{2.24}$$

Finally, we note that these relative entropies satisfy

$$\tilde{D}_\alpha(P\|Q) \le \bar{D}_\alpha(P\|Q) \tag{2.25}$$

for all $\alpha \in [0,2]$ and positive operators P and Q.

2.3.5. Rényi conditional entropies

We can use the relative entropies defined above to define conditional entropies, similar to how the von Neumann conditional entropy is derived from the relative entropy (Eq. 2.14 and 2.15).

Definition 2.14. *For the subnormalised state ρ_{AB}, we can define the sandwiched Rényi conditional entropies as*

$$\tilde{H}^{\uparrow}_{\alpha}(A|B)_{\rho} := \sup_{\sigma_B} -\tilde{D}_{\alpha}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{2.26}$$

$$\tilde{H}^{\downarrow}_{\alpha}(A|B)_{\rho} := -\tilde{D}_{\alpha}(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) \tag{2.27}$$

for $\alpha \in [\frac{1}{2},1) \cup (1,\infty]$ and the Petz Rényi conditional entropies as

$$\bar{H}^{\uparrow}_{\alpha}(A|B)_{\rho} := \sup_{\sigma_B} -\bar{D}_{\alpha}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \tag{2.28}$$

$$\bar{H}^{\downarrow}_{\alpha}(A|B)_{\rho} := -\bar{D}_{\alpha}(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) \tag{2.29}$$

for $\alpha \in [0,1) \cup (1,2]$. The supremum in the definition for $\tilde{H}^{\uparrow}_{\alpha}$ and $\bar{H}^{\uparrow}_{\alpha}$ is over all quantum states σ_B on register B.

In the limit $\alpha \to 1$, all four of these conditional entropies converge to the von Neumann conditional entropy $H(A|B)$.

These conditional entropies satisfy a data processing inequality similar to the von Neumann entropy (see, for example [Tom16, Corollary 5.1]). For a channel $\Phi_{B\to B'}$ and state ρ_{AB}, we have

$$\mathbb{H}_{\alpha}(A|B')_{\Phi(\rho)} \geq \mathbb{H}_{\alpha}(A|B)_{\rho} \tag{2.30}$$

where $\mathbb{H}_{\alpha}$ can be any one of $\{\bar{H}^{\uparrow}_{\alpha}, \bar{H}^{\downarrow}_{\alpha}, \tilde{H}^{\uparrow}_{\alpha}, \tilde{H}^{\downarrow}_{\alpha}\}$ for α lying in the appropriate domain of definition for the entropy.

2.3.6. Min-entropy

The conditional entropy $\tilde{H}^{\uparrow}_{\infty}$ is called the min-entropy and is usually written as $H_{\min}$. It characterises the amount of randomness one can extract from a classical register which is correlated to an adversary's register. It can be equivalently defined as:

Definition 2.15 ([Ren06]). *For a subnormalised state ρ_{AB}, the min-entropy of register A given register B is defined as*

$$H_{\min}(A|B)_{\rho} := \sup\left\{\lambda \in \mathbb{R} : \text{ there exists state } \sigma_B \text{ such that } \rho_{AB} \leq e^{-\lambda}\,\mathbb{1}_A \otimes \sigma_B\right\}. \tag{2.31}$$

Another version of the min-entropy, that will be relevant in this thesis is given by $\tilde{H}^{\downarrow}_{\infty}$. We denote it here using $H^{\downarrow}_{\min}$. It can alternatively be defined as:

Definition 2.16. *For a subnormalised state ρ_{AB}, the $H^{\downarrow}_{\min}$ of register A given register B is defined as*

$$H^{\downarrow}_{\min}(A|B)_{\rho} := \sup\left\{\lambda \in \mathbb{R} : \rho_{AB} \leq e^{-\lambda}\, \mathbb{1}_A \otimes \rho_B\right\} \tag{2.32}$$

$$= -\log \left\| \rho_B^{-\frac{1}{2}} \rho_{AB} \rho_B^{-\frac{1}{2}} \right\|_{\infty} \tag{2.33}$$

For a classical distribution p_{AB}, the min-entropies above simplify to

$$H_{\min}(A|B)_p = -\log \sum_b p(b) \max_a p(a|b) \tag{2.34}$$

$$H^{\downarrow}_{\min}(A|B)_p = -\log \max_{a,b} p(a|b). \tag{2.35}$$

The expressions for these entropies in the classical case clearly demonstrates the difference between $H_{\min}$ and $H^{\downarrow}_{\min}$. Since $H_{\min}$ is the negative log of the guessing probability [**KRS09**], it averages over the values of B. Whereas $H^{\downarrow}_{\min}$ selects the worst possible value of B and then calculates the guessing probability for this value.

2.3.7. Max-entropy

The conditional entropy $\tilde{H}^{\uparrow}_{1/2}$ is called the max-entropy and is usually written as $H_{\max}$. The max-entropy essentially characterises the performance of source and channel coding in the one-shot setting.

Definition 2.17. *For a subnormalised state ρ_{AB}, the max-entropy of register A given register B is defined as*

$$H_{\max}(A|B)_{\rho} := \sup_{\sigma_B} \log F(\rho_{AB}, \mathbb{1}_A \otimes \sigma_B). \tag{2.36}$$

2.3.8. Smoothed entropies

Typically in applications, one uses smoothed versions of one-shot entropies. This allows one to consider finite error for applications, and it also results in tighter bounds.

We begin by defining an ϵ-ball around a state.

Definition 2.18. *For a subnormalised state ρ, and $0 \leq \epsilon < \sqrt{\mathrm{tr}(\rho)}$, we define the ϵ-ball around ρ as*

$$B_\epsilon(\rho) := \{\tilde{\rho} \geq 0 : P(\rho, \tilde{\rho}) \leq \epsilon \text{ and } \mathrm{tr}\,\tilde{\rho} \leq 1\}. \tag{2.37}$$

Using this, we define smoothed versions of $D_{\max}$, $H_{\min}$, $H^{\downarrow}_{\min}$ and $H_{\max}$ as follows.

Definition 2.19. *We define the smooth max-relative entropy between a subnormalised state ρ and a positive operator Q for $0 \leq \epsilon < \sqrt{\mathrm{tr}(\rho)}$ as*

$$D^{\epsilon}_{\max}(\rho\|Q) := \inf_{\tilde{\rho} \in B_\epsilon(\rho)} D_{\max}(\tilde{\rho}\|Q). \tag{2.38}$$

Definition 2.20 ([Ren06]). *For a subnormalised state ρ_{AB}, and $0 \leq \epsilon < \sqrt{\mathrm{tr}(\rho)}$, we define the smooth min-entropy of register A given register B as*

$$H^{\epsilon}_{\min}(A|B)_\rho := \sup_{\tilde{\rho} \in B_\epsilon(\rho)} H_{\min}(A|B)_{\tilde{\rho}}. \tag{2.39}$$

Definition 2.21. *For a subnormalised state ρ_{AB}, and $0 \leq \epsilon < \sqrt{\mathrm{tr}(\rho)}$, we define the $H^{\downarrow,\epsilon}_{\min}$ of register A given register B as*

$$H^{\downarrow,\epsilon}_{\min}(A|B)_\rho := \sup_{\tilde{\rho} \in B_\epsilon(\rho)} H^{\downarrow}_{\min}(A|B)_{\tilde{\rho}}. \tag{2.40}$$

Definition 2.22. *For a subnormalised state ρ_{AB}, and $0 \leq \epsilon < \sqrt{\mathrm{tr}(\rho)}$, we define the smooth max-entropy of register A given register B as*

$$H^{\epsilon}_{\max}(A|B)_\rho := \inf_{\tilde{\rho} \in B_\epsilon(\rho)} H_{\max}(A|B)_{\tilde{\rho}}. \tag{2.41}$$

[TRSS10, Lemma 20] relates the two smooth min-entropies above as:

$$H^{\epsilon}_{\min}(A|B)_\rho - \log\left(\frac{2}{\epsilon^2} + \frac{1}{1-\epsilon}\right) \leq H^{\downarrow,2\epsilon}_{\min}(A|B)_\rho \leq H^{2\epsilon}_{\min}(A|B)_\rho. \tag{2.42}$$

We primarily use $H^{\epsilon}_{\min}$ as the smooth min-entropy in this thesis. In Chapter 5, we work with $H^{\downarrow,\epsilon}_{\min}$ to establish a universal chain rule for the smooth min-entropy.

The following lemma (originally proven in [TCR09]) relates the smooth min and max-entropies with the Rényi conditional entropies.

Lemma 2.23 ([DFR20, Lemma B.10]). *For a quantum state ρ and $\alpha \in (1,2]$ and $\epsilon \in (0,1)$, we have*

$$H^{\epsilon}_{\min}(A|B) \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_\rho - \frac{g(\epsilon)}{\alpha - 1} \tag{2.43}$$

$$H^{\epsilon}_{\max}(A|B) \leq \tilde{H}^{\downarrow}_{1/\alpha}(A|B)_\rho + \frac{g(\epsilon)}{\alpha - 1} \tag{2.44}$$

for $g(\epsilon) := -\log(1-\sqrt{1-\epsilon^2})$.

This lemma also shows that for $\alpha \in (1,2]$ the sandwiched conditional entropy $\tilde{H}^{\uparrow}_{\alpha}$ behaves like the min-entropy, whereas for $\alpha \in [1/2,1)$, it behaves like the max-entropy.

One of the key advantages of smoothing is that for a given ϵ, the regularisations of the smoothed quantities equal their von Neumann counterparts [TCR09, Tom12, TH13]. Specifically,

$$\frac{1}{n} D^{\epsilon}_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho\|\sigma) \tag{2.45}$$

$$\frac{1}{n} H^{\epsilon}_{\min}(A_1^n|B_1^n)_{\rho_{AB}^{\otimes n}} = H(A|B)_{\rho} \tag{2.46}$$

$$\frac{1}{n} H^{\epsilon}_{\max}(A_1^n|B_1^n)_{\rho_{AB}^{\otimes n}} = H(A|B)_{\rho} \tag{2.47}$$

This allows one to derive the i.i.d. results as special cases of their one-shot generalisations.

2.4. Privacy amplification

Privacy amplification is one of the most important primitives in cryptography. It allows a party holding a string, which might be correlated with an adversary's information, to extract a key that is completely random from the adversary's viewpoint. This primitive is often used at the end of cryptographic protocols like QKD. The amount of randomness, which can be extracted from a register is quantified by the smooth min-entropy as the leftover hashing lemma shows below.

Definition 2.24. *A family of functions $\mathcal{F}$ from register X to Z is called two universal if for all $x \neq x'$ and a $F \in \mathcal{F}$ chosen uniformly at random, we have*

$$\Pr[F(x) = F(x')] \leq \frac{1}{|Z|}. \tag{2.48}$$

Lemma 2.25 (Leftover hashing lemma [RK05, TRSS10])**.** *For a set of 2-universal hash functions $\mathcal{F}$ from A to Z and a classical-quantum state $\rho_{AE} = \sum_a p(a) |a\rangle\langle a| \otimes \rho_E^{(a)}$, let ρ_{ZEF} be the state produced by applying a random hash function $F \in \mathcal{F}$ to register A to produce the output register Z. This state satisfies*

$$\frac{1}{2} \left\| \rho_{ZEF} - \tau_Z \otimes \rho_{EF} \right\| \leq 2\epsilon + e^{\frac{1}{2}(\log|Z| - H^{\epsilon}_{\min}(A|E))} \tag{2.49}$$

where $\tau_Z := \frac{1}{|Z|} \mathbb{1}_Z$ is the completely mixed state on register Z.

So, if one chooses a family of hash functions with output length $|Z| = H_{\min}^{\epsilon}(A|E)_{\rho} - 2\log 1/\epsilon$, the output state ρ_{ZEF} is 3ϵ close to $\tau_Z \otimes \rho_{EF}$. Thus, one can extract $H_{\min}^{\epsilon}(A|E)_{\rho} - O(1)$ amount of randomness independent of the adversary for the state ρ.

2.5. Substate theorem

The quantum substate theorem stated below is one of the major results used to bound the smooth max-relative entropy in this thesis. This theorem serves as a tool to convert bounds on the measured relative entropy to bounds on the smooth max-relative entropy.

Theorem 2.26 (Quantum substate theorem [JRS02,JN11])**.** *Let ρ and σ be two normalised states on the same Hilbert space. Then for any $\epsilon \in (0,1)$, we have*

$$D_{\max}^{\sqrt{\epsilon}}(\rho\|\sigma) \leq \frac{D_m(\rho\|\sigma)+1}{\epsilon} + \log\frac{1}{1-\epsilon}. \tag{2.50}$$

Usually, the above theorem is stated with the relative entropy $D(\rho\|\sigma)$ on the right-hand side instead of the measured relative entropy $D_m(\rho\|\sigma)$. Since, in the proofs in Chapter 5 we are only able to derive bounds on the measured relative entropy, we need the stronger version stated above. It is, therefore, instructive to understand how it is possible to use $D_m(\rho\|\sigma)$ in the bound above.

[JRS02,JN11] actually prove a bound on $D_{\max}^{\sqrt{\epsilon}}(\rho\|\sigma)$ in terms of a divergence they call the *observational divergence* D_{obs}, which is defined as

$$D_{obs}(\rho\|\sigma) := \sup\left\{\operatorname{tr}(P\rho)\log\frac{\operatorname{tr}(P\rho)}{\operatorname{tr}(P\sigma)} : \text{ for } 0 \leq P \leq \mathbb{1} \text{ such that } \operatorname{tr}(P\sigma) \neq 0\right\}. \tag{2.51}$$

[JN11, Theorem 1] proves that

$$D_{\max}^{\sqrt{\epsilon}}(\rho\|\sigma) \leq \frac{D_{\text{obs}}(\rho\|\sigma)}{\epsilon} + \log\frac{1}{1-\epsilon}. \tag{2.52}$$

Observe that for any binary measurement $(P, \mathbb{1}-P)$, we have

$$\begin{aligned} D_m(\rho\|\sigma) &\geq \operatorname{tr}(P\rho)\log\frac{\operatorname{tr}(P\rho)}{\operatorname{tr}(P\sigma)} + (1-\operatorname{tr}(P\rho))\log\frac{1-\operatorname{tr}(P\rho)}{1-\operatorname{tr}(P\sigma)} \\ &\geq \operatorname{tr}(P\rho)\log\frac{\operatorname{tr}(P\rho)}{\operatorname{tr}(P\sigma)} + (1-\operatorname{tr}(P\rho))\log(1-\operatorname{tr}(P\rho)) \\ &\geq \operatorname{tr}(P\rho)\log\frac{\operatorname{tr}(P\rho)}{\operatorname{tr}(P\sigma)} - 1 \end{aligned}$$

which implies, that

$$D_{obs}(\rho\|\sigma) \leq D_m(\rho\|\sigma) + 1. \tag{2.53}$$

This allows us to use $D_m(\rho\|\sigma)$ in the bound for the substate theorem.

2.6. Entropy accumulation theorem (EAT)

As mentioned in Chapter 1, unlike the von Neumann entropy, one cannot generally decompose the smooth min-entropy of an n-partite system into the smooth min-entropies of its n subsystems. The entropy accumulation theorem (EAT) [DFR20] allows one to break the smooth min-entropy of a large system produced by a sequential process into the sum of its parts, like a chain rule, under certain conditions. This technique is very powerful and allows one to adapt security proofs based on the i.i.d. assumption to the one-shot setting. As a result, it is able to provide tight bounds on the rates of cryptographic applications.

2.6.1. Quantum Markov chains

EAT requires certain Markov chain conditions to be true. We begin by defining a Markov chain.

Definition 2.27. *A state ρ_{ABC} is said to obey the Markov chain $A \leftrightarrow B \leftrightarrow C$ if there exist Hilbert spaces $\{a_j\}_j$ and $\{c_j\}_j$ such that B satisfies the following isomorphism under the isometry $V : B \to \left(\bigoplus_j a_j\right) \otimes \left(\bigoplus_k c_k\right)$*

$$B \cong \bigoplus_j a_j \otimes c_j$$

and

$$V \rho_{ABC} V^\dagger = \bigoplus_j \rho_{Aa_j} \otimes \rho_{c_j C}.$$

Equivalently, ρ_{ABC} satisfies $A \leftrightarrow B \leftrightarrow C$ if and only if $I(A : C|B)_\rho = 0$.

For a state ρ_{ABC} satisfying the Markov chain $A \leftrightarrow B \leftrightarrow C$, the information theoretic characterisation $I(A : C|B)_\rho = 0$ implies that the register C holds no information about register A in addition to the information contained in register B.

2.6.2. Overview

The entropy accumulation theorem (EAT) [DFR20] provides a tight and simple lower bound for the smooth min-entropy $H^\epsilon_{\min}(A_1^n|B_1^n E)_\rho$ of sequential processes, under certain Markov chain conditions. The state $\rho_{A_1^n B_1^n E}$ in the setting for EAT is produced by a sequential process of the form shown in Fig. 2.1. The process begins with the registers R_0 and E. In the context of a cryptographic protocol, the register R_0 is usually held by the honest parties, whereas the register E is held by the adversary. Then, in each round $k \in [n]$

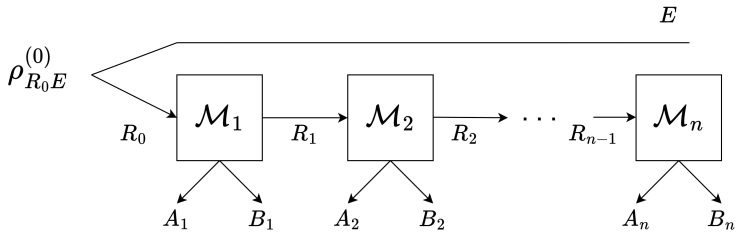

Fig. 2.1. The setting for the entropy accumulation theorem.

of the process, a channel $\mathcal{M}_k : R_{k-1} \to A_k B_k R_k$ is applied on the register R_{k-1} to produce the registers A_k, B_k and R_k. The registers A_1^n usually contain a partially secret raw key and the registers B_1^n contain the side information about A_1^n revealed to the adversary during the protocol. It can be seen that for cryptographic applications if one can prove a lower bound for $H^\epsilon_{\min}(A_1^n|B_1^n E)$, then one can also create secret key of almost the same length using privacy amplification.

EAT requires that for every $k \in [n]$, the side information B_k satisfies the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$, that is, the side information revealed in the k^{th} round does not reveal anything more about the secret registers of the previous rounds than was already known to the adversary through $B_1^{k-1}E$. Under this assumption, EAT provides a lower bound for the smooth min-entropy. We state the result in the following theorem.

Theorem 2.28 ([DFR20]). *If a state $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E})$ produced through the sequential process shown in Fig. 2.1, satisfies the Markov chain conditions $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$ for every $k \in [n]$, then*

$$H^\epsilon_{\min}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^{n} \inf_{\omega_{R_{k-1}\tilde{R}}} H(A_k|B_k\tilde{R})_{\mathcal{M}_k(\omega)} - c\sqrt{n} \tag{2.54}$$

where the infimum is taken over all input states to the channels $\mathcal{M}_k$ and $c > 0$ is a constant depending only on $|A|$ (size of registers A_k) and ϵ.

2.6.3. EAT with testing

EAT is usually applied while conditioning on the outcome of certain statistics. This helps ensure that the infimums occuring in the right-hand side of the bound in Eq. 2.54 are non-zero. We present the framework for testing in EAT here and restate the theorem with testing as Theorem 2.30. We use the same framework for our approximate generalisations of the theorem as well.

In this section, we will consider the channels $\mathcal{M}_k$ which map registers R_{k-1} to $A_kB_kX_kR_k$ such that X_k is a classical value which is determined using the registers A_k and B_k (We assume that all X_k have the same alphabet $\mathcal{X}$). Concretely, suppose that for every k, there exist a channel $\mathcal{T}_k : A_kB_k \to A_kB_kX_k$ of the form

$$\mathcal{T}_k(\omega_{A_kB_k}) = \sum_{y,z} \Pi^{(y)}_{A_k} \otimes \Pi^{(z)}_{B_k} \omega_{A_kB_k} \Pi^{(y)}_{A_k} \otimes \Pi^{(z)}_{B_k} \otimes |f(y,z)\rangle\langle f(y,z)|_{X_k} \tag{2.55}$$

where $\{\Pi^{(y)}_{A_k}\}_y$ and $\{\Pi^{(z)}_{B_k}\}_z$ are orthogonal projectors and f is some deterministic function which uses the measurement results y and z to create the output register X_k. We assume that every $\mathcal{M}_k = \mathcal{T}_k \circ \mathcal{M}'_k$.

Let $\mathbb{P}$ be the set of probability distributions over the alphabet of X registers, $\mathcal{X}$. Let R be any register isomorphic to R_{k-1}. For a probability $q \in \mathbb{P}$ and a channel $\mathcal{N}_k : R_{k-1} \to A_kB_kX_kR_k$, we define the set of states

$$\Sigma_k(q|\mathcal{N}_k) := \{\nu_{A_kB_kX_kR_kR} = \mathcal{N}_k(\omega_{R_{k-1}R}) : \text{ for a state } \omega_{R_{k-1}R} \text{ such that } \nu_{X_k} = q\} \tag{2.56}$$

as the states that are compatible with the output distribution q on register X_k when input to the channel $\mathcal{N}_k$.

We use this notation to define a min tradeoff function, which will be used to bound the min-entropy in the theorem.

Definition 2.29. *A function $f : \mathbb{P} \to \mathbb{R}$ is called a min-tradeoff function for the channels $\{\mathcal{N}_k\}_{k=1}^n$ if for every $k \in [n]$, it satisfies*

$$f(q) \leq \inf_{\nu \in \Sigma_k(q|\mathcal{N}_k)} H(A_k|B_kR)_\nu. \tag{2.57}$$

For a string $x_1^n \in \mathcal{X}^n$, we define its frequency distribution $\text{freq}(x_1^n) \in \mathbb{P}$ as

$$\text{freq}(x_1^n)(a) := \frac{|\{i \in [n] : x_i = a\}|}{n}. \tag{2.58}$$

With the above definitions and notation in hand we can state the entropy accumulation theorem in its full generality.

Theorem 2.30 ([DFR20, Theorem 4.4]). *Let the channels $\{\mathcal{M}_i\}_{i=1}^n$ be as described above and the state $\rho_{A_1^nB_1^nX_1^nE} := \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E})$ be produced through a sequential process similar to Fig. 2.1. Suppose, ρ satisfies the Markov chain conditions $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$ for every $k \in [n]$. Let f be an affine min-tradeoff function for all the channels $\{\mathcal{M}_i\}_{i=1}^n$ and let $0 < \epsilon < 1$. Then, for any event $\Omega \subseteq \mathcal{X}^n$ such that for every $x_1^n \in \Omega$, $f(\text{freq}(x_1^n)) \geq h$, we have*

$$H^\epsilon_{\min}(A_1^n|B_1^nE)_{\rho|\Omega} \geq nh - c\sqrt{n}$$

for $c = 2\left(\log(1+2|A|) + \|\nabla f\|_\infty + 1\right)\sqrt{1 - 2\log(\epsilon \Pr_\rho(\Omega))}$*, where* $|A|$ *is the maximum dimension of the systems* A_i *and* $\Pr_\rho(\Omega)$ *is the probability of the event* Ω *for the state* ρ*.*

The entropy accumulation theorem is fairly general and has found many applications in cryptography. It has been applied to prove the security of sequential DI-QKD [AFDF+18], randomness expansion [AFRV19], to prove bounds on quantum random access codes [DFR20], and also to prove bounds on Lyapunov exponents [SFR21].

Using recently proven chain rules [FF21] for *channel* sandwiched Rényi relative entropies, [MFSR24] have been able to relax the Markov conditions required in the theorem above. Their *generalized* entropy accumulation theorem allows for the adversary's registers to also evolve in every round. The only condition, imposed on the channels, $\mathcal{M}_i$, is that they satisfy a kind of *non-signalling* property for channels, which requires that information is not leaked by the honest parties devices to the adversary. This generalisation can be applied to scenarios like blind randomness generation [MS17, FM18, MFSR24, MR22], where one cannot apply the original entropy accumulation theorem because the Markov condition is not satisfied.

2.7. Quantum key distribution (QKD)

Key distribution is a primitive in cryptography, which allows two honest parties, Alice and Bob to share a secret key in the presence of an adversary, Eve. Classically, key distribution cannot be accomplished without some assumption on the capabilities of the adversary. Typically, it is assumed that a certain problem is computationally "hard" even for the adversary. However, quantum mechanics allows two parties to implement key distribution unconditionally [BB84, Ben92, SP00]. These *quantum key distribution* (QKD) protocols rely on the fact that it is not possible to measure an unknown quantum state without disturbing it.

We will discuss (a variant of) the BB84 protocol in detail in order to understand QKD. It is presented as Protocol 2.1. For the protocol, it is assumed that Alice and Bob have access to a classical authenticated channel and a (insecure) quantum channel. In particular, Eve can access and arbitrarily modify the states sent over the quantum channel, but she cannot change the contents of the classical authenticated channel. She can only read these contents.

BB84 QKD protocol[a]

Parameters:

- n is the number of qubits sent
- $\mu \in (0,1)$ is the probability of measurement in the X basis $\{|+\rangle, |-\rangle\}$.
- $e \in (0, \frac{1}{2})$ is the maximum error tolerated

Protocol:

(1) For every $1 \leq i \leq n$ perform the following steps:
   (a) Alice chooses a random bit b_i and with probability $1-\mu$ encodes it in the Z basis and with probability μ in the X basis.
   (b) Alice sends her encoded qubit to Bob.
   (c) Bob measures the qubit in the Z basis with probability $1-\mu$ and X basis with probability μ. He records the output as b'_i.

(2) **Sifting:** Alice and Bob share their choice of bases for all the rounds and discard the rounds where their choices are different. We use $\mathcal{X}$ ($\mathcal{Z}$) to denote the set of indices where Alice encoded the bit in basis X (Z) and Bob measured it in the X (Z) basis.

(3) **Parameter estimation:** Alice and Bob announce their string $b_{\mathcal{X}}$ and $b'_{\mathcal{X}}$ (b_i's for the set $\mathcal{X}$) and compute $\hat{e} = \frac{1}{|\mathcal{X}|}\sum_{i\in\mathcal{X}} b_i \oplus b'_i$. They abort if $\hat{e} > e$.

(4) **Information reconciliation:** Alice and Bob use an information reconciliation procedure, which lets Bob obtain a guess for Alice's raw key $\hat{b}_{\mathcal{Z}}$ (if the information reconciliation protocol succeeds $\hat{b}_{\mathcal{Z}} = b_{\mathcal{Z}}$).

(5) **Raw key validation:** Alice selects a random hash function from a 2-universal family and sends it along with $\text{hash}(b_{\mathcal{Z}})$ to Bob. If $\text{hash}(b_{\mathcal{Z}}) \neq \text{hash}(\hat{b}_{\mathcal{Z}})$ Bob aborts the protocol.

(6) **Privacy amplification:** Alice and Bob use a privacy amplification protocol to create a secret key.

[a]Strictly speaking, this is a modern variant of the BB84 protocol. However, we will refer to it as the BB84 protocol for this thesis.

Protocol 2.1

The protocol uses classical bits $\{0,1\}$ encoded in the X basis (as the states $\{|+\rangle, |-\rangle\}$) and Z basis (as the states $\{|0\rangle, |1\rangle\}$) (also see [Wie83]). In a single step of the protocol, Alice randomly chooses a bit b_i and encodes it randomly in the X or Z basis. She then sends her qubit to Bob, who randomly measures it in the X or Z basis. If Bob measures the qubit in the same basis in which Alice encoded it, then he would also measure the bit b_i assuming no

adversary or noise interferes with the protocol. This step is repeated multiple times during the protocol. Then Alice and Bob publicly share their bases choices and discard the rounds where Bob's measurement basis do not match Alice's encoding basis. They use the rounds, where Alice encoded using the X basis to estimate the noise during the protocol. If the noise is higher than a certain threshold, they abort. In order for Eve to gain information about Alice's raw key $b_{\mathcal{Z}}$, Eve has to interact with the qubits sent by Alice. The idea for proving security for this QKD protocol is that if Eve interferes too much with Alice's qubits, then the noise measured by Bob is high. Otherwise, the amount of information gained by Eve is small enough that Alice and Bob can extract a secret key. Alice and Bob perform information reconciliation and key validation after parameter estimation to make sure that they hold the same strings. At this point, Eve may have partial information about these strings. So, in the final stage of the protocol, Alice and Bob use privacy amplification with appropriate parameters to make the keys independent of Eve's state. We describe this classical post-processing in more detail in Sec. 2.7.2.

2.7.1. Security definition for QKD

In a QKD protocol, Alice and Bob are honest parties and Eve is considered to be an adversarial eavesdropper. Alice and Bob have access to a classical public authenticated channel. In addition, they have access to an unsecure quantum channel. As mentioned above, the attack model we consider for QKD is that Eve can arbitrarily modify the messages sent over the quantum channel. At the end of the protocol, Alice receives the key K_A and Bob the key K_B and suppose that Eve holds the register E. All three parties also receive an abort flag.

Definition 2.31 (Security of QKD [PR14])**.** *A QKD protocol is said to be* (ϵ_c, ϵ_s)*-secure if for every strategy of the adversary, Eve, the following two properties hold:*

(1) Correctness*: Alice and Bob's keys are equal with a high probability*[(1)]

$$\Pr[K_A \neq K_B] \leq \epsilon_c. \tag{2.59}$$

(2) Secret*: Alice's key is almost uniform and independent of Eve's registers*

$$\frac{\Pr(\neg abort)}{2} \left\| \rho_{K_A E|\neg abort} - \frac{1}{|K_A|} \sum_k |k\rangle\langle k| \otimes \rho_{E|\neg abort} \right\|_1 \leq \epsilon_s. \tag{2.60}$$

We note that a QKD protocol which is secure according to the definition above can be securely composed with other protocols [PR14].

[(1)] Note that we can assume that Alice and Bob trivially output the same key when the protocol aborts.

2.7.2. Classical post-processing and QKD proof requirements

Information reconciliation, raw key validation, and privacy amplification are used to process the raw keys created by Alice and Bob's measurement outcomes in QKD protocols like Protocol 2.1. We provide a brief description of these protocols and how they are used in QKD. For more details, we refer the reader to [**Ren06**, Chapter 6] and [**AF20**, Section 4.2.2].

An information reconciliation (IR) protocol [**BS94**] allows two parties holding correlated strings to derive a common string, while minimising communication. Raw key validation simply uses a random hash function from a 2-universal family to verify that Alice and Bob hold the same raw key. Raw key validation is often included in the information reconciliation protocol itself. We describe it as a separate step here, since it makes correctness evident and aids our explanation.

[**Ren06**, Lemma 6.3.4] demonstrates a one-way protocol from Alice to Bob for which the number of bits communicated during information reconciliation and raw key validation, which we will simply refer to as leak_{IR}, can be bounded as

$$\text{leak}_{\text{IR}} \leq H^{\epsilon}_{\max}(X|Y) + O(1) \tag{2.61}$$

where X and Y are Alice and Bob's raw keys respectively, and the parameter $\epsilon \in [0,1]$.

In QKD protocols, information reconciliation ensures correctness. For example, in Protocol 2.1, adversarial interference and noise may cause the raw keys generated by Alice and Bob to be unequal. Through information reconciliation, Bob can adjust his raw key to match Alice's.

Privacy amplification, described in Sec. 2.4, allows Alice and Bob to extract a random secret key from their shared raw key obtained through information reconciliation. The process involves selecting a random 2-universal hash function and applying it to their raw key. When parameters are appropriately chosen, Lemma 2.25 guarantees that the resulting key is independent of the adversary's state.

Let's suppose that Alice and Bob's raw keys before applying classical post-processing are A_1^n and B_1^n, and Eve's state is E. If Alice sends message C to Bob during information reconciliation and raw key validation, then according to Lemma 2.25, the QKD protocol's key length is lower bounded (up to a constant) by:

$$H^{\epsilon}_{\min}(A_1^n|EC)_{\rho} \geq H^{\epsilon}_{\min}(A_1^n|E)_{\rho} - \text{leak}_{\text{IR}} \tag{2.62}$$

using the dimension bound [Tom16, Lemma 6.8]. The information reconciliation cost can be bounded as

$$\text{leak}_{\text{IR}} \leq H_{\max}^{\epsilon'}(A_1^n|B_1^n)_{\rho_{\text{honest}}} + O(1). \tag{2.63}$$

$H_{\max}^{\epsilon'}$ in the bound above is evaluated on ρ_{honest}, which represents the state produced at the end of the QKD rounds in the absence of an adversary. In the presence of an adversary, this amount of communication may not suffice for successful information reconciliation, and in that case the raw key validation step would fail with high probability (see [Ren06, Section 6.3] for a detailed discussion), which is permissible according to the security definition.

This term is typically straightforward to bound as the form of ρ_{honest} is known. It can usually be expressed as $nf(e)$, where e quantifies the protocol noise and f is a small function such that $f(e) \to 0$ as $e \to 0$. Therefore, while proving security for QKD, the main challenge lies in establishing a linear lower bound for $H_{\min}^{\epsilon}(A_1^n|E)_\rho$ in cases where the protocol does not abort.

If one is able to prove that

$$H_{\min}^{\epsilon}(A_1^n|E)_\rho - H_{\max}^{\epsilon'}(A_1^n|B_1^n)_{\rho_{\text{honest}}} \geq rn - O(1) \tag{2.64}$$

for some $r > 0$, then Alice and Bob can produce a secure key of length $(rn - O(1))$ according to Lemma 2.25. The *key rate* for a QKD protocol is defined as its key length divided by the number of rounds. In this case, it asymptotically tends to r.

2.8. Device-independent quantum key distribution (DIQKD)

The security of QKD protocols, like BB84, can be mathematically proven. However, practical implementations of these QKD protocols are vulnerable to side-channel attacks, where Eve exploits imperfections in the preparation and detection devices to extract additional information [LWW+10, SRK+15]. For example, Eve could send a laser signal towards Alice's preparation equipment, and analyse its reflection to determine the preparation bases, potentially compromising the protocol's security. These attacks stem from the impossibility of completely and accurately modelling Alice and Bob's devices. To circumvent such problems, *device-independent QKD (DIQKD)* [MY98, MY04, BHK05, PAB+09] (also see [Eke91]) has been developed. DIQKD protocols allow Alice and Bob to use *untrusted* devices (or equivalently, devices prepared by the eavesdropper) to implement key distribution. The security of such protocols relies solely on the properties of quantum

mechanics.

Typically, in such protocols, Alice and Bob play multiple non-local games using some shared quantum state. A non-local game is defined as:

Definition 2.32 (Non-local game)**.** *A non-local game* $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \Pi_{XY}, V)$ *is a game played between two cooperating parties called Alice and Bob, who are not allowed to communicate while playing the game. Questions* $x \in \mathcal{X}$ *and* $y \in \mathcal{Y}$ *are sampled according to the distribution* Π_{XY} *and distributed to Alice and Bob respectively. Alice and Bob reply with answers* $a \in \mathcal{A}$ *and* $b \in \mathcal{B}$ *using some predetermined strategy. Their strategy can be classical, in which case their answers depend only on their questions and some shared randomness, or it can be quantum, in which case they measure some pre-shared quantum state to determine their answers. Alice and Bob win the game if they satisfy the predicate* $V(x, y, a, b)$*. The maximum winning probability for classical strategies is denoted using* $\omega_c(G)$ *and for quantum strategies using* $\omega_q(G)$*.*

One of the most prominent examples of a non-local game is the CHSH game [**CHSH69**]. For the CHSH game, we have $\mathcal{X} = \mathcal{Y} = \mathcal{A} = \mathcal{B} = \{0,1\}$, Π_{XY} is the uniform distribution on all possible questions and the predicate $V(x,y,a,b) = \neg[a \oplus b \oplus (x \wedge y)]$. For this game, $\omega_c(G) = \frac{3}{4}$ and $\omega_q(G) = \frac{2+\sqrt{2}}{4}$. The ideal quantum strategy uses a Bell state between Alice and Bob, and projective measurements for Alice and Bob (see for example [**Wat18**, Example 6.59]). This strategy is also unique up to isometries [**SW88**, **MYS12**].

The existence of non-local games which one can win with a higher probability using quantum mechanics than with classical strategies is one of the most celebrated results in quantum information. It provides a means to test the quantumness of our reality. Moreover, if two parties sharing a quantum state are able to win such a game with a probability exceeding the classical winning probability, then one can place bounds on the amount of information any third party has about their answers to the game, irrespective of the imperfections in the measurement devices used during the game [**MAG06**, **MS17**, **BFF21**]. This fact is used to ensure the security of DIQKD protocols.

During a DIQKD protocol, Alice and Bob treat their measurement devices and quantum states as black boxes that produce outputs corresponding to the inputs of the non-local game. If the games for a DIQKD protocol are played sequentially, that is the inputs for the $(k+1)^{\text{th}}$ game are provided to the devices after the answers for the k^{th} game are provided by the device, then we call the protocol *sequential*. It should be noted that for

these protocols, in general, after each round of the protocol, Alice and Bob's questions are revealed to Eve and the state Eve chooses for the next round can depend on these questions too. In particular, the answers for the k^{th} round of the game depend only on the questions in the previous rounds. Sequential protocols are a good model for DIQKD protocols, which distribute 'entanglement on the fly'. As we will see these DIQKD protocols are well understood and their rates are close to optimal.

On the other hand, if the protocol assumes that Alice and Bob submit all questions to their devices at the same time and then the device uses all of them to produce the answers, the protocol is termed *parallel*. In a parallel protocol, the answers for the k^{th} round of the game can depend on the questions for all the rounds. This makes their analysis much more difficult.

In the following, we will discuss the security definition for DIQKD. We also present a sequential DIQKD protocol and sketch its proof of security using entropy accumulation. These largely follow [**AF20**]. We present a parallel DIQKD protocol along with a proof of security in Chapter 6.

2.8.1. Security definition for DIQKD

The security definition for DIQKD is the same as that for QKD (also see [**AF20**, Section 11.3]). The key distinction lies in the attack model considered for the adversary, Eve. In standard QKD, the quantum devices used by Alice and Bob in their laboratories are assumed to be honest and behave exactly as specified. In contrast, DIQKD assumes these quantum devices are untrusted. Concretely speaking, the adversary is allowed to choose both the quantum devices used by Alice and Bob and their shared quantum states.

On the other hand, similar to QKD, Alice and Bob both have access to trusted classical devices for post-processing and generating randomness, as well as an authenticated classical channel. Finally, we assume that Alice and Bob can prevent information from leaking from their laboratories to Eve (this assumption is necessary, as otherwise Eve could simply obtain Alice's key even after successful protocol execution; also see [**AF20**, Section 4.2.5]).

2.8.2. Protocol for sequential DIQKD

The sequential DIQKD protocol used in [**AF20**] is presented as Protocol 2.2. For this protocol, Bob uses a quantum device, which can take questions from {0,1,2} and produce corresponding answers. In the absence of an adversary, Alice and Bob share a Bell state.

Sequential DIQKD protocol

Parameters:

- $n \geq 1$ is the number of rounds in the protocol
- $\gamma \in (0,1]$ is the fraction of test rounds
- ω_{exp} is the winning threshold for the CHSH games

Protocol:

(1) For every $1 \leq i \leq n$ perform the following steps:
   (a) Alice sends one half of a Bell state to Bob.
   (b) Alice chooses a random $T_i \in \{0,1\}$ such that $\Pr[T_i = 1] = \gamma$.
   (c) Alice sends T_i to Bob.
   (d) If $T_i = 0$, Alice and Bob set the questions $(X_i,Y_i) = (0,2)$, otherwise they sample (X_i,Y_i) uniformly at random from $\{0,1\}$.
   (e) Alice and Bob use their device with the questions (X_i,Y_i) and obtain the outputs A_i, B_i.

(2) Alice announces her questions X_1^n to Bob.

(3) **Information reconciliation:** Alice and Bob use an information reconciliation protocol, which lets Bob obtain the raw key $\tilde{A}_1^n$ (if the information reconciliation protocol succeeds $A_1^n = \tilde{A}_1^n$). In case the information reconciliation protocol aborts, they abort the QKD protocol too.

(4) **Parameter estimation:** Bob uses B_1^n and $\tilde{A}_1^n$ to compute the average winning probability ω_{avg} on the test rounds. He aborts if $\omega_{\text{avg}} < \omega_{\text{exp}}$.

(5) **Privacy amplification:** Alice and Bob use a privacy amplification protocol to create a secret key K from A_1^n (using $\tilde{A}_1^n$ for Bob).

Protocol 2.2

They use the measurements for the optimal strategy for the CHSH game when given questions in $\{0,1\}$. For the question $y = 2$, Bob's device should measure the Bell state using the same measurement it uses for Alice's question $x = 0$, so that Alice and Bob's measurements are perfectly correlated for the question $(x,y) = (0,2)$. The rounds with this question correspond to *generation rounds*, that is, rounds which are used to create the raw key. The rest of the rounds are used for parameter estimation and are called *test rounds*. It is noteworthy that Alice's device cannot differentiate between a test round and a generation round.

2.8.3. Proof sketch for sequential DIQKD

Security against i.i.d. attacks

Under the i.i.d. assumption (also called the collective attack case), Eve distributes the same state for each round of the protocol and chooses the devices such that they use the same measurement in each round too. This is a highly restrictive model, but it is usually the first step towards proving a general security proof. The security of DIQKD under this assumption was proved by [PAB+09]. Their main result was a bound on the von Neumann entropy of Alice's answers given Eve's information for a single round of the CHSH game. We state this result in the form used by [AF20] here.

Lemma 2.33 ([AF20, Lemma 5.3]). *Suppose that the honest players, Alice and Bob play the CHSH game with quantum devices provided by the adversary, Eve. Let register E be Eve's part of the initial quantum state, X and Y be Alice and Bob's questions for the game, and A and B Alice and Bob answers. If the quantum devices provided by Eve win the game with a probability $\omega \in \left[\frac{3}{4}, \frac{2+\sqrt{2}}{4}\right]$, we have*

$$H(A|XYE) \geq \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16\omega(\omega-1)+3}\right) \tag{2.65}$$

where $h(.)$ represents the binary entropy function

Under i.i.d. attacks, it is straightforward to see that the parameter estimation step in Protocol 2.2 allows Alice and Bob to ensure that the state provided to them by Eve wins the CHSH game with a high probability. Then, security follows simply by using the above bound along with the equipartition property (Eq. 2.46) for the smooth min-entropy. The intuition for the security proof against general attacks is similar. However, Alice and Bob can no longer meaningfully estimate a winning probability for the states distributed between them, since these could all be different and even be correlated across rounds.

Security against general attacks

In this section, we briefly describe how entropy accumulation (Theorem 2.30) can be used to prove the smooth min-entropy lower bound required for the security of sequential DIQKD in the presence of an adversary. We use the variables defined in Protocol 2.2 and follow the proof given in [AFDF+18].

To establish the security of Protocol 2.2 and show it generates a key of length $\Omega(n)$, we need only prove that the entropy $H^{\epsilon}_{\min}(A_1^n|EX_1^nY_1^nT_1^n)$ is lower bounded by $\Omega(n)$. The

information leaked during information reconciliation protocol is small and can be addressed as described in Sec. 2.7.2.

For brevity, we consider a simpler case of sequential DIQKD, where Eve distributes the states and devices to Alice and Bob for all the games at the beginning of the protocol. The state ρ_{R_0E} in the entropy accumulation (EAT) setting for this protocol is the initial quantum state shared between Alice, Bob and Eve. Here the register E contains information about the initial state kept by Eve. R_0 can be decomposed as two registers E_A and E_B, where E_A is held by Alice and E_B by Bob. The channels $\mathcal{M}_i$ in the EAT setting correspond to one round of Step 1 of Protocol 2.2. Therefore, the channel $\mathcal{M}_i$ samples T_i, X_i and Y_i and then uses Alice and Bob's devices to produce the answers A_i and B_i. In the language of Theorem 2.30, we set $A_i \leftarrow A_iB_i$, $B_i \leftarrow T_iX_iY_i$, $E \leftarrow E$ and $X_i \leftarrow V(X_i, Y_i, A_i, B_i)$, where V is the winning predicate for the CHSH game. We condition the output state on the protocol not aborting. This determines the event Ω in Theorem 2.30.

For a binary probability distribution, q and the channel $\mathcal{M}_i$, the set $\Sigma_i(q|\mathcal{M}_i)$ consists of the set of input states for CHSH strategy used by the channel $\mathcal{M}_i$, which win the game with probability $q(1)$. Thus, using the single round bound in Lemma 2.33, we have

$$\inf_{\nu\in\Sigma_i(q|\mathcal{M}_i)} H(A_iB_i|X_iY_iT_i\tilde{R}_{i-1}) \geq \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16q(1)(q(1)-1)+3}\right)$$

as long as $q(1) > 3/4$. This bound can be used to derive a min-tradeoff function for EAT. Since, the protocol aborts when the winning probability on the test rounds is less than ω_{exp}, we can restrict ourselves to distributions q such that $q(1) \geq \omega_{\text{exp}}$, and it can be shown that[(2)]

$$h := \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16\omega_{\text{exp}}(\omega_{\text{exp}}-1)+3}\right) \tag{2.66}$$

can be used to lower bound the min-tradeoff function for EAT. Theorem 2.30 then shows that

$$H^{\epsilon}_{\min}(A_1^nB_1^n|EX_1^nY_1^nT_1^n) \geq nh - O(\sqrt{n}).$$

We can then use a simple chain rule and the fact that A_1^n and B_1^n differ only on a small set of positions to derive a similar bound for $H^{\epsilon}_{\min}(A_1^n|EX_1^nY_1^nT_1^n)$.

[(2)] We overload the symbol h here, since it is also used to denote the binary entropy function. Its meaning, however, should be clear from context.

Chapter 3

Smooth min-entropy lower bounds for approximation chains

3.1. Introduction

In this chapter, we consider scenarios consisting of approximation chains of a state along with additional conditions, and prove smooth min-entropy lower bounds for these. The techniques developed here form the foundation for analysing approximation chains in the remainder of the thesis.

We begin by considering the scenario of *approximately independent registers*, that is, a state $\rho_{A_1^n B}$, which for every $1 \leq k \leq n$ satisfies

$$\left\| \rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1} B} \right\|_1 \leq \epsilon. \tag{3.1}$$

for some small $\epsilon > 0$ and arbitrarily large n (in particular $n \gg \frac{1}{\epsilon}$). That is, for every k, the system A_k is almost independent of the system B and everything else which came before it. For simplicity, let us further assume that for all k the state $\rho_{A_k} = \rho_{A_1}$. Intuitively, one expects that the smooth min-entropy (with the smoothing parameter depending on ϵ and not on n)[(1)] for such a state will be large and close to $\approx n(H(A_1) - g'(\epsilon))$ (for some small function $g'(\epsilon)$). However, it is not possible to prove this result using techniques, which rely only on the triangle inequality and smoothing. The triangle inequality, in general, can only be used to bound the trace distance between $\rho_{A_1^n B}$ and $\otimes_{k=1}^n \rho_{A_k} \otimes \rho_B$ by $n\epsilon$, which will result in a trivial bound when $n \gg \frac{1}{\epsilon}$ [(2)]. Instead, we show that a bound on the entropic distance

[(1)] The smoothing parameter must depend on ϵ in such a scenario. This can be seen by considering the probability distribution $P_{A_1^n B}$ such that B is 0 with probability ϵ and 1 otherwise and A_1^n is a random n-bit string if $B = 1$ and constant if $B = 0$.

[(2)] Consider the distribution $Q_{A_1^{2n} B_1^{2n}}$, where for every $i \in [2n]$, the bit B_i is chosen independently and is equal to 0 with probability ϵ and is 1 otherwise. The bit A_i is chosen randomly if $B_i = 1$, otherwise it is

given by the smooth max-relative entropy between these two states can be used to prove a lower bound for the smooth min-entropy in this scenario.

While an upper bound of $n\epsilon$ is trivial and meaningless for the trace distance for large n, it is still a meaningful bound for the relative entropy between two states, which is unbounded in general. We can show that the above approximation conditions (Eq. 3.1) also imply that relative entropy distance between $\rho_{A_1^n B}$ and $\otimes_{k=1}^n \rho_{A_k} \otimes \rho_B$ is $nf(\epsilon)$ for some small function $f(\epsilon)$. The substate theorem [**JRS02**] allows us to transform this relative entropy bound into a smooth max-relative entropy bound. For two general states ρ_{AB} and η_{AB}, such that $d := D^{\delta}_{\max}(\rho_{AB}\|\eta_{AB})$, we can easily bound the smooth min-entropy of ρ in terms of the min-entropy of η by observing that

$$\rho_{AB} \approx_{\delta} \tilde{\rho}_{AB} \leq e^{d}\eta_{AB} \leq e^{-(H_{\min}(A|B)_{\eta}-d)}\, \mathbb{1}_A \otimes \sigma_B \tag{3.2}$$

for some state σ_B, which satisfies $D_{\max}(\eta_{AB}\|\, \mathbb{1}_A \otimes \sigma_B) = -H_{\min}(A|B)_{\eta}$. This implies that

$$H^{\delta}_{\min}(A|B)_{\rho} \geq H_{\min}(A|B)_{\eta} - D^{\delta}_{\max}(\rho_{AB}\|\eta_{AB}) \tag{3.3}$$

We call this an *entropic triangle inequality*, since it is based on the triangle inequality property of $D_{\max}$. We can further improve this smooth min-entropy triangle inequality to (Lemma 3.5)

$$H^{\epsilon+\delta}_{\min}(A|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\eta} - \frac{\alpha}{\alpha-1} D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) - \frac{g_1(\delta,\epsilon)}{\alpha-1} \tag{3.4}$$

for some function g_1, $\epsilon + \delta < 1$ and $1 < \alpha \leq 2$. Our general strategy for the scenarios considered in this thesis is to first bound the "one-shot information theoretic" distance (the smooth max-relative entropy distance) between the real state ρ ($\rho_{A_1^n B}$ in the above scenario) and a virtual, but *nicer* state, η ($\otimes_{k=1}^n \rho_{A_k} \otimes \rho_B$ above) by $nf(\epsilon)$ for some small $f(\epsilon)$. Then, we use Eq. 3.4 above to reduce the problem of bounding the smooth min-entropy on state ρ to that of bounding an α-Rényi entropy on the state η. Using this strategy, in Corollary 3.10, we prove that for states satisfying the approximately independent registers assumptions, we have for $\delta = O\left(\epsilon \log \frac{|A|}{\epsilon}\right)$ that

$$H^{\delta^{\frac{1}{4}}}_{\min}(A_1^n|B)_{\rho} \geq n\left(H(A_1)_{\rho} - O(\delta^{\frac{1}{4}})\right) - O\left(\frac{1}{\delta^{3/4}}\right). \tag{3.5}$$

Another scenario we consider here is that of approximate entropy accumulation. As discussed in Sec. 2.6, in the setting for entropy accumulation, a sequence of channels $\mathcal{M}_k : R_{k-1} \to A_k B_k R_k$ for $1 \leq k \leq n$ sequentially act on a state $\rho^{(0)}_{R_0 E}$ to produce the state $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E})$. It is assumed that the channels $\mathcal{M}_k$ are such that the Markov

chosen to be equal to A_{i-1}. In this case, Q_{A_k} is the uniformly random distribution for bits and Eq. 3.1 is satisfied. Let $I = |\{i \in [n] : A_{2i-1} = A_{2i}\}|$. Then, for $Q_{A_1^{2n} B_1^{2n}}$, this value concentrates around $\frac{n(1+\epsilon)}{2}$, whereas for $\prod_{i=1}^{2n} Q_{A_i} \cdot Q_{B_1^{2n}}$, it concentrates around $\frac{n}{2}$. This shows that $\left\| Q_{A_1^{2n} B_1^{2n}} - \prod_{i=1}^{2n} Q_{A_i} \cdot Q_{B_1^{2n}} \right\|_1 \to 2$.

chain $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$ is satisfied for every k. Under these conditions, the entropy accumulation theorem [DFR20], provides a tight lower bound for the smooth min-entropy $H_{\min}^{\delta}(A_1^n|B_1^nE)$. We consider an approximate version of this setting in this chapter where the channels $\mathcal{M}_k$ themselves do not necessarily satisfy the Markov chain condition, but they can be ϵ-approximated by a sequence of channels $\mathcal{M}'_k$, which satisfies certain Markov chain conditions. Such relaxations are important to understand the behaviour of cryptographic protocols, like device-independent quantum key distribution [AFDF+18], implemented with imperfect devices [JK23, Tan23]. Once again we can model this scenario as an approximation chain: for every $1 \leq k \leq n$, the state produced in the k^{th} step satisfies

$$\begin{aligned}\rho_{A_1^kB_1^kE} &= \mathrm{tr}_{R_k} \circ \mathcal{M}_k \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E})\right) \\ &\approx_\epsilon \mathrm{tr}_{R_k} \circ \mathcal{M}'_k \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E})\right) := \sigma^{(k)}_{A_1^kB_1^kE}.\end{aligned}$$

Moreover, the assumptions on the channel $\mathcal{M}'_k$ guarantee that the state $\sigma^{(k)}_{A_1^kB_1^kE}$ satisfies the Markov chain condition $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$. This implies that the chain rules and bounds used for entropy accumulation apply to this state too, and hence we can expect that the smooth min-entropy is large for it similar to the original setting.

To prove this, roughly speaking, we use the chain rules for divergences [FF21] to show that the divergence distance between the states $\rho_{A_1^nB_1^nE} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E})$ and the virtual state $\sigma_{A_1^nB_1^nE} = \mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(\rho^{(0)}_{R_0E})$ is relatively small, and then reduce the problem of lower bounding the smooth min-entropy of $\rho_{A_1^nB_1^nE}$ to that of lower bounding an α-Rényi entropy of $\sigma_{A_1^nB_1^nE}$, which can be done by using the chain rules developed for entropy accumulation[(3)]. In Theorem 3.12, we show the following smooth min-entropy lower bound for the state $\rho_{A_1^nB_1^nE}$ for sufficiently small ϵ and an arbitrary $\delta > 0$

$$H_{\min}^{\delta}(A_1^n|B_1^nE)_\rho \geq \sum_{k=1}^{n} \inf_{\omega} H(A_k|B_k\tilde{R}_{k-1})_{\mathcal{M}'_k(\omega)} - nO(\epsilon^{1/24}) - O\left(\frac{1}{\epsilon^{1/24}}\right) \tag{3.6}$$

where the infimum is over all possible input states $\omega_{R_{k-1}\tilde{R}_{k-1}}$ for reference register $\tilde{R}_{k-1}$ isomorphic to R_{k-1}, and the dimensions $|A|$ and $|B|$ are assumed constant while using the asymptotic notation.

[(3)] The channel divergence bounds we are able to prove are too weak for this idea to work as stated here. The actual proof is more complicated. However, this idea works in the classical case.

3.2. Entropic triangle inequality for the smooth min-entropy

In this section, we derive a simple entropic triangle inequality (Lemma 3.5) for the smooth min-entropy of the form in Eq. 3.4. This lemma is a direct consequence of the following triangle inequality for $\tilde{D}_\alpha$ (see [CMH17, Theorem 3.1] for a triangle inequality, which changes the second argument of $\tilde{D}_\alpha$).

Lemma 3.1. *Let ρ and η be subnormalised states and Q be a positive operator, then for $\alpha > 1$, we have*

$$\tilde{D}_\alpha(\rho\|Q) \leq \tilde{D}_\alpha(\eta\|Q) + \frac{\alpha}{\alpha-1} D_{\max}(\rho\|\eta) + \frac{1}{\alpha-1}\log\frac{\operatorname{tr}(\eta)}{\operatorname{tr}(\rho)}$$

and for $\alpha < 1$ if one of $\tilde{D}_\alpha(\eta\|Q)$ and $D_{\max}(\rho\|\eta)$ is finite (otherwise we cannot define their difference), we have

$$\tilde{D}_\alpha(\rho\|Q) \geq \tilde{D}_\alpha(\eta\|Q) - \frac{\alpha}{1-\alpha} D_{\max}(\rho\|\eta) - \frac{1}{1-\alpha}\log\frac{\operatorname{tr}(\eta)}{\operatorname{tr}(\rho)}.$$

Proof. If $D_{\max}(\rho\|\eta) = \infty$, then both statements are true trivially. Otherwise, we have that $\rho \leq e^{D_{\max}(\rho\|\eta)}\eta$ and also $\rho \ll \eta$. Now, if $\rho \not\ll Q$ then $\eta \not\ll Q$. Hence, for $\alpha > 1$ if $\tilde{D}_\alpha(\rho\|Q) = \infty$, then $\tilde{D}_\alpha(\eta\|Q) = \infty$, which means the lemma is also satisfied in this condition. For $\alpha < 1$, if $\tilde{D}_\alpha(\rho\|Q) = \infty$, then the lemma is also trivially satisfied. For the remaining cases we have,

$$\begin{aligned} e^{(\alpha-1)\tilde{D}_\alpha(\rho\|Q)} &= \frac{\operatorname{tr}\left(Q^{-\frac{\alpha-1}{2\alpha}}\rho Q^{-\frac{\alpha-1}{2\alpha}}\right)^\alpha}{\operatorname{tr}(\rho)} \\ &\leq \frac{\operatorname{tr}\left(Q^{-\frac{\alpha-1}{2\alpha}} e^{D_{\max}(\rho\|\eta)}\eta Q^{-\frac{\alpha-1}{2\alpha}}\right)^\alpha}{\operatorname{tr}(\rho)} \\ &= \frac{\operatorname{tr}(\eta)}{\operatorname{tr}(\rho)} e^{\alpha D_{\max}(\rho\|\eta)} e^{(\alpha-1)\tilde{D}_\alpha(\eta\|Q)} \end{aligned}$$

where we used the fact that $\operatorname{tr}(f(X))$ is monotone increasing if the function f is monotone increasing. Dividing by $(\alpha-1)$ now gives the result. □

We define smooth α-Rényi conditional entropy as follows to help us amplify the above inequality.

Definition 3.2 (ϵ-smooth α-Rényi conditional entropy)**.** *For $\alpha \in (1,\infty]$ and $\epsilon \in [0,1]$, we define the ϵ-smooth α-Rényi conditional entropy as*

$$\tilde{H}_\alpha^{\uparrow,\epsilon}(A|B)_\rho := \sup_{\tilde{\rho}_{AB}\in B_\epsilon(\rho_{AB})} \tilde{H}_\alpha^{\uparrow}(A|B)_{\tilde{\rho}}. \tag{3.7}$$

Lemma 3.3. *For $\alpha \in (1,\infty]$ and $\epsilon \in [0,1)$, and states ρ_{AB} and η_{AB} we have*

$$\tilde{H}^{\uparrow,\epsilon}_{\alpha}(A|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\eta} - \frac{\alpha}{\alpha-1} D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) - \frac{1}{\alpha-1}\log\frac{1}{1-\epsilon^2}.$$

Proof. Let $\tilde{\rho}_{AB} \in B_{\epsilon}(\rho_{AB})$ be a subnormalised state such that $D_{\max}(\tilde{\rho}_{AB}\|\eta_{AB}) = D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB})$. Using Lemma 3.1 for $\alpha > 1$, we have that for every state σ_B, we have

$$\tilde{D}_{\alpha}(\tilde{\rho}_{AB}\|\mathbb{1}_A \otimes \sigma_B) \leq \tilde{D}_{\alpha}(\eta_{AB}\|\mathbb{1}_A \otimes \sigma_B) + \frac{\alpha}{\alpha-1} D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) + \frac{1}{\alpha-1}\log\frac{1}{1-\epsilon^2} \tag{3.8}$$

where we used the fact that $\tilde{\rho}_{AB} \in B_{\epsilon}(\rho_{AB})$ which implies that $\mathrm{tr}(\tilde{\rho}_{AB}) \geq 1-\epsilon^2$. Since, the above bound is true for arbitrary states σ_B, we can multiply it by -1 and take the supremum to derive

$$\tilde{H}^{\uparrow}_{\alpha}(A|B)_{\tilde{\rho}} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\eta} - \frac{\alpha}{\alpha-1} D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) - \frac{1}{\alpha-1}\log\frac{1}{1-\epsilon^2}.$$

The desired bound follows by using the fact that $\tilde{H}^{\uparrow,\epsilon}_{\alpha}(A|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\tilde{\rho}}$. □

Lemma 3.4. *For a state ρ_{AB}, $\epsilon \in [0,1)$, and $\delta \in (0,1)$ such that $\epsilon + \delta < 1$ and $\alpha \in (1,2]$, we have*

$$H^{\epsilon+\delta}_{\min}(A|B)_{\rho} \geq \tilde{H}^{\uparrow,\epsilon}_{\alpha}(A|B)_{\rho} - \frac{g_0(\delta)}{\alpha-1}$$

where $g_0(x) := -\log(1-\sqrt{1-x^2})$.

Proof. First, note that

$$H^{\epsilon+\delta}_{\min}(A|B)_{\rho} \geq \sup_{\tilde{\rho}\in B_{\epsilon}(\rho_{AB})} H^{\delta}_{\min}(A|B)_{\tilde{\rho}}. \tag{3.9}$$

To prove this, consider a $\tilde{\rho}_{AB} \in B_{\epsilon}(\rho_{AB})$ and $\rho'_{AB} \in B_{\delta}(\tilde{\rho}_{AB})$ such that $H_{\min}(A|B)_{\rho'} = H^{\delta}_{\min}(A|B)_{\tilde{\rho}}$. Then, using the triangle inequality for the purified distance, we have

$$\begin{aligned} P(\rho_{AB},\rho'_{AB}) &\leq P(\rho_{AB},\tilde{\rho}_{AB}) + P(\tilde{\rho}_{AB},\rho'_{AB}) \\ &\leq \epsilon + \delta \end{aligned}$$

which implies that $H^{\epsilon+\delta}_{\min}(A|B)_{\rho} \geq H_{\min}(A|B)_{\rho'} = H^{\delta}_{\min}(A|B)_{\tilde{\rho}}$. Since, this is true for all $\tilde{\rho} \in B_{\epsilon}(\rho_{AB})$ the bound in Eq. 3.9 is true.

Using this, we have

$$\begin{aligned} H^{\epsilon+\delta}_{\min}(A|B)_{\rho} &\geq \sup_{\tilde{\rho}\in B_{\epsilon}(\rho_{AB})} H^{\delta}_{\min}(A|B)_{\tilde{\rho}} \\ &\geq \sup_{\tilde{\rho}\in B_{\epsilon}(\rho_{AB})} \left\{ \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\tilde{\rho}} - \frac{g_0(\delta)}{\alpha-1} \right\} \end{aligned}$$

$$= \tilde{H}^{\uparrow,\epsilon}_{\alpha}(A|B)_{\rho} - \frac{g_0(\delta)}{\alpha - 1}$$

where we have used Lemma 2.23[(4)] in the second step. □

We can combine these two lemmas to derive the following result.

Lemma 3.5. *For $\alpha \in (1,2]$, $\epsilon \in [0,1)$, and $\delta \in (0,1)$ such that $\epsilon + \delta < 1$ and two states ρ_{AB} and η_{AB}, we have*

$$H^{\epsilon+\delta}_{\min}(A|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\eta} - \frac{\alpha}{\alpha - 1} D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) - \frac{g_1(\delta,\epsilon)}{\alpha - 1} \tag{3.10}$$

where $g_1(x,y) := -\log(1 - \sqrt{1-x^2}) - \log(1-y^2)$.

Proof. We can combine Lemmas 3.3 and 3.4 as follows to derive the bound in the lemma:

$$\begin{aligned} H^{\epsilon+\delta}_{\min}(A|B)_{\rho} &\geq \tilde{H}^{\uparrow,\epsilon}_{\alpha}(A|B)_{\rho} - \frac{g_0(\delta)}{\alpha - 1} \\ &\geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\eta} - \frac{\alpha}{\alpha - 1} D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) - \frac{1}{\alpha - 1}\left(g_0(\delta) + \log \frac{1}{1-\epsilon^2}\right). \end{aligned}$$

□

We also note the simple triangle inequality proven in the Introduction (Eq. 3.2) here for use in the future.

Lemma 3.6. *For two states ρ_{AB} and η_{AB} and $\epsilon \in [0,1)$, we have*

$$H^{\epsilon}_{\min}(A|B)_{\rho} \geq H_{\min}(A|B)_{\eta} - D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}). \tag{3.11}$$

We can use the asymptotic equipartition theorem for the smooth max-relative entropy and smooth min-entropy (Eq. 2.45 and 2.46) to derive the following novel triangle inequality for the von Neumann conditional entropy. Although we do not use this inequality in this thesis, we believe it is interesting and may prove useful in the future.

Corollary 3.7. *For $\alpha \in (1,2]$ and states ρ_{AB} and η_{AB}, we have that*

$$H(A|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\eta} - \frac{\alpha}{\alpha - 1} D(\rho_{AB}\|\eta_{AB}). \tag{3.12}$$

[(4)] This lemma is also valid for subnormalised states as long as $\delta \in (0, \sqrt{2\operatorname{tr}(\tilde{\rho}) - \operatorname{tr}(\tilde{\rho})^2})$ according to [DFR20, Lemma B.4].

Proof. Using Lemma 3.5 with $\alpha \in (1,2]$, the states $\rho_{AB}^{\otimes n}$, and $\eta_{AB}^{\otimes n}$ and any $\epsilon > 0$ and $\delta > 0$ satisfying the conditions for the lemma, we get

$$\begin{aligned} H_{\min}^{\epsilon+\delta}(A_1^n|B_1^n)_{\rho^{\otimes n}} &\geq \tilde{H}_\alpha^\uparrow(A_1^n|B_1^n)_{\eta^{\otimes n}} - \frac{\alpha}{\alpha-1} D_{\max}^\epsilon(\rho_{AB}^{\otimes n}\|\eta_{AB}^{\otimes n}) - \frac{g_1(\delta,\epsilon)}{\alpha-1} \\ \Rightarrow \frac{1}{n} H_{\min}^{\epsilon+\delta}(A_1^n|B_1^n)_{\rho^{\otimes n}} &\geq \tilde{H}_\alpha^\uparrow(A|B)_\eta - \frac{\alpha}{\alpha-1}\frac{1}{n} D_{\max}^\epsilon(\rho_{AB}^{\otimes n}\|\eta_{AB}^{\otimes n}) - \frac{1}{n}\frac{g_1(\delta,\epsilon)}{\alpha-1}. \end{aligned}$$

Taking the limit of the above for $n \to \infty$, we get

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} H_{\min}^{\epsilon+\delta}(A_1^n|B_1^n)_{\rho^{\otimes n}} \geq \tilde{H}_\alpha^\uparrow(A|B)_\eta - \lim_{n\to\infty} \frac{\alpha}{\alpha-1}\frac{1}{n} D_{\max}^\epsilon(\rho_{AB}^{\otimes n}\|\eta_{AB}^{\otimes n}) - \lim_{n\to\infty} \frac{1}{n}\frac{g_1(\delta,\epsilon)}{\alpha-1} \\ \Rightarrow &H(A|B)_\rho \geq \tilde{H}_\alpha^\uparrow(A|B)_\eta - \frac{\alpha}{\alpha-1} D(\rho_{AB}\|\eta_{AB}) \end{aligned}$$

which proves the claim. □

3.3. Approximately independent registers

In this section, we introduce our technique for using the smooth min-entropy triangle inequality for analysing approximation chains by studying a state $\rho_{A_1^n B}$ such that for every $k \in [n]$

$$\left\| \rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1} B} \right\|_1 \leq \epsilon. \tag{3.13}$$

We assume that the registers A_k all have the same dimension equal to $|A|$. One should think of the registers A_k as the secret information produced during some protocol, which also provides the register B to an adversary. We would like to prove that $H_{\min}^{f(\epsilon)}(A_1^n|B)$ is large (lower bounded by $\Omega(n)$) under the above *approximate independence conditions* for some reasonably small function f of ϵ and close to $nH(A_1)$, if we assume the states ρ_{A_k} are identical.

Let us first examine the case where the state ρ above is classical. We use the standard notation for probability distributions to address elements of ρ, so that $\rho(a_1^n, b) := \langle a_1^n, b| \rho_{A_1^n B} |a_1^n, b\rangle$, where $|a_1^n, b\rangle$ is standard basis vector. To show that in this case the smooth min-entropy is high, we will show that the set where the conditional probability $\rho(a_1^n|b) := \frac{\rho(a_1^n b)}{\rho(b)}$ can be large, has a small probability using the Markov inequality. We will use the following lemma for this purpose.

Lemma 3.8. *Suppose p, q are probability distributions on $\mathcal{X}$ such that $\frac{1}{2}\|p-q\|_1 \leq \epsilon$, then $S \subseteq \mathcal{X}$ defined as $S := \{x \in \mathcal{X} : p(x) \leq (1+\epsilon^{1/2})q(x)\}$ is such that $q(S) \geq 1 - \epsilon^{1/2}$ and $p(S) \geq 1 - \epsilon^{1/2} - \epsilon$.*

Proof. Let $S^c := \mathcal{X} \setminus S$, where S is the set defined above. If $q(S^c) = 0$, the statement in the lemma is satisfied. If $q(S^c) > 0$, we have that

$$\begin{aligned}
\epsilon \geq \frac{1}{2}\left\|p - q\right\|_1 &= \max_{H \subseteq \mathcal{X}} |p(H) - q(H)| \\
&\geq q(S^c)\left|\frac{p(S^c)}{q(S^c)} - 1\right| \\
&\geq q(S^c)\left(\frac{p(S^c)}{q(S^c)} - 1\right) \\
&= q(S^c)\left(\frac{\sum_{x \in S^c} p(x)}{\sum_{x \in S^c} q(x)} - 1\right) \\
&\geq q(S^c)\left(\frac{\sum_{x \in S^c}(1 + \epsilon^{\frac{1}{2}})q(x)}{\sum_{x \in S^c} q(x)} - 1\right) \\
&= q(S^c)\epsilon^{\frac{1}{2}}
\end{aligned}$$

which implies that $q(S^c) \leq \epsilon^{\frac{1}{2}}$. Now, the statement in the lemma follows. □

We will also assume for the sake of simplicity that ρ_{A_k} are identical for all $k \in [n]$. Using the lemma above, for every $k \in [n]$, we know that the set

$$\begin{aligned}
B_k :&= \left\{(a_1^n, b) : \rho(a_1^k, b) > (1 + \sqrt{\epsilon})\rho(a_1^{k-1}, b)\rho(a_k)\right\} \\
&= \left\{(a_1^n, b) : \rho(a_k | a_1^{k-1}, b) > (1 + \sqrt{\epsilon})\rho(a_k)\right\}
\end{aligned}$$

satisfies $\Pr_\rho(B_k) \leq 2\sqrt{\epsilon}$. We can now define $L = \sum_{k=1}^n \chi_{B_k}$, which is a random variable that simply counts the number of bad sets B_k an element (a_1^n, b) belongs to. Using the Markov inequality, we have

$$\Pr_\rho\left[L > n\epsilon^{\frac{1}{4}}\right] \leq \frac{\mathbb{E}_\rho[L]}{n\epsilon^{\frac{1}{4}}} \leq 2\epsilon^{\frac{1}{4}}.$$

We can define the bad set $\mathcal{B} := \left\{(a_1^n, b) : L(a_1^n, b) > n\epsilon^{\frac{1}{4}}\right\}$. Using this, we can define the subnormalised distribution $\tilde{\rho}_{A_1^n B}$ as

$$\tilde{\rho}_{A_1^n B}(a_1^n, b) = \begin{cases} \rho_{A_1^n B}(a_1^n, b) & (a_1^n, b) \notin \mathcal{B} \\ 0 & \text{else} \end{cases}.$$

We have $P(\tilde{\rho}_{A_1^n B}, \rho_{A_1^n B}) \leq 2\epsilon^{1/8}$. Further, note that for every $(a_1^n, b) \notin \mathcal{B}$, we have

$$\begin{aligned}
\rho(a_1^n | b) &= \prod_{k=1}^n \rho(a_k | a_1^{k-1}, b) \\
&= \prod_{k:(a_1^n, b) \notin B_k} \rho(a_k | a_1^{k-1}, b) \prod_{k:(a_1^n, b) \in B_k} \rho(a_k | a_1^{k-1}, b) \\
&\leq (1 + \sqrt{\epsilon})^n \prod_{k:(a_1^n, b) \notin B_k} \rho_{A_k}(a_k)
\end{aligned}$$

$$\leq (1+\sqrt{\epsilon})^n e^{-n(1-\epsilon^{\frac{1}{4}})H_{\min}(A_1)}$$

where in the third line we have used the fact that if $(a_1^n, b) \notin B_k$, then $\rho(a_k|a_1^{k-1}b) \leq (1+\sqrt{\epsilon})\rho_{A_k}(a_k)$ and in the last line we have used the fact that for $(a_1^k, b) \notin \mathcal{B}$, we have $|\{k \in [n] : (a_1^n,b) \notin B_k\}| = n - L(a_1^n, b) \geq n(1-\epsilon^{\frac{1}{4}})$, that all the states ρ_{A_k} are identical and $e^{-H_{\min}(A_k)} = \max_{a_k} \rho_{A_k}(a_k)$. Note that we have essentially proven and used a $D_{\max}$ bound above. This proves the following lower bound for the smooth min-entropy of ρ

$$H_{\min}^{2\epsilon^{\frac{1}{8}}}(A_1^n|B) \geq n(1-\epsilon^{\frac{1}{4}})H_{\min}(A_1) - n\log(1+\sqrt{\epsilon}). \tag{3.14}$$

The right-hand side above can be improved to get the Shannon entropy H instead of the min-entropy $H_{\min}$. However, we will not pursue this here, since this bound is sufficient for the purpose of our discussion.

Although, we do not generalise this classical argument to the quantum case yet (we will do it in Sec. 5.4), it provides a great amount of insight into the approximately independent registers problem. Two important examples of distributions, which satisfy the approximate independence conditions above were mentioned in Footnotes (1) and (2) earlier. To create the first distribution, we flip a biased coin B, which is 0 with probability ϵ and 1 otherwise. If $B = 0$, then A_1^n is set to the constant all zero string otherwise it is sampled randomly and independently. For this distribution, once the bad event $(B = 0)$ is removed, the new distribution has a high min-entropy. On the other hand, for the second distribution, $Q_{A_1^{2n}B_1^{2n}}$, we have that the random bits B_i are chosen independently, with each being equal to 0 with probability ϵ and 1 otherwise. If the bit B_i is 0, then A_i is set equal to A_{i-1} otherwise it is sampled independently. In this case, there is no small probability (small as a function of ϵ) event, that one can simply remove, so that the distribution becomes i.i.d. However, we expect that with high probability the number of $B_i = 0$ is close to $2n\epsilon$. Given that the distribution samples all the other A_i independently, the smooth min-entropy for the distribution should be close to $2n(1-\epsilon)H(A_1)$. The above argument shows that any distribution satisfying the approximate independence conditions in Eq. 3.13 can be handled by combining the methods used for these two example distributions, that is, deleting the bad part of the distribution and recognising that the probability for every element in the rest of the space behaves independently on average.

The classical argument above is difficult to generalise to quantum states primarily because the quantum equivalents of Lemma 3.8 are not as nice and simple. Furthermore, quantum conditional probabilities themselves are also difficult to use. We will develop the tools to generalise this argument in Chapter 5. Luckily for this problem, the substate theorem

serves as the perfect tool for developing a smooth max-relative entropy bound, which can then be used with the min-entropy triangle inequality. The quantum substate theorem (Theorem 2.26) provides an upper bound on the smooth max relative entropy $D^{\epsilon}_{\max}(\rho\|\sigma)$ between two states in terms of their relative entropy $D(\rho\|\sigma)$:

$$D^{\sqrt{\epsilon}}_{\max}(\rho\|\sigma) \leq \frac{D(\rho\|\sigma)+1}{\epsilon} + \log\frac{1}{1-\epsilon}. \tag{3.15}$$

In this section, we will also frequently use the multipartite mutual information [**Wat60**, **Hor94**, **CMS02**]. For a state $\rho_{X_1^n}$, the multipartite mutual information between the registers $(X_1, X_2, \cdots, X_n)$ is defined as

$$I(X_1 : X_2 : \cdots : X_n)_\rho := D(\rho_{X_1^n}\|\rho_{X_1}\otimes\rho_{X_2}\otimes\cdots\otimes\rho_{X_n}). \tag{3.16}$$

In other words, it is the relative entropy between $\rho_{X_1^n}$ and $\rho_{X_1}\otimes\rho_{X_2}\otimes\cdots\otimes\rho_{X_n}$. It can easily be shown that the multipartite mutual information satisfies the following identities:

$$I(X_1 : X_2 : \cdots : X_n)_\rho = \sum_{k=1}^{n} H(X_k)_\rho - H(X_1\cdots X_n)_\rho \tag{3.17}$$

$$= \sum_{k=2}^{n} I(X_k : X_1^{k-1}). \tag{3.18}$$

Going back to proving a bound for the quantum approximately independent registers problem, note that using the Alicki-Fannes-Winter (AFW) bound [**AF04**, **Win16**] for mutual information [**Wil13**, Theorem 11.10.4], Eq. 3.13 implies that for every $k \in [n]$

$$I(A_k : A_1^{k-1}B)_\rho \leq \epsilon\log|A| + g_2\left(\frac{\epsilon}{2}\right) \tag{3.19}$$

where $g_2(x) := (x+1)\log(x+1) - x\log(x)$. With this in mind, we can now focus our efforts on proving the following theorem.

Theorem 3.9. *Let registers A_k have dimension $|A|$ for all $k \in [n]$. Suppose a quantum state $\rho_{A_1^nB}$ is such that for every $k \in [n]$, we have*

$$I(A_k : A_1^{k-1}B)_\rho \leq \epsilon \tag{3.20}$$

for some $0 < \epsilon < 1$. Then, we have that

$$H^{\epsilon^{\frac{1}{4}}+\epsilon}_{\min}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H(A_k)_\rho - 3n\epsilon^{\frac{1}{4}}\log(1+2|A|) - \frac{2\log(1+2|A|)}{\epsilon^{3/4}} - \frac{2\log(1+2|A|)}{\epsilon^{1/4}}\left(\log(1-\sqrt{\epsilon}) + g_1(\epsilon, \epsilon^{\frac{1}{4}})\right) \tag{3.21}$$

where $g_1(x,y) := -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$. In particular, when all the states ρ_{A_k} are identical, we have

$$H^{\epsilon^{\frac{1}{4}}+\epsilon}_{\min}(A_1^n|B)_\rho \geq n\left(H(A_1)_\rho - 3\epsilon^{\frac{1}{4}}\log(1+2|A|)\right)$$

$$- \frac{2\log(1+2|A|)}{\epsilon^{3/4}} - \frac{2\log(1+2|A|)}{\epsilon^{1/4}} \left(\log(1-\sqrt{\epsilon}) + g_1(\epsilon, \epsilon^{\frac{1}{4}})\right). \tag{3.22}$$

Proof. First note that we have,

$$\begin{aligned} I(A_1 : A_2 : \cdots : A_n : B) &= D(\rho_{A_1^n B} \| \bigotimes_{k=1}^{n} \rho_{A_k} \otimes \rho_B) \\ &= \sum_{k=1}^{n} I(A_k : A_1^{k-1} B) \\ &\leq n\epsilon. \end{aligned}$$

Using the substate theorem, we now have

$$\begin{aligned} D_{\max}^{\epsilon^{\frac{1}{4}}} \left(\rho_{A_1^n B} \middle\| \bigotimes_{k=1}^{n} \rho_{A_k} \otimes \rho_B\right) &\leq \frac{D(\rho_{A_1^n B} \| \bigotimes_{k=1}^{n} \rho_{A_k} \otimes \rho_B) + 1}{\sqrt{\epsilon}} - \log(1-\sqrt{\epsilon}) \\ &\leq n\sqrt{\epsilon} + \frac{1}{\sqrt{\epsilon}} - \log(1-\sqrt{\epsilon}). \end{aligned} \tag{3.23}$$

We now define the auxiliary state $\eta_{A_1^n B} := \bigotimes_{k=1}^{n} \rho_{A_k} \otimes \rho_B$. Using Lemma 3.5, for $\alpha \in (1,2)$, we can transform the smooth min-entropy into an α-Rényi entropy on the auxiliary product state $\eta_{A_1^n B}$ as follows:

$$\begin{aligned} & H_{\min}^{\epsilon^{\frac{1}{4}}+\epsilon}(A_1^n|B)_\rho \\ &\geq \tilde{H}_\alpha^\uparrow(A_1^n|B)_\eta - \frac{\alpha}{\alpha-1} D_{\max}^{\epsilon^{\frac{1}{4}}}(\rho_{A_1^n B} \| \eta_{A_1^n B}) - \frac{g_1(\epsilon, \epsilon^{\frac{1}{4}})}{\alpha-1} \\ &= \sum_{k=1}^{n} \tilde{H}_\alpha^\uparrow(A_k)_\rho - \frac{\alpha}{\alpha-1} D_{\max}^{\epsilon^{\frac{1}{4}}}(\rho_{A_1^n B} \| \eta_{A_1^n B}) - \frac{g_1(\epsilon, \epsilon^{\frac{1}{4}})}{\alpha-1} \\ &\geq \sum_{k=1}^{n} H(A_k)_\rho - n(\alpha-1)\log^2(1+2|A|) - \frac{\alpha}{\alpha-1} D_{\max}^{\epsilon^{\frac{1}{4}}}(\rho_{A_1^n B} \| \eta_{A_1^n B}) - \frac{g_1(\epsilon, \epsilon^{\frac{1}{4}})}{\alpha-1} \\ &\geq \sum_{k=1}^{n} H(A_k)_\rho - n(\alpha-1)\log^2(1+2|A|) - \frac{\alpha}{\alpha-1} n\sqrt{\epsilon} - \frac{\alpha}{\alpha-1}\frac{1}{\sqrt{\epsilon}} - \frac{\alpha}{\alpha-1}\log(1-\sqrt{\epsilon}) - \frac{g_1(\epsilon, \epsilon^{\frac{1}{4}})}{\alpha-1}. \end{aligned}$$

In the third line above, we have used [**DFR20**, Lemma B.9] (which is an improvement of [**TCR09**, Lemma 8]), which is valid as long as $\alpha < 1 + \frac{1}{\log(1+2|A|)}$. We will select $\alpha = 1 + \frac{\epsilon^{1/4}}{\log(1+2|A|)}$ for which the above α bound is satisfied, this gives us

$$\begin{aligned} H_{\min}^{\epsilon^{\frac{1}{4}}+\epsilon}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H(A_k)_\rho - 3n\epsilon^{\frac{1}{4}}\log(1+2|A|) - \frac{2\log(1+2|A|)}{\epsilon^{3/4}} \\ - \frac{2\log(1+2|A|)}{\epsilon^{1/4}} \left(\log(1-\sqrt{\epsilon}) + g_1(\epsilon, \epsilon^{\frac{1}{4}})\right). \end{aligned}$$

□

We can now plug the bound in Eq. 3.19 to derive the following Corollary.

Corollary 3.10. *Let registers A_k have dimension $|A|$ for all $k \in [n]$. Suppose a quantum state $\rho_{A_1^n B}$ is such that for every $k \in [n]$, we have*

$$\left\| \rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1} B} \right\|_1 \leq \epsilon. \tag{3.24}$$

Then, we have that for $\delta = \epsilon \log |A| + g_2\left(\frac{\epsilon}{2}\right)$ such that $0 < \delta < 1$,

$$\begin{aligned} H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B)_\rho \geq \sum_{k=1}^n H(A_k)_\rho - 3n\delta^{\frac{1}{4}} \log(1+2|A|) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}} \left(\log(1-\sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}}) \right) \end{aligned} \tag{3.25}$$

where $g_1(x,y) = -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$ and $g_2(x) = (x+1)\log(x+1) - x\log(x)$. In particular, when all the states ρ_{A_k} are identical, we have

$$\begin{aligned} H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B)_\rho \geq n \left(H(A_1)_\rho - 3\delta^{\frac{1}{4}} \log(1+2|A|) \right) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}} \left(\log(1-\sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}}) \right). \end{aligned} \tag{3.26}$$

3.3.1. Weak approximate asymptotic equipartition

We can modify the proof of Theorem 3.9 to prove a *weak* approximate asymptotic equipartition property (AEP).

Theorem 3.11. *Let registers A_k have dimension $|A|$ for all $k \in [n]$ and the registers B_k have dimension $|B|$ for all $k \in [n]$. Suppose a quantum state $\rho_{A_1^n B_1^n E}$ is such that for every $k \in [n]$, we have*

$$\left\| \rho_{A_1^k B_1^k E} - \rho_{A_k B_k} \otimes \rho_{A_1^{k-1} B_1^{k-1} E} \right\|_1 \leq \epsilon. \tag{3.27}$$

Then, we have that for $\delta = \epsilon \log (|A||B|) + g_2\left(\frac{\epsilon}{2}\right)$ such that $0 < \delta < 1$,

$$\begin{aligned} H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^n H(A_k|B_k)_\rho - 3n\delta^{\frac{1}{4}} \log(1+2|A|) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}} \left(\log(1-\sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}}) \right) \end{aligned} \tag{3.28}$$

where $g_1(x,y) = -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$ and $g_2(x) = (x+1)\log(x+1) - x\log(x)$. In particular, when all the states $\rho_{A_k B_k}$ are identical, we have

$$\begin{aligned} H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B_1^n E)_\rho \geq n \left(H(A_1|B_1)_\rho - 3\delta^{\frac{1}{4}} \log(1+2|A|) \right) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}} \left(\log(1-\sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}}) \right). \end{aligned} \tag{3.29}$$

Proof. To prove this, we use the auxiliary state $\eta_{A_1^n B_1^n E} := \bigotimes \rho_{A_k B_k} \otimes \rho_E$. Then, we have

$$\begin{aligned} D(\rho_{A_1^n B_1^n E} \| \eta_{A_1^n B_1^n E}) &= I(A_1 B_1 : A_2 B_2 : \cdots : A_n B_n : E)_\rho \\ &= \sum_{k=1}^{n} I(A_k B_k : A_1^{k-1} B_1^{k-1} E)_\rho \\ &\leq n \left(\epsilon \log (|A||B|) + g\left(\frac{\epsilon}{2}\right) \right) = n\delta \end{aligned}$$

where we used the AFW bound for mutual information in the last line [**Wil13**, Theorem 11.10.4]. The rest of the proof follows the proof of Theorem 3.9, only difference being that now we have $\tilde{H}_\alpha^\uparrow(A_1^n|B_1^n E)_\eta = \sum_{k=1}^n \tilde{H}_\alpha^\uparrow(A_k|B_k)_\rho$. □

We call this generalisation *weak* because the smoothing term (δ) depends on size of the side information $|B|$. In Appendix A.5, we show that under the assumptions of the theorem, some sort of bound on the dimension of the registers B is necessary otherwise one cannot have a non-trivial bound on the smooth min-entropy.

3.3.2. Simple security proof for sequential device-independent quantum key distribution

The approximately independent register scenario and the associated min-entropy lower bound can be used to provide simple "proof of concept" security proofs for cryptographic protocols. In this section, we briefly sketch a proof for sequential device-independent quantum key distribution (DIQKD) to demonstrate this idea. We consider the sequential DIQKD protocol presented in Protocol 2.2.

For simplicity, we assume Eve (the adversary) distributes a state $\rho^{(0)}_{E_A E_B E}$ between Alice and Bob at the beginning of the protocol. Alice and Bob then use their states sequentially as given in Protocol 2.2. The k^{th} round of the protocol produces the questions X_k, Y_k and T_k, the answers A_k and B_k and transforms the shared state from $\rho^{(k-1)}_{E_A E_B E}$ to $\rho^{(k)}_{E_A E_B E}$.

Given the questions and answers of the previous rounds, the state shared between Alice and Bob and their devices in each round can be viewed as a device for playing the CHSH game. Suppose in the k^{th} round, the random variables produced in the previous $k-1$ rounds are $r_{k-1} := x_1^{k-1}, y_1^{k-1}, t_1^{k-1}, a_1^{k-1}, b_1^{k-1}$ and that the state shared between Alice and Bob is $\rho^{(k-1)}_{E_A E_B E|r_{k-1}}$. We can then define $\Pr[W_k|r_{k-1}]$ to be the winning probability of the CHSH game played by Alice and Bob using the state and their devices in the k^{th} round. Note that Alice's device cannot distinguish whether the CHSH game is played in a round or is used for key generation. We can further take an average over all the previous round's random

variables to derive the probability of winning the k^{th} game

$$\Pr[W_k] = \mathbb{E}_{r_{k-1}}\left[\Pr[W_k|r_{k-1}]\right]. \tag{3.30}$$

Alice and Bob randomly sample a subset of the rounds (using the random variable T_k) and play the CHSH game on this subset. If the average winning probability of CHSH game on this subset is small, they abort the protocol. For simplicity and brevity, we will assume here that the state $\rho^{(0)}_{E_A E_B E}$ distributed between Alice and Bob at the start of the protocol by Eve has an average winning probability at least ω_{exp}, that is,

$$\frac{1}{n}\sum_{k=1}^{n}\Pr[W_k] \geq \omega_{\text{exp}} - \delta \tag{3.31}$$

for some small $\delta > 0$[(5)].

For any shared state $\sigma_{E_A E_B E}$ (where E_A is held by Alice, E_B is held by Bob and E is held by the adversary) and local measurement devices, if Alice and Bob win the CHSH game with a probability $\omega \in (\frac{3}{4}, \frac{2+\sqrt{2}}{4}]$, then Alice's answer A to the game is random given the questions X,Y and the register E held by adversary. This is quantified by the following entropic bound (Lemma 2.33)

$$H(A|XYE) \geq f(\omega) = \begin{cases} \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16\omega(\omega-1)+3}\right) & \text{if } \omega \in [\frac{3}{4}, \frac{2+\sqrt{2}}{4}] \\ 0 & \text{if } \omega \in [0, \frac{3}{4}) \end{cases} \tag{3.32}$$

where $h(\cdot)$ is the binary entropy. The function f is convex over the interval $\left[0, \frac{2+\sqrt{2}}{4}\right]$. We plot it in the interval $[\frac{3}{4}, \frac{2+\sqrt{2}}{4}]$ in Fig. 3.1.

For $\epsilon > 0$, we choose the parameter $\omega_{\text{exp}} \in [\frac{3}{4} + \delta, \frac{2+\sqrt{2}}{4}]$ to be large enough so that

$$\log(2) - f(\omega_{\text{exp}} - \delta) = h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16(\omega_{\text{exp}} - \delta)(\omega_{\text{exp}} - \delta - 1) + 3}\right) \leq \epsilon^4. \tag{3.33}$$

We will now use Eq. 3.32 to bound the von Neumann entropy of the answers given Eve's information for the sequential DIQKD protocol. We have

$$\begin{aligned} H(A_1^n|X_1^n Y_1^n T_1^n E) &= \sum_{k=1}^{n} H(A_k|A_1^{k-1} X_1^n Y_1^n T_1^n E) \\ &\overset{(1)}{=} \sum_{k=1}^{n} H(A_k|A_1^{k-1} X_1^k Y_1^k T_1^k E) \\ &\overset{(2)}{=} \sum_{k=1}^{n} H(A_k|X_k Y_k R_{k-1} E) \end{aligned}$$

[(5)]By modifying the CHSH game played by Alice and Bob to the 3CHSH game (Sec. 6.2) and measuring the winning probability for this game, one can show that either this assumption is true or the protocol aborts with high probability.

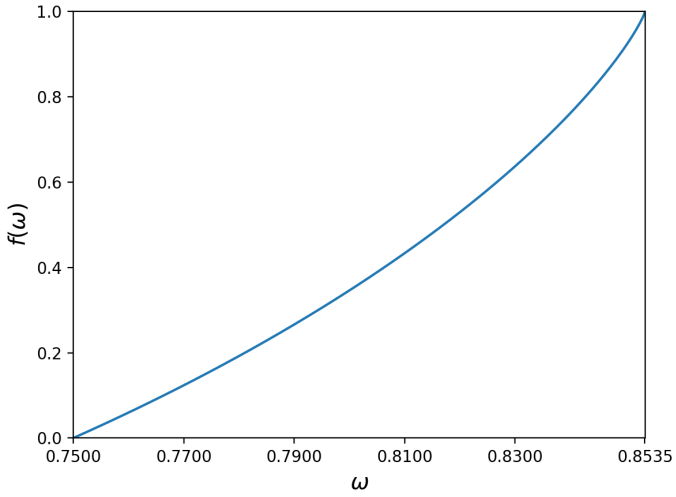

Fig. 3.1. The lower bound in Eq. 3.32 for the interval $[\frac{3}{4}, \frac{2+\sqrt{2}}{4}]$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \mathbb{E}_{r_{k-1}\sim R_{k-1}} \left[H(A_k|X_kY_kE)_{\rho^{(k)}_{|r_{k-1}}} \right] \\
&\overset{(3)}{\geq} \sum_{k=1}^{n} \mathbb{E}_{r_{k-1}\sim R_{k-1}} \left[f\left(\Pr[W_k|r_{k-1}]\right) \right] \\
&\geq \sum_{k=1}^{n} f\left(\Pr[W_k]\right) \\
&\geq nf\left(\frac{1}{n} \sum_{k=1}^{n} \Pr[W_k] \right) \\
&\geq nf(\omega_{\text{exp}} - \delta) \geq n(\log(2) - \epsilon^4)
\end{aligned}
$$

where in (1) we have used the fact that the questions sampled in the rounds after the k^{th} round are independent of the random variables in the previous rounds, in (2) we use the fact that Alice's answers are independent of the random variable T_k given the question X_k and we also grouped the random variables generated in the previous round into the random variable $R_{k-1} := A_1^{k-1}B_1^{k-1}X_1^{k-1}Y_1^{k-1}T_1^{k-1}$, in (3) we use the bound in Eq. 3.32, and in the next two steps we use convexity of f. If instead of the von Neumann entropy on the left-hand side above we had the smooth min-entropy, then the bound above would be sufficient to prove the security of DIQKD. However, this argument cannot be easily generalised to the smooth min-entropy because a chain rule like the one used in the first step does not exist for the smooth min-entropy (entropy accumulation [DFR20, MFSR24] generalises exactly such an argument). We can use the argument used for the approximately independent register case to transform this von Neumann entropy bound to a smooth min-entropy bound.

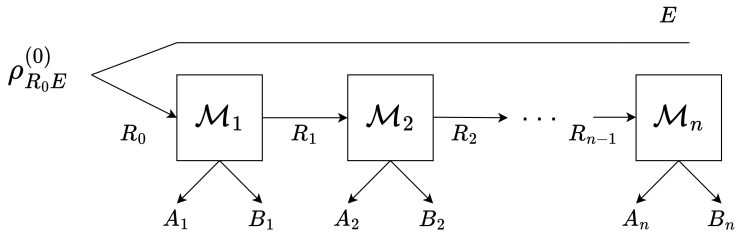

Fig. 3.2. The setting for the EAT and approximate EAT.

This bound results in the following bound on the multipartite mutual information

$$\begin{aligned} I(A_1 : \cdots : A_n : X_1^n Y_1^n T_1^n E) &= \sum_{k=1}^{n} H(A_k) + H(X_1^n Y_1^n T_1^n E) - H(A_1^n X_1^n Y_1^n T_1^n E) \\ &= \sum_{k=1}^{n} H(A_k) - H(A_1^n | X_1^n Y_1^n T_1^n E) \\ &\leq n \log(2) - n(\log(2) - \epsilon^4) = n\epsilon^4 \end{aligned}$$

where we have used the dimension bound $H(A_k) \leq 1$ for every $k \in [n]$. This is the same as the multipartite mutual information bound we derived while analysing approximately independent registers in Theorem 3.9. We can simply use the smooth min-entropy bound derived there here as well. This gives us the bound

$$\begin{aligned} H_{\min}^{2\epsilon}(A_1^n | X_1^n Y_1^n T_1^n E) &\geq \sum_{k=1}^{n} H(A_k) - 3n\epsilon \log 5 - O\left(\frac{1}{\epsilon^3}\right) \\ &= n(\log(2) - 3\epsilon \log 5) - O\left(\frac{1}{\epsilon^3}\right) \end{aligned} \tag{3.34}$$

where we have used the fact that the answers A_k can always be assumed to be uniformly distributed [**PAB+09**, **AF20**]. For every $\epsilon > 0$, we can choose a sufficiently large n so that this bound is large and positive.

We note that this method is only able to provide "proof of concept" or existence type security proofs. This proof method couples the value of the security parameter for privacy amplification ϵ with the average winning probability, which is not desirable. The parameter ϵ is chosen according to the security requirements of the protocol and is typically very small. For such values of ϵ, the average winning probability of the protocol will have to be extremely close to the maximum and we cannot realistically expect practical implementations to achieve such high winning probabilities. However, we do expect that this method will make it easier to create "proof of concept" type proofs for new cryptographic protocols in the future.

3.4. Approximate entropy accumulation

In this section, we will prove our first approximate version of the entropy accumulation theorem (EAT) using the entropic triangle inequality. We discussed EAT in Sec. 2.6. In the setting for entropy accumulation, a sequence of channels $\mathcal{M}_k : R_{k-1} \to A_k B_k R_k$ for $1 \leq k \leq n$ sequentially act on a state $\rho^{(0)}_{R_0 E}$ to produce the state $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E})$ (see Fig. 3.2). It is assumed that the channels $\mathcal{M}_k$ are such that the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k$ is satisfied for every k. Under this assumption, EAT provides the following lower bound for the smooth min-entropy

$$H^{\epsilon}_{\min}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^{n} \inf_{\omega_{R_{k-1}\tilde{R}}} H(A_k|B_k\tilde{R})_{\mathcal{M}_k(\omega)} - c\sqrt{n} \tag{3.35}$$

where the infimum is taken over all input states to the channels $\mathcal{M}_k$ and $c > 0$ is a constant depending only on $|A|$ (size of registers A_k) and ϵ. We will state and prove an approximate version of EAT. Consider the sequential process in Fig. 3.2 again. Now, suppose that the channels $\mathcal{M}_k$ do not necessarily satisfy the Markov chain conditions mentioned above, but each of the channels $\mathcal{M}_k$ can be ϵ-approximated by $\mathcal{M}'_k$ which satisfy the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k$ for a certain collection of inputs. The approximate entropy accumulation theorem below provides a lower bound on the smooth min-entropy in such a setting. The proof of this theorem again uses the technique based on the smooth min-entropy triangle inequality developed in the previous section. In this setting too, we have an approximation chain. For each $k \in [n]$,

$$\begin{aligned}\rho_{A_1^k B_1^k E} &= \mathrm{tr}_{R_k} \circ \mathcal{M}_k \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \right) \\ &\approx_\epsilon \mathrm{tr}_{R_k} \circ \mathcal{M}'_k \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \right) := \sigma^{(k)}_{A_1^k B_1^k E}.\end{aligned}$$

According to the Markov chain assumption for the channels $\mathcal{M}'_k$, the state $\sigma^{(k)}_{A_1^k B_1^k E}$, satisfies the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k$. Therefore, we expect that the register A_k adds some entropy to the smooth min-entropy $H^{\epsilon'}_{\min}(A_1^n|B_1^n E)_\rho$ and that the information leaked through B_1^n is not too large. We show that this is indeed the case in the approximate entropy accumulation theorem.

The approximate entropy accumulation theorem can be used to analyse and prove the security of cryptographic protocols under certain imperfections. The entropy accumulation theorem itself is used to prove the security of sequential device-independent quantum key distribution (DIQKD) protocols (Sec. 2.8.3). In these protocols, the side information B_k produced during each of the rounds are the questions used during the round to play a non-local game, like the CHSH game. In the ideal case, these questions are sampled independently of everything which came before. As an example of an imperfection, we can imagine that

crosstalk between the memory storing the secret bits A_1^{k-1} and the device producing the questions may lead to a small correlation between the side information produced during the k^{th} round and the secret bits A_1^{k-1} (also see [JK23, Tan23]). The approximate entropy accumulation theorem below can be used to prove security of DIQKD under such imperfections. We do not, however, pursue this example here and leave applications of this theorem for future work. In Sec. 3.4.5, we modify this Theorem to incorporate testing for EAT.

Theorem 3.12. *For $k \in [n]$, let the registers A_k and B_k be such that $|A_k| = |A|$ and $|B_k| = |B|$. For $k \in [n]$, let $\mathcal{M}_k$ be channels from $R_{k-1} \to R_k A_k B_k$ and*

$$\rho_{A_1^n B_1^n E} = \mathrm{tr}_{R_n} \circ \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \tag{3.36}$$

be the state produced by applying these maps sequentially. Suppose the channels $\mathcal{M}_k$ are such that for every $k \in [n]$, there exists a channel $\mathcal{M}'_k$ from $R_{k-1} \to R_k A_k B_k$ such that

(1) $\mathcal{M}'_k$ ϵ-approximates $\mathcal{M}_k$ in the diamond norm:

$$\frac{1}{2} \left\| \mathcal{M}_k - \mathcal{M}'_k \right\|_\diamond \leq \epsilon \tag{3.37}$$

(2) For every choice of a sequence of channels $\mathcal{N}_i \in \{\mathcal{M}_i, \mathcal{M}'_i\}$ for $i \in [k-1]$, the state $\mathcal{M}'_k \circ \mathcal{N}_{k-1} \circ \cdots \circ \mathcal{N}_1(\rho^{(0)}_{R_0 E})$ satisfies the Markov chain

$$A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k. \tag{3.38}$$

Then, for $0 < \delta, \epsilon_1, \epsilon_2 < 1$ such that $\epsilon_1 + \epsilon_2 < 1$, $\alpha \in \left(1, 1 + \frac{1}{\log(1+2|A|)}\right)$ and $\beta > 1$, we have

$$\begin{aligned} H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^n \inf_{\omega_{R_{k-1}\tilde{R}}} & H(A_k|B_k\tilde{R})_{\mathcal{M}'_k(\omega)} - n(\alpha-1)\log^2(1+2|A|) \\ & - \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(e^{\frac{\alpha-1}{\alpha} 2\log(|A||B|)} - 1\right)\right) \\ & - \frac{\alpha}{\alpha-1} n z_\beta(\epsilon,\delta) - \frac{1}{\alpha-1}\left(g_1(\epsilon_2,\epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta-1}\right). \end{aligned} \tag{3.39}$$

where

$$z_\beta(\epsilon,\delta) := \frac{\beta+1}{\beta-1} \log\left(\left(1 + \sqrt{(1-\delta)\epsilon}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\beta}\right)^{\frac{1}{\beta+1}}\right) \tag{3.40}$$

and $g_1(x,y) = -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$ and the infimum in Eq. 3.39 is taken over all input states $\omega_{R_{k-1}\tilde{R}}$ to the channels $\mathcal{M}'_k$ where $\tilde{R}$ is a reference register ($\tilde{R}$ can be chosen isomorphic to R_{k-1}).

For the choice of $\beta = 2$, $\delta = \epsilon^{\frac{1}{8}}$, we have

$$z_2(\epsilon,\delta) \leq 3\log\left(\left(1 + \epsilon^{\frac{1}{2}}\right)^{\frac{2}{3}} + \epsilon^{\frac{1}{12}}\right).$$

We also have that

$$\log\left(1+\delta e^{\frac{\alpha-1}{\alpha}2\log(|A||B|)}\right) \leq (|A||B|)^2\epsilon^{\frac{1}{8}}.$$

Finally, if we define $\epsilon_r := (|A||B|)^2\epsilon^{\frac{1}{8}} + 3\log\left(\left(1+\epsilon^{\frac{1}{2}}\right)^{\frac{2}{3}} + \epsilon^{\frac{1}{12}}\right)$, and choose $\alpha = \sqrt{\epsilon_r}$, we get the bound

$$H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n|B_1^nE)_\rho \geq \sum_{k=1}^{n} \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}'_k(\omega_{R_k\tilde{R}_k})} - n\sqrt{\epsilon_r}(\log^2(1+2|A|)+2) - \frac{1}{\sqrt{\epsilon_r}}\left(g_1(\epsilon_2,\epsilon_1)+2g_0(\epsilon_1)\right) \quad (3.41)$$

The entropy loss per round in the above bound behaves as $\sim \epsilon^{1/24}$. This dependence on ϵ is indeed very poor. In comparison, we can carry out a similar proof argument for classical probability distributions to get a dependence of $O(\sqrt{\epsilon})$ (Theorem A.13). The exponent of ϵ in our bound seems to be almost a factor of 12 off from the best possible bound. Roughly speaking, while carrying out the proof classically, we can bound the relevant channel divergences in the proof by $O(\epsilon)$, whereas in Eq. 3.41, we were only able to bound the channel divergence by $\sim \epsilon^{1/12}$. This leads to the deterioration of performance we see here as compared to the classical case. We will discuss this further in Sec. 3.4.6.

3.4.1. Proof idea

In order to prove this theorem, we will use a channel divergence based chain rule. Given any divergence $\mathbb{D}$, we can define the (stabilised) channel divergence based on $\mathbb{D}$ between two channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ as [**CMW16**, **LKDW18**]

$$\mathbb{D}(\mathcal{N}\,\|\,\mathcal{M}) := \sup_{\rho_{AR}} \mathbb{D}(\mathcal{N}_{A\to B}(\rho_{AR})\|\,\mathcal{M}_{A\to B}(\rho_{AR})) \quad (3.42)$$

where R is reference register of arbitrary size ($|R| = |A|$ can be chosen when $\mathbb{D}$ satisfies the data processing inequality).

Recently proven chain rules for α-Rényi relative entropy [**FF21**, Corollary 5.2] state that for $\alpha > 1$ and states ρ_A and σ_A, and channels $\mathcal{E}_{A\to B}$ and $\mathcal{F}_{A\to B}$, we have

$$\tilde{D}_\alpha(\mathcal{E}_{A\to B}(\rho_A)\|\mathcal{F}_{A\to B}(\sigma_A)) \leq \tilde{D}_\alpha(\rho_A\|\sigma_A) + \tilde{D}_\alpha^{\text{reg}}(\mathcal{E}_{A\to B}\|\mathcal{F}_{A\to B}) \quad (3.43)$$

where $\tilde{D}_\alpha^{\text{reg}}(\mathcal{E}_{A\to B}\|\mathcal{F}_{A\to B}) := \lim_{n\to\infty}\frac{1}{n}\tilde{D}_\alpha(\mathcal{E}_{A\to B}^{\otimes n}\|\mathcal{F}_{A\to B}^{\otimes n})$.

Now observe that if we were guaranteed that for the maps in Theorem 3.12 above, $\tilde{D}_\alpha^{\text{reg}}(\mathcal{M}_k\,\|\,\mathcal{M}'_k) \leq \epsilon$ for every k for some $\alpha > 1$. Then, we could use the chain rule in Eq. 3.43

as follows

$$\begin{aligned}
&\tilde{D}_\alpha(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E}) \| \mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(\rho^{(0)}_{R_0E})) \\
&\qquad \leq \tilde{D}_\alpha(\mathcal{M}_{n-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E}) \| \mathcal{M}'_{n-1} \circ \cdots \circ \mathcal{M}'_1(\rho^{(0)}_{R_0E})) + \tilde{D}^{\mathrm{reg}}_\alpha(\mathcal{M}_n \| \mathcal{M}'_n) \\
&\qquad \leq \cdots \\
&\qquad \leq \tilde{D}_\alpha(\rho^{(0)}_{R_0E} \| \rho^{(0)}_{R_0E}) + \sum_{k=1}^{n} \tilde{D}^{\mathrm{reg}}_\alpha(\mathcal{M}_k \| \mathcal{M}'_k) \\
&\qquad \leq n\epsilon.
\end{aligned}$$

Once we have the above result we can simply use the well known relation between smooth max-relative entropy and α-Rényi relative entropy [**Tom16**, Proposition 6.5] to get the bound

$$\begin{aligned}
&D^{\epsilon'}_{\max}(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E}) \| \mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(\rho^{(0)}_{R_0E})) \\
&\qquad \leq \tilde{D}_\alpha(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E}) \| \mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(\rho^{(0)}_{R_0E})) + \frac{g_0(\epsilon')}{\alpha - 1} \\
&\qquad \leq n\epsilon + O(1).
\end{aligned}$$

This bound can subsequently be used in Lemma 3.5 to relate the smooth min-entropy of the real state $\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0E})$ with the α–Rényi conditional entropy of the auxiliary state $\mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(\rho^{(0)}_{R_0E})$, for which we can use the original entropy accumulation theorem.

In order to prove Theorem 3.12, we broadly follow this idea. However, the condition $\|\mathcal{M}_k - \mathcal{M}'_k\|_\diamond \leq \epsilon$ does not lead to any kind of bound on $\tilde{D}^{\mathrm{reg}}_\alpha$ or any other channel divergence. We will get around this issue by instead using mixed channels $\mathcal{M}^\delta_k := (1-\delta)\mathcal{M}'_k + \delta\mathcal{M}_k$. Also, instead of trying to bound channel divergence in terms of $\tilde{D}^{\mathrm{reg}}_\alpha$, we will bound the $D^{\#}_\alpha$ (defined in the next section) channel divergence and use its chain rule. We develop the relevant α-Rényi divergence bounds for this divergence in the next two subsections and then prove the theorem above in Sec 3.4.4.

3.4.2. Divergence bound for approximately equal states

We will use the sharp Rényi divergence $D^{\#}_\alpha$ defined in Ref. [**FF21**] (see [**BSD21**] for the following equivalent definition) in this section. For $\alpha > 1$ and two positive operators P and Q, it is defined

$$D^{\#}_\alpha(P\|Q) := \min_{A \geq P} \hat{D}_\alpha(A\|Q) \tag{3.44}$$

where $\hat{D}_\alpha(A\|Q)$ is the α-Rényi geometric divergence [Mat18]. For $\alpha > 1$, it is defined as

$$\hat{D}_\alpha(A\|Q) = \begin{cases} \frac{1}{\alpha-1} \log \operatorname{tr}\left(Q\left(Q^{-\frac{1}{2}} A Q^{-\frac{1}{2}}\right)^\alpha\right) & \text{if } A \ll Q \\ \infty & \text{otherwise.} \end{cases} \tag{3.45}$$

A in the optimisation above is any operator $A \geq P$. In general, such an operator A is unnormalised. We will prove a bound on $D_\alpha^\#$ between two states in terms of the distance between them and their max-relative entropy. In order to prove this bound, we require the following simple generalisation of the pinching inequality (see, for example, [Tom16, Sec. 2.6.3]).

Lemma 3.13 (Asymmetric pinching)**.** *For $t > 0$, a positive semidefinite operator $X \geq 0$ and orthogonal projections Π and $\Pi_\perp = \mathbb{1} - \Pi$, we have that*

$$X \leq (1+t)\Pi X \Pi + \left(1 + \frac{1}{t}\right) \Pi_\perp X \Pi_\perp. \tag{3.46}$$

Proof. We will write the positive matrix X as the block matrix

$$X = \begin{pmatrix} X_1 & X_2 \\ X_2^* & X_3 \end{pmatrix}$$

where the blocks are partitioned according to the direct sum $\operatorname{im}(\Pi) \oplus \operatorname{im}(\Pi_\perp)$. Then, the statement in the lemma is equivalent to proving that

$$\begin{pmatrix} X_1 & X_2 \\ X_2^* & X_3 \end{pmatrix} \leq \begin{pmatrix} (1+t)X_1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \left(1+\frac{1}{t}\right) X_3 \end{pmatrix}$$

which is equivalent to proving that

$$0 \leq \begin{pmatrix} tX_1 & -X_2 \\ -X_2^* & \frac{1}{t} X_3 \end{pmatrix}.$$

This is true because

$$\begin{pmatrix} tX_1 & -X_2 \\ -X_2^* & \frac{1}{t} X_3 \end{pmatrix} = \begin{pmatrix} -t^{1/2} & 0 \\ 0 & t^{-1/2} \end{pmatrix} \begin{pmatrix} X_1 & X_2 \\ X_2^* & X_3 \end{pmatrix} \begin{pmatrix} -t^{1/2} & 0 \\ 0 & t^{-1/2} \end{pmatrix} \geq 0$$

since $X \geq 0$. □

Lemma 3.14. *Let $\epsilon > 0$ and $\alpha \in (1, \infty)$, ρ and σ be two normalised quantum states on the Hilbert space $\mathbb{C}^n$ such that $\frac{1}{2}\|\rho - \sigma\|_1 \leq \epsilon$ and also $D_{\max}(\rho\|\sigma) \leq d < \infty$, then we have the bound*

$$D_\alpha^\#(\rho\|\sigma) \leq \frac{\alpha+1}{\alpha-1} \log\left((1+\sqrt{\epsilon})^{\frac{\alpha}{\alpha+1}} + \left(e^{\alpha d}\sqrt{\epsilon}\right)^{\frac{1}{\alpha+1}}\right). \tag{3.47}$$

Note: For a fixed $\alpha \in (1, \infty)$, this upper bound tends to zero as $\epsilon \to 0$. On the other hand, for a fixed $\epsilon \in (0,1)$, the upper bound tends to infinity as $\alpha \to 1$ (that is, the bound becomes trivial). In Appendix A.2, we show that a bound of this form for $D_\alpha^\#$ necessarily diverges for $\epsilon > 0$ as $\alpha \to 1$.

Proof. Since, $D_{\max}(\rho\|\sigma) < \infty$, we have that $\rho \ll \sigma$. We can assume that σ is invertible. If it was not, then we could always restrict our vector space to the subspace $\mathrm{supp}(\sigma)$.

Let $\rho - \sigma = P - Q$, where $P \geq 0$ is the positive part of the matrix $\rho - \sigma$ and $Q \geq 0$ is its negative part. We then have that $\mathrm{tr}(P) = \mathrm{tr}(Q) \leq \epsilon$.

Further, let

$$\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}} = \sum_{i=1}^{n} \lambda_i |x_i\rangle \langle x_i| \tag{3.48}$$

be the eigenvalue decomposition of $\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}}$. Define the real vector $q \in \mathbb{R}^n$ as

$$q(i) := \langle x_i| \sigma |x_i\rangle .$$

Note that q is a probability distribution. Observe that

$$\begin{aligned}
\mathbb{E}_{I\sim q}[\lambda_I] &= \sum_{i=1}^{n} \lambda_i \langle x_i| \sigma |x_i\rangle \\
&= \mathrm{tr}\left(\sigma \sum_{i=1}^{n} \lambda_i |x_i\rangle \langle x_i|\right) \\
&= \mathrm{tr}\left(\sigma \sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}}\right) \\
&= \mathrm{tr}(P) \\
&\leq \epsilon.
\end{aligned}$$

Also, observe that $\lambda_i \geq 0$ for all $i \in [n]$ because $\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}} \geq 0$. Let's define

$$S := \{i \in [n] : \lambda_i \leq \sqrt{\epsilon}\}. \tag{3.49}$$

Since, $\lambda_i \geq 0$ for all $i \in [n]$, we can use the Markov inequality to show:

$$\begin{aligned}
\Pr_q(I \in S^c) &= \Pr_q(\lambda_I > \sqrt{\epsilon}) \\
&\leq \frac{\mathbb{E}_{I\sim q}[\lambda_I]}{\sqrt{\epsilon}} \\
&\leq \sqrt{\epsilon}.
\end{aligned}$$

Thus, if we define the projectors $\Pi := \sum_{i\in S} |x_i\rangle \langle x_i|$ and $\Pi_\perp := \sum_{i\in S^c} |x_i\rangle \langle x_i| = \mathbb{1} - \Pi$, we have

$$\mathrm{tr}(\sigma \Pi_\perp) = \sum_{i\in S^c} \langle x_i| \sigma |x_i\rangle$$

$$= \Pr_q(I \in S^c)$$
$$\leq \sqrt{\epsilon}. \tag{3.50}$$

Moreover, by the definition of set S (Eq. 3.49) we have

$$\Pi\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}}\Pi = \sum_{i\in S} \lambda_i |x_i\rangle\langle x_i|$$
$$\leq \sqrt{\epsilon}\Pi \tag{3.51}$$

and using $D_{\max}(\rho\|\sigma) \leq d$, we have that

$$\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} \leq e^d \mathbb{1}\,. \tag{3.52}$$

Now, observe that since $\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} \geq 0$, for an arbitrary $t > 0$, using Lemma 3.13 we have

$$\begin{aligned}
\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} &\leq (1+t)\Pi\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\Pi + \left(1+\frac{1}{t}\right)\Pi_\perp\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\Pi_\perp \\
&\leq (1+t)\Pi\left(\mathbb{1} + \sigma^{-\frac{1}{2}}P\sigma^{-\frac{1}{2}}\right)\Pi + \left(1+\frac{1}{t}\right)e^d\Pi_\perp \\
&\leq (1+t)(1+\sqrt{\epsilon})\Pi + \left(1+\frac{1}{t}\right)e^d\Pi_\perp
\end{aligned}$$

where we have used $\rho \leq \sigma + P$ to bound the first term and Eq. 3.52 to bound the second term in the second line, and Eq. 3.51 to bound $\Pi\sigma^{-\frac{1}{2}}P\sigma^{-\frac{1}{2}}\Pi$ in the last step.

We will define $A_t := (1+t)(1+\sqrt{\epsilon})\sigma^{\frac{1}{2}}\Pi\sigma^{\frac{1}{2}} + \left(1+\frac{1}{t}\right)e^d\sigma^{\frac{1}{2}}\Pi_\perp\sigma^{\frac{1}{2}}$. Above, we have shown that $A_t \geq \rho$ for every $t > 0$. Therefore, for each $t > 0$, $D^{\#}_\alpha(\rho\|\sigma) \leq \hat{D}_\alpha(A_t\|\sigma)$. We will now bound $\hat{D}_\alpha(A_t\|\sigma)$ for $\alpha \in (1,\infty)$ as:

$$\begin{aligned}
\hat{D}_\alpha(A_t\|\sigma) &= \frac{1}{\alpha-1}\log \operatorname{tr}\left(\sigma\left(\sigma^{-\frac{1}{2}}A_t\sigma^{-\frac{1}{2}}\right)^\alpha\right) \\
&= \frac{1}{\alpha-1}\log \operatorname{tr}\left(\sigma\left((1+t)(1+\sqrt{\epsilon})\Pi + \left(1+\frac{1}{t}\right)e^d\Pi_\perp\right)^\alpha\right) \\
&= \frac{1}{\alpha-1}\log \operatorname{tr}\left(\sigma\left((1+t)^\alpha(1+\sqrt{\epsilon})^\alpha\Pi + \left(1+\frac{1}{t}\right)^\alpha e^{d\alpha}\Pi_\perp\right)\right) \\
&= \frac{1}{\alpha-1}\log\left((1+t)^\alpha(1+\sqrt{\epsilon})^\alpha \operatorname{tr}(\sigma\Pi) + \left(1+\frac{1}{t}\right)^\alpha e^{d\alpha}\operatorname{tr}(\sigma\Pi_\perp)\right) \\
&\leq \frac{1}{\alpha-1}\log\left((1+t)^\alpha(1+\sqrt{\epsilon})^\alpha + \left(1+\frac{1}{t}\right)^\alpha e^{d\alpha}\sqrt{\epsilon}\right)
\end{aligned}$$

where in the last line we use $\operatorname{tr}(\sigma\Pi) \leq 1$ and $\operatorname{tr}(\sigma\Pi_\perp) \leq \sqrt{\epsilon}$ (Eq. 3.50). Finally, since $t > 0$ was arbitrary, we can choose the $t > 0$ which minimizes the right-hand side. For this choice of $t_{\min} = \left(\frac{e^{\alpha d}\sqrt{\epsilon}}{(1+\sqrt{\epsilon})^\alpha}\right)^{\frac{1}{\alpha+1}}$, we get

$$\hat{D}_\alpha(A_{t_{\min}}\|\sigma) \leq \frac{\alpha+1}{\alpha-1}\log\left((1+\sqrt{\epsilon})^{\frac{\alpha}{\alpha+1}} + e^{\frac{\alpha}{\alpha+1}d}\epsilon^{\frac{1}{2(\alpha+1)}}\right)$$

which proves the required bound. □

3.4.3. Bounding the channel divergence for two channels close to each other

Suppose there are two channels $\mathcal{N}$ and $\mathcal{M}$ mapping registers from the space A to B such that $\frac{1}{2}\|\mathcal{N}-\mathcal{M}\|_\diamond \leq \epsilon$. In general, the channel divergence between two such channels can be infinite because there may be states ρ such that $\mathcal{N}(\rho) \not\ll \mathcal{M}(\rho)$. In order to get around this issue, we will use the δ–mixed channel, $\mathcal{M}_\delta$. For $\delta \in (0,1)$, we define $\mathcal{M}_\delta$ as

$$\mathcal{M}_\delta := (1-\delta)\mathcal{M} + \delta\mathcal{N}.$$

This guarantees that $D_{\max}(\mathcal{N}\|\mathcal{M}_\delta) \leq \log\frac{1}{\delta}$, which is enough to ensure that the divergences we are interested in are finite. Moreover, by mixing $\mathcal{M}$ with $\mathcal{N}$, we only decrease the distance:

$$\begin{aligned}\frac{1}{2}\|\mathcal{M}_\delta-\mathcal{N}\|_\diamond &= \frac{1}{2}\|(1-\delta)\mathcal{M}+\delta\mathcal{N}-\mathcal{N}\|_\diamond \\ &= (1-\delta)\frac{1}{2}\|\mathcal{M}-\mathcal{N}\|_\diamond \\ &\leq (1-\delta)\epsilon. \end{aligned} \tag{3.53}$$

We will now show that $D^\#_\alpha(\mathcal{N}\|\mathcal{M}_\delta)$ is small for an appropriately chosen δ. By the definition of channel divergence, we have that

$$D^\#_\alpha(\mathcal{N}\|\mathcal{M}_\delta) = \sup_{\rho_{AR}} D^\#_\alpha(\mathcal{N}(\rho_{AR})\|\mathcal{M}_\delta(\rho_{AR}))$$

where R is an arbitrary reference system ($\mathcal{N},\mathcal{M}_\delta$ map register A to register B). We will show that for every ρ_{AR}, $D^\#_\alpha(\mathcal{N}(\rho_{AR})\|\mathcal{M}_\delta(\rho_{AR}))$ is small. Note that

$$\begin{aligned}\mathcal{M}_\delta(\rho_{AR}) &= (1-\delta)\mathcal{M}(\rho_{AR}) + \delta\mathcal{N}(\rho_{AR}) \\ &\geq \delta\mathcal{N}(\rho_{AR})\end{aligned}$$

which implies that $D_{\max}(\mathcal{N}(\rho_{AR})\|\mathcal{M}_\delta(\rho_{AR})) \leq \log\frac{1}{\delta}$. Also, using Eq. 3.53 have that

$$\frac{1}{2}\|\mathcal{M}_\delta(\rho_{AR})-\mathcal{N}(\rho_{AR})\|_1 \leq (1-\delta)\epsilon.$$

Using Lemma 3.14, we have for every $\alpha \in (1,\infty)$

$$D^\#_\alpha(\mathcal{N}(\rho_{AR})\|\mathcal{M}_\delta(\rho_{AR})) \leq \frac{\alpha+1}{\alpha-1}\log\left(\left(1+\sqrt{(1-\delta)\epsilon}\right)^{\frac{\alpha}{\alpha+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\alpha}\right)^{\frac{1}{\alpha+1}}\right).$$

Since, this is true for all ρ_{AR}, for every $\alpha \in (1,\infty)$ we have

$$D^\#_\alpha(\mathcal{N}\|\mathcal{M}_\delta) \leq \frac{\alpha+1}{\alpha-1}\log\left(\left(1+\sqrt{(1-\delta)\epsilon}\right)^{\frac{\alpha}{\alpha+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\alpha}\right)^{\frac{1}{\alpha+1}}\right).$$

Note that since δ was arbitrary, we can choose it appropriately to make sure that the above bound is small, for example by choosing $\delta = \epsilon^{\frac{1}{4\alpha}}$, we get the bound

$$D_\alpha^\#(\mathcal{N}\,\|\,\mathcal{M}_\delta) \leq \frac{\alpha+1}{\alpha-1}\log\left((1+\sqrt{\epsilon})^{\frac{\alpha}{\alpha+1}} + \epsilon^{\frac{1}{4(\alpha+1)}}\right)$$

which is a small function of ϵ in the sense that it tends to 0 as $\epsilon \to 0$. We summarise the bound derived above in the following lemma.

Lemma 3.15. *Let $\epsilon > 0$. Suppose channels $\mathcal{N}$ and $\mathcal{M}$ from register A to B are such that $\frac{1}{2}\|\mathcal{N}-\mathcal{M}\|_\diamond \leq \epsilon$. For $\delta \in (0,1)$, we can define the mixed channel $\mathcal{M}_\delta := (1-\delta)\mathcal{M} + \delta\mathcal{N}$. Then, for every $\alpha \in (1,\infty)$, we have the following bound on the channel divergence*

$$D_\alpha^\#(\mathcal{N}\,\|\,\mathcal{M}_\delta) \leq \frac{\alpha+1}{\alpha-1}\log\left(\left(1+\sqrt{(1-\delta)\epsilon}\right)^{\frac{\alpha}{\alpha+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\alpha}\right)^{\frac{1}{\alpha+1}}\right). \tag{3.54}$$

3.4.4. Proof of the approximate entropy accumulation theorem

We use the mixed channels defined in the previous section to define the auxiliary state $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E}^{(0)})$ for our proof. It is easy to show using the divergence bounds in Sec. 3.4.3 and the chain rule for $D_\alpha^\#$ entropies that the relative entropy distance between the real state and this choice of the auxiliary state is small. However, the state $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E}^{(0)})$ does not necessarily satisfy the Markov chain conditions required for entropy accumulation. Thus, we also need to reprove the entropy lower bound on this state by modifying the approach used in the proof of the original entropy accumulation theorem.

Proof of Theorem 3.12. Using Lemma 3.15, for every $\delta \in (0,1)$ and for each $k \in [n]$ we have that for every $\beta > 1$, the mixed maps $\mathcal{M}_k^\delta := (1-\delta)\,\mathcal{M}_k' + \delta\,\mathcal{M}_k$ satisfy

$$\begin{aligned} D_\beta^\#(\mathcal{M}_k\,\|\,\mathcal{M}_k^\delta) &\leq \frac{\beta+1}{\beta-1}\log\left(\left(1+\sqrt{(1-\delta)\epsilon}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\beta}\right)^{\frac{1}{\beta+1}}\right) \\ &:= z_\beta(\epsilon,\delta) \end{aligned} \tag{3.55}$$

where we defined the right-hand side above as $z_\beta(\epsilon,\delta)$. This can be made "small" by choosing $\delta = \epsilon^{\frac{1}{4\beta}}$ as was shown in the previous section. We use these maps to define the auxiliary state as

$$\sigma_{A_1^nB_1^nE} := \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E}^{(0)}). \tag{3.56}$$

Now, we have that for $\beta > 1$ and $\epsilon_1 > 0$

$$D_{\max}^{\epsilon_1}(\rho_{A_1^nB_1^nE}\|\sigma_{A_1^nB_1^nE})$$

$$
\begin{aligned}
&\leq \tilde{D}_\beta(\rho_{A_1^n B_1^n E} \| \sigma_{A_1^n B_1^n E}) + \frac{g_0(\epsilon_1)}{\beta - 1} \\
&\leq D_\beta^{\#}(\rho_{A_1^n B_1^n E} \| \sigma_{A_1^n B_1^n E}) + \frac{g_0(\epsilon_1)}{\beta - 1} \\
&= D_\beta^{\#}(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \| \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho^{(0)}_{R_0 E})) + \frac{g_0(\epsilon_1)}{\beta - 1} \\
&\leq D_\beta^{\#}(\mathcal{M}_{n-1} \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \| \mathcal{M}_{n-1}^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho^{(0)}_{R_0 E})) + D_\beta^{\#}(\mathcal{M}_n \| \mathcal{M}_n^\delta) + \frac{g_0(\epsilon_1)}{\beta - 1} \\
&\leq \cdots \\
&\leq \sum_{k=1}^{n} D_\beta^{\#}(\mathcal{M}_k \| \mathcal{M}_k^\delta) + \frac{g_0(\epsilon_1)}{\beta - 1} \\
&\leq n z_\beta(\epsilon, \delta) + \frac{g_0(\epsilon_1)}{\beta - 1}
\end{aligned} \tag{3.57}
$$

where the first line follows from [Tom16, Proposition 6.5], the second line follows from [FF21, Proposition 3.4], fourth line follows from the chain rule for $D_\beta^{\#}$ [FF21, Proposition 4.5], and the last line follows from Eq. 3.55.

For $\epsilon_2 > 0$ and $\alpha \in (1, 1 + \frac{1}{\log(1+2|A|)})$, we can plug the above in the bound provided by Lemma 3.5 to get

$$
\begin{aligned}
H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n|B_1^n E)_\rho \geq \tilde{H}_\alpha^{\uparrow}(A_1^n|B_1^n E)_\sigma &- \frac{\alpha}{\alpha - 1} n z_\beta(\epsilon, \delta) \\
&- \frac{1}{\alpha - 1}\left(g_1(\epsilon_2, \epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta - 1}\right).
\end{aligned} \tag{3.58}
$$

We have now reduced our problem to lower bounding $\tilde{H}_\alpha^{\uparrow}(A_1^n|B_1^n E)_\sigma$. Note that we cannot directly use the entropy accumulation here, since the mixed maps $\mathcal{M}_k^\delta = (1-\delta)\,\mathcal{M}_k' + \delta\,\mathcal{M}_k$, which means that with δ probability the B_k register may be correlated with A_1^{k-1} even given $B_1^{k-1}E$, and it may not satisfy the Markov chain required for entropy accumulation.

The application of the maps $\mathcal{M}_k^\delta$ can be viewed as applying the channel $\mathcal{M}_k'$ with probability $1-\delta$ and the channel $\mathcal{M}_k$ with probability δ. We can define the channels $\mathcal{N}_k$ which map the registers R_{k-1} to $R_k A_k B_k C_k$, where C_k is a binary register. The action of $\mathcal{N}_k$ can be defined as:

(1) Sample the classical random variable $C_k \in \{0,1\}$ independently. $C_k = 1$ with probability $1-\delta$ and 0 otherwise.

(2) If $C_k = 1$ apply the map $\mathcal{M}_k'$ on R_{k-1}, else apply $\mathcal{M}_k$ on R_{k-1}.

Let us call $\theta_{A_1^n B_1^n C_1^n E} = \mathcal{N}_n \circ \cdots \circ \mathcal{N}_1(\rho^{(0)}_{R_0 E})$. Clearly $\mathrm{tr}_{C_1^n}\left(\theta_{A_1^n B_1^n C_1^n E}\right) = \sigma_{A_1^n B_1^n E}$. Thus, we have

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_{\sigma} &= \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_{\theta} \\ &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n C_1^n E)_{\theta}. \end{aligned} \tag{3.59}$$

We will now focus on lower bounding $\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n C_1^n E)_{\theta}$. Using [Tom16, Proposition 5.1], we have that

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n C_1^n E)_{\theta} = \frac{\alpha}{1-\alpha}\log \sum_{c_1^n} \theta(c_1^n) \exp\left(\frac{1-\alpha}{\alpha}\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_{\theta_{|c_1^n}}\right).$$

We will show that for a given c_1^n, the conditional entropy $\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_{\theta_{|c_1^n}}$ accumulates whenever the "good" map $\mathcal{M}'_k$ is used and loses some entropy for the rounds where the "bad" map $\mathcal{M}_k$ is used. The fact that c_1^n contains far more 1s than 0s with a large probability then allows us to prove a lower bound on $\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n C_1^n E)_{\theta}$.

Claim 3.16. *Define* $h_k := \inf_{\omega} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_k\tilde{R}_{k-1})_{\mathcal{M}'_k(\omega)}$ *where the infimum is over all states* $\omega_{R_{k-1}\tilde{R}_{k-1}}$ *for a register* $\tilde{R}_{k-1}$*, which is isomorphic to* R_{k-1}*, and* $s := \log(|A||B|^2)$*. Then, we have*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_{\theta_{|c_1^n}} \geq \sum_{k=1}^{n} \left(\delta(c_k,1)h_k - \delta(c_k,0)s\right) \tag{3.60}$$

where $\delta(x,y)$ *is the Kronecker delta function (*$\delta(x,y) = 1$ *if* $x = y$ *and* 0 *otherwise).*

Proof. We will prove the statement

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^k|B_1^k E)_{\theta_{|c_1^k}} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\theta_{|c_1^{k-1}}} + \left(\delta(c_k,1)h_k - \delta(c_k,0)s\right)$$

then the claim will follow inductively. We will consider two cases: when $c_k = 0$ and when $c_k = 1$. First suppose, $c_k = 0$ then $\theta_{A_1^k B_1^k E|c_1^k} = \mathrm{tr}_{R_k} \circ \mathcal{M}_k^{R_{k-1}\to R_k A_k B_k}\left(\theta_{R_{k-1}A_1^{k-1}B_1^{k-1}E|c_1^k}\right)$. In this case, we have

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1^k|B_1^k E)_{\theta_{|c_1^k}} &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^{k-1}|B_1^k E)_{\theta_{|c_1^k}} - \log|A| \\ &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\theta_{|c_1^k}} - \log\left(|A||B|^2\right) \\ &= \tilde{H}^{\uparrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\theta_{|c_1^{k-1}}} - s\end{aligned}$$

where in the first line we have used the dimension bound in Lemma A.8, in the second line we have used the dimension bound in Lemma A.10 and in the last line we have used $\theta_{A_1^{k-1}B_1^{k-1}E|c_1^k} = \theta_{A_1^{k-1}B_1^{k-1}E|c_1^{k-1}}$.

Now, suppose that $c_k = 1$. In this case, we have that $\theta_{A_1^k B_1^k E|c_1^k} = \mathrm{tr}_{R_k} \circ \mathcal{M}'_k\left(\theta_{R_{k-1}A_1^{k-1}B_1^{k-1}E|c_1^k}\right)$ and since $\theta_{R_{k-1}A_1^{k-1}B_1^{k-1}E|c_1^k} = \Phi_{k-1} \circ \Phi_{k-2} \cdots \circ \Phi_1(\rho^{(0)}_{R_0 E})$ where each of the $\Phi_i \in \{\mathcal{M}_i, \mathcal{M}'_i\}$, using

the hypothesis of the theorem we have that the state $\theta_{A_1^k B_1^k E|c_1^k} = \mathcal{M}'_k\left(\theta_{R_{k-1}A_1^{k-1}B_1^{k-1}E|c_1^k}\right)$ satisfies the Markov chain

$$A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k.$$

Now, using Corollary A.7 (the $\tilde{H}^{\uparrow}_{\alpha}$ counterpart for [DFR20, Corollary 3.5], which is the main chain rule used for proving entropy accumulation), we have

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1^k|B_1^kE)_{\theta_{|c_1^k}} &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\theta_{|c_1^k}} + \inf_{\omega} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_k\tilde{R}_{k-1})_{\mathcal{M}'_k(\omega)} \\ &= \tilde{H}^{\uparrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\theta_{|c_1^{k-1}}} + h_k\end{aligned}$$

where in the last line we have again used $\theta_{A_1^{k-1}B_1^{k-1}E|c_1^k} = \theta_{A_1^{k-1}B_1^{k-1}E|c_1^{k-1}}$. Combining these two cases, we have

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^k|B_1^kE)_{\theta_{|c_1^k}} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\theta_{|c_1^{k-1}}} + (\delta(c_k,1)h_k - \delta(c_k,0)s)\,. \tag{3.61}$$

Using this bound n times starting with $\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^nE)_{\theta_{|c_1^n}}$ gives us the bound required in the claim. □

For the sake of clarity let $l_k(c_k) := (\delta(c_k,1)h_k - \delta(c_k,0)s)$. We will now evaluate

$$\begin{aligned}\sum_{c_1^n}\theta(c_1^n)\exp\left(\frac{1-\alpha}{\alpha}\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^nE)_{\theta_{|c_1^n}}\right) &\leq \sum_{c_1^n}\theta(c_1^n)\exp\left(\frac{1-\alpha}{\alpha}\sum_{k=1}^n l_k(c_k)\right) \\ &= \sum_{c_1^n}\prod_{k=1}^n \theta(c_k)e^{\frac{1-\alpha}{\alpha}l_k(c_k)} \\ &= \prod_{k=1}^n\sum_{c_k}\theta(c_k)e^{\frac{1-\alpha}{\alpha}l_k(c_k)}.\end{aligned} \tag{3.62}$$

Then, we have

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^nC_1^nE)_{\theta} &= \frac{\alpha}{1-\alpha}\log\sum_{c_1^n}\theta(c_1^n)\exp\left(\frac{1-\alpha}{\alpha}\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^nE)_{\theta_{|c_1^n}}\right). \\ &\geq \frac{\alpha}{1-\alpha}\sum_{k=1}^n\log\sum_{c_k}\theta(c_k)e^{\frac{1-\alpha}{\alpha}l_k(c_k)} \\ &= \frac{\alpha}{1-\alpha}\sum_{k=1}^n\log\left((1-\delta)e^{\frac{1-\alpha}{\alpha}h_k} + \delta e^{-\frac{1-\alpha}{\alpha}s}\right) \\ &= \sum_{k=1}^n h_k - \frac{\alpha}{\alpha-1}\sum_{k=1}^n\log\left(1-\delta+\delta e^{\frac{\alpha-1}{\alpha}(s+h_k)}\right) \\ &\geq \sum_{k=1}^n h_k - \frac{\alpha}{\alpha-1}n\log\left(1+\delta\left(e^{\frac{\alpha-1}{\alpha}(s+\log|A|)}-1\right)\right)\end{aligned} \tag{3.63}$$

where in the second line we have used Eq. 3.62 and in the last line we have used the fact that $h_k \leq \log|A|$ for all $k \in [n]$.

We restricted the choice of α to the region $\left(1, 1+\frac{1}{\log(1+2|A|)}\right)$ in the theorem, so that we can now use [DFR20, Lemma B.9] to transform the above to

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^nC_1^nE)_{\theta} \geq \sum_{k=1}^{n} \inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} H(A_k|B_k\tilde{R}_{k-1})_{\mathcal{M}'_k(\omega)} - n(\alpha-1)\log^2(1+2|A|) - \frac{\alpha}{\alpha-1} n \log\left(1+\delta\left(e^{\frac{\alpha-1}{\alpha}2\log(|A||B|)}-1\right)\right). \tag{3.64}$$

Putting Eq. 3.58, Eq. 3.59, and Eq. 3.64 together, we have

$$\begin{aligned} H^{\epsilon_1+\epsilon_2}_{\min}(A_1^n|B_1^nE)_{\rho} \geq \sum_{k=1}^{n} \inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} & H(A_k|B_k\tilde{R}_{k-1})_{\mathcal{M}'_k(\omega)} - n(\alpha-1)\log^2(1+2|A|) \\ & - \frac{\alpha}{\alpha-1} n \log\left(1+\delta\left(e^{\frac{\alpha-1}{\alpha}2\log(|A||B|)}-1\right)\right) \\ & - \frac{\alpha}{\alpha-1} n z_{\beta}(\epsilon,\delta) \frac{1}{\alpha-1}\left(g_1(\epsilon_2,\epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta-1}\right). \end{aligned}$$

□

3.4.5. Testing for approximate EAT

We will now incorporate testing into the approximate entropy accumulation theorem proven in Theorem 3.12. Testing enables one to prove a lower bound on the smooth min-entropy of a state produced by the process in Fig. 2.1 conditioned on the output of a classical event. This is particularly useful for proving tight and practical bounds in cryptographic protocols.

In this section, we will consider the channels $\mathcal{M}_k$ and $\mathcal{M}'_k$ which map registers R_{k-1} to $A_kB_kX_kR_k$ such that X_k is a classical value which is determined using the registers A_k and B_k. Concretely, suppose that for every k, there exist a channel $\mathcal{T}_k : A_kB_k \to A_kB_kX_k$ of the form

$$\mathcal{T}_k(\omega_{A_kB_k}) = \sum_{y,z} \Pi^{(y)}_{A_k} \otimes \Pi^{(z)}_{B_k} \omega_{A_kB_k} \Pi^{(y)}_{A_k} \otimes \Pi^{(z)}_{B_k} \otimes |x(y,z)\rangle\langle x(y,z)|_{X_k} \tag{3.65}$$

where $\{\Pi^{(y)}_{A_k}\}_y$ and $\{\Pi^{(z)}_{B_k}\}_z$ are orthogonal projectors and $x(\cdot)$ is some deterministic function which uses the measurements y and z to create the output register X_k.

In order to define the min-tradeoff functions, we let $\mathbb{P}$ be the set of probability distributions over the alphabet of X registers. Let R be any register isomorphic to R_{k-1}. For a probability $q \in \mathbb{P}$ and a channel $\mathcal{N}_k : R_{k-1} \to A_kB_kX_kR_k$, we also define the set

$$\Sigma_k(q|\mathcal{N}_k) := \{\nu_{A_kB_kX_kR_kR} = \mathcal{N}_k(\omega_{R_{k-1}R}) : \text{ for a state } \omega_{R_{k-1}R} \text{ such that } \nu_{X_k} = q\}. \tag{3.66}$$

Definition 3.17. *A function $f : \mathbb{P} \to \mathbb{R}$ is called a min-tradeoff function for the channels $\{\mathcal{N}_k\}_{k=1}^n$ if for every $k \in [n]$, it satisfies*

$$f(q) \leq \inf_{\nu \in \Sigma_k(q|\mathcal{N}_k)} H(A_k|B_kR)_\nu. \tag{3.67}$$

We will also need the definitions of the following simple properties of the min-tradeoff functions for our entropy accumulation theorem:

$$\mathrm{Max}(f) \coloneqq \max_{q \in \mathbb{P}} f(q) \tag{3.68}$$

$$\mathrm{Min}(f) \coloneqq \min_{q \in \mathbb{P}} f(q) \tag{3.69}$$

$$\mathrm{Min}_\Sigma(f) \coloneqq \min_{q : \Sigma(q) \neq \varnothing} f(q) \tag{3.70}$$

$$\mathrm{Var}(f) \coloneqq \max_{q : \Sigma(q) \neq \varnothing} \sum_x q(x) f(\delta_x)^2 - \left(\sum_x q(x) f(\delta_x) \right)^2 \tag{3.71}$$

where $\Sigma(q) \coloneqq \bigcup_k \Sigma_k(q)$ and δ_x is the distribution with unit weight on the alphabet x.

Theorem 3.18. *For $k \in [n]$, let the registers A_k and B_k be such that $|A_k| = |A|$ and $|B_k| = |B|$. For $k \in [n]$, let $\mathcal{M}_k$ be channels from $R_{k-1} \to R_k A_k B_k X_k$ and*

$$\rho_{A_1^n B_1^n X_1^n E} = \mathrm{tr}_{R_n} \circ \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \tag{3.72}$$

be the state produced by applying these maps sequentially. Further, let $\mathcal{M}_k$ be such that $\mathcal{M}_k = \mathcal{T}_k \circ \mathcal{M}_k^{(0)}$ for $\mathcal{T}_k$ defined in Eq. 3.65 and some channel $\mathcal{M}_k^{(0)} : R_{k-1} \to R_k A_k B_k$. Suppose the channels $\mathcal{M}_k$ are such that for every $k \in [n]$, there exists a channel $\mathcal{M}'_k$ from $R_{k-1} \to R_k A_k B_k X_k$ such that

(1) $\mathcal{M}'_k = \mathcal{T}_k \circ \mathcal{M}_k'^{(0)}$ for some channel $\mathcal{M}_k'^{(0)} : R_{k-1} \to R_k A_k B_k$.

(2) $\mathcal{M}'_k$ ϵ-approximates $\mathcal{M}_k$ in the diamond norm:

$$\frac{1}{2} \left\| \mathcal{M}_k - \mathcal{M}'_k \right\|_\diamond \leq \epsilon \tag{3.73}$$

(3) For every choice of a sequence of channels $\mathcal{N}_i \in \{\mathcal{M}_i, \mathcal{M}'_i\}$ for $i \in [k-1]$, the state $\mathcal{M}'_k \circ \mathcal{N}_{k-1} \circ \cdots \circ \mathcal{N}_1(\rho^{(0)}_{R_0 E})$ satisfies the Markov chain

$$A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k. \tag{3.74}$$

Then, for an event Ω defined using X_1^n, an affine min-tradeoff function f for $\{\mathcal{M}'_k\}_{k=1}^n$ such that for every $x_1^n \in \Omega$, $f(freq(x_1^n)) \geq h$, for parameters $0 < \delta, \epsilon_1, \epsilon_3 < 1$ and $\epsilon_2 \coloneqq 2\sqrt{\frac{\epsilon_1}{\Pr_\rho(\Omega)}}$ such that $\epsilon_2 + \epsilon_3 < 1$, $\alpha \in (1,2)$, and $\beta > 1$, we have

$$\begin{aligned} H_{\min}^{\epsilon_2+\epsilon_3}(A_1^n|B_1^n E)_{\rho|\Omega} \geq nh &- \frac{(\alpha-1)\log(2)}{2} \left(\log(2|A|^2+1) + \log(2)\sqrt{2 + Var(f)} \right)^2 - n(\alpha-1)^2 K_\alpha \\ &- \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(e^{\frac{\alpha-1}{\alpha} 2(\log(|A||B|) + Max(f) - Min(f) + 1)} - 1\right)\right) \end{aligned}$$

$$- \frac{\alpha}{\alpha - 1} n z_\beta(\epsilon, \delta) - \frac{1}{\alpha - 1} \left(\alpha \log \frac{1}{\Pr_\rho(\Omega) - \epsilon_1} + g_1(\epsilon_3, \epsilon_2) + \frac{\alpha g_0(\epsilon_1)}{\beta - 1} \right). \tag{3.75}$$

where

$$z_\beta(\epsilon, \delta) := \frac{\beta + 1}{\beta - 1} \log \left(\left(1 + \sqrt{(1-\delta)\epsilon}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\beta} \right)^{\frac{1}{\beta+1}} \right) \tag{3.76}$$

$$K_\alpha := \frac{1}{6(2-\alpha)^3} e^{(\alpha-1)(2\log|A| + (Max(f) - Min_\Sigma(f)))} \log^3 \left(e^{(2\log|A| + (Max(f) - Min_\Sigma(f)))} + e^2 \right) \tag{3.77}$$

and $g_1(x,y) = -\log(1 - \sqrt{1-x^2}) - \log(1-y^2)$.

We provide a proof for this theorem in Appendix A.8.

3.4.6. Limitations and further improvements

As we pointed out previously, the dependence of the entropy loss per round on ϵ is very poor (behaves as $\sim \epsilon^{1/24}$) in Theorem 3.12. The classical version of this theorem has a much better dependence of $O(\sqrt{\epsilon})$ on ϵ (see Theorem A.13). The reason for the poor performance of the quantum version is that our bound on the channel divergence (Lemma 3.15) is very weak compared to the bound we can use classically. It should be noted, however, that if Lemma 3.15 were to be improved in the future, one could simply plug the new bound into our proof and derive an improvement for Theorem 3.12.

A better bound on the channel divergence would have an additional benefit. It could simplify the proof and the Markov chain assumption in our theorem. In particular, it would be much easier to carry out the proof if the mixed channels $\mathcal{M}_k^\delta$ were defined as $(1-\delta)\,\mathcal{M}_k' + \delta \tau_{A_k B_k} \otimes \mathrm{tr}_{A_k B_k} \circ \mathcal{M}_k$ (which is what is done classically), where $\tau_{A_k B_k}$ is the completely mixed state on registers $A_k B_k$. Here, instead of mixing the channel $\mathcal{M}_k'$ with $\mathcal{M}_k$, we mix it with $\tau_{A_k B_k} \otimes \mathrm{tr}_{A_k B_k} \circ \mathcal{M}_k$, which also keeps $D_{\max}(\mathcal{M}_k \,||\, \mathcal{M}_k^\delta)$ small enough. Moreover, this definition ensures that the registers B_k produced by the map $\mathcal{M}_k^\delta$ always satisfy the Markov chain conditions. If it were possible to show that the divergence between the real state $\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})$ and the auxiliary state $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0 E}^{(0)})$ is small for this definition of $\mathcal{M}_k^\delta$, then one could directly use the entropy accumulation theorem for lower bounding the entropy for the auxiliary state. We cannot do this in our proof as this definition of the mixed channel $\mathcal{M}_k^\delta$ also increases the distance from the original channel $\mathcal{M}_k$ to $\epsilon + 2\delta$ and this makes the upper bound in Lemma 3.14 large (finite even in the limit $\epsilon \to 0$).

It seems that it should be possible to weaken the assumptions for approximate entropy accumulation. The classical equivalent of this theorem (Theorem A.13) for instance can be proven very easily and requires a much weaker approximation assumption. It would be interesting if one could remove the "memory" registers R_k from the assumptions required for approximate entropy accumulation, since these are not typically accessible to the users in applications.

Another troubling feature of the approximate entropy accumulation theorem seems to be that it assumes that the size of the side information registers B_k is constant. One might wonder if this is necessary, since continuity bounds like the Alicki-Fannes-Winter (AFW) inequality do not depend on the size of the side information. It turns out that a bound on the side information size is indeed necessary in this case. We show a simple classical example to demonstrate this in Appendix A.5. The necessity of such a bound also rules out a similar approximate extension of the generalised entropy accumulation theorem (GEAT) [**MFSR24**].

Chapter 4

Proving security of BB84 under source correlations

4.1. Introduction

Protocols in quantum cryptography often require an honest party to produce multiple independent quantum states. As an example, quantum key distribution (QKD) protocols [BB84, Ben92] and bit commitment protocols [KWW12, LMT20] all require the honest participant, Alice to produce an independently and randomly chosen quantum state from a set of states in every round of the protocol. The security proofs for these protocols also rely on the fact that the quantum state produced in each round of the protocol is independent of the other rounds. However, this is a difficult property to enforce practically. All physical devices have an internal memory, which is difficult to characterise and control. This memory can cause the quantum states produced in different rounds to be correlated with one another. For example, when implementing BB84 states using the polarisation of light, if the polariser is in the horizontal polarisation ($|0\rangle$) for round k, and it is switched to the $\Pi/4$-diagonal polarisation ($|+\rangle$) in the $(k+1)^{\text{th}}$ round, then it is plausible that the state produced in the $(k+1)^{\text{th}}$ round is "tilted" towards the horizontal (that is, has a larger component along $|0\rangle$ than $|1\rangle$) simply due to the inertia of switching the polariser. Such correlations between different rounds caused by an imperfect source are called *source correlations*. Security proofs for cryptographic protocols need to consider such correlations in order to be relevant in the real world.

An extensive line of research has led to techniques for proving the security of QKD protocols with a perfect source [SP00, Ren06, Koa09, TL17, DFR20, MR22]. However,

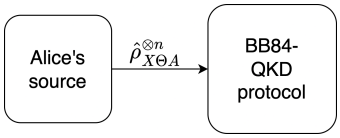

Fig. 4.1. Quantum input for the BB84 protocol with a perfect source.

almost all of these techniques rely on *source purification*[(1)]– the fact that the security of this protocol is equivalent to one where Alice sends out one half of a Bell state in each round and randomly measures her half. When the states produced by Alice's source are correlated across different rounds, this equivalence step fails and one can no longer use these methods. In this chapter, we use the entropic triangle inequality to reduce the security of the BB84 QKD protocol with source correlations to that of the BB84 protocol with a perfect source. With this reduction, one can simply use one of the many security analysis methods developed to complete the security proof[(2)]. We demonstrate our technique using the BB84 protocol, although we believe it is quite general and can be applied to other cryptographic protocols as well.

In the BB84 protocol, the only quantum state to the protocol is provided by Alice. The protocol can be represented as in Fig. 4.1. If the source is imperfect and the BB84 protocol is directly performed on the state produced by such a source as in Fig. 4.1, it is difficult to analyse the protocol and provide good security guarantees[(3)]. Instead, we propose and analyse the setup presented in Fig. 4.2. Here the source is tested during the execution of the protocol using a simple procedure and the protocol run is only declared valid if the test passes. The source test randomly measures the output of the source on a small sample of the rounds in the preparation basis and aborts if the relative deviation of the observed output from the expected output is more than some small threshold ϵ. Practically, this test

[(1)]Only [MR22] does not use source purification, but it still requires that the states in each round be produced independently.

[(2)] Following the reduction, any security proof technique for QKD which can bound $\tilde{H}_\alpha^\uparrow$ of Alice's raw key given Eve's side information can be used to complete the proof. The assumptions for the security of the protocol will be a combination of the assumptions required for this security proof and the assumptions used during the source test presented in Protocol 4.2.

[(3)] The following example demonstrates the difficulty. Imagine a source which at the start of the QKD protocol flips a coin C, which is 0 with probability ϵ_s. If $C = 1$, the source produces the qubit states perfectly, otherwise if $C = 0$ it encodes 0 whenever a key generation basis is used. The state produced by this source will be $O(\epsilon_s)$ close to the perfect state in each round. It will also not abort during parameter estimation. However, with probability ϵ_s no key is produced between Alice and Bob. In this situation, we would like the protocol's secrecy error to remain arbitrarily small and its abort probability to be $\approx \epsilon_s$. For this we need to be able to somehow identify the $C = 0$ bad case.

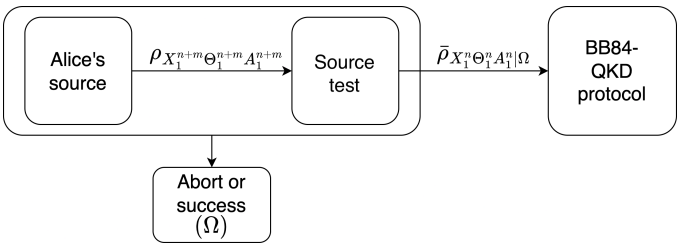

Fig. 4.2. The setup for performing the BB84 protocol with a source test.

can be carried out concurrently with the BB84 protocol and no quantum memory is required.

Roughly speaking, the output of the source test has at most ϵ error in each round. This scenario can be viewed through the framework of approximation chains: after the source test, one can show that almost every partial state $\rho_{X_1^k\Theta_1^kA_1^k|\Omega}$ can be approximated by a state of the form $\rho_{X_1^{k-1}\Theta_1^{k-1}A_1^{k-1}|\Omega} \otimes \hat{\rho}_{X\Theta A}$, where $\hat{\rho}_{X\Theta A}$ is the perfect state. While one can formalise this connection, we do not pursue it as it does not yield additional insights. The randomised testing actually ensures that the deviations in a round are not correlated with those in the other rounds (as they are in the example in Footnote (3)). This allows us to achieve an arbitrary smoothing parameter (and protocol error), in contrast to the analysis for the approximately independent registers problem in the previous chapter. This is crucial for practical QKD protocols.

For the security analysis of the protocol depicted in Fig. 4.2, we do not require any additional assumptions on the source. Assumptions are only required on the measurements used for the source test. In Section 4.3, we present the security analysis assuming perfect measurements and then in Section 4.4, we demonstrate how this analysis can be modified to incorporate imperfect measurements. It is worth noting that these measurements are used at a much smaller rate than the source, so it should be easier to implement them almost perfectly than it is to do the same for the source. In comparison, [**PCLN+22**], which is one of the most comprehensive treatments of source imperfections and source correlation, makes multiple complex assumptions about Alice's source (also see [**CLPK+23**]). Among these, it assumes that the state produced by Alice in the k^{th} round can only be correlated to the states produced in the ℓ_c rounds preceding it, where ℓ_c is some known constant. Moreover, it also assumes that Alice's quantum states are not entangled across different rounds. These are both, as noted by the authors in [**PCLN+22**], very strong assumptions, which cannot be guaranteed in practical setups. Importantly, it is also not possible to accurately estimate the parameter ℓ_c experimentally.

The source test also addresses the challenge of characterising the source for QKD [MST+19, KZF19, HML+23]. Most theoretical descriptions of QKD protocol require the source to operate almost perfectly. Thus, in order to implement these protocols one needs to characterise the source beforehand. Since we show security of BB84 as long as the source test succeeds, no prior characterisation is required for the source in our protocol. However, one still needs to characterise the measurements used in the source test. As mentioned above, this might be easier since the measurements are used at a much smaller rate.

4.2. Quantum sampling

In order to use the entropic triangle inequality effectively, we need a way to bound the $D^{\epsilon}_{\max}$ between the output of the source test and the almost perfect source state. We will use results from [BF10] for this task. [BF10] studies how sampling techniques can be used to estimate the *relative weight* of a string classically as well as quantumly. In particular, it essentially generalises the Hoeffding-Serfling random sampling bounds to quantum states. The main result of this paper has been summarised as the theorem below. To state it, we first need to define the relative weight of a string. For an alphabet $\mathcal{X}$ and a string $x_1^n \in \mathcal{X}^n$, the relative weight of x_1^n is the frequency of non-zero x_i, that is,

$$\omega(x_1^n) := \frac{1}{n}|\{i \in [n] : x_i \neq 0\}|. \tag{4.1}$$

Further, for a string x_1^n and a subset $S \subseteq [n]$, let x_S refer to the string $(x_i)_{i\in S}$.

Theorem 4.1 (Quantum sampling [BF10])**.** *Let Ψ be a sampling strategy which takes a string a_1^n, selects a random subset $\Gamma \subseteq [n]$ using the probability distribution p_Γ, a random seed K with probability p_K and produces an estimate $f(\Gamma, a_\Gamma, K)$ for the relative weight of the rest of the string $a_{\bar{\Gamma}}$. We can define the set of strings for which this strategy provides a δ-correct estimate for $\delta > 0$ given the choices $\Gamma = \gamma$ and $K = \kappa$ as*

$$B^{\delta}_{\gamma\kappa}(\Psi) := \{a_1^n : |\omega(a_{\bar{\gamma}}) - f(\gamma, a_\gamma, \kappa)| < \delta\} \tag{4.2}$$

where $\bar{\gamma}$ is the complement of the set γ in $[n]$. The classical maximum error probability for this strategy Ψ is defined as

$$\epsilon^{\delta}_{cl} := \max_{a_1^n} \Pr_{\Gamma K}[a_1^n \notin B^{\delta}_{\Gamma K}(\Psi)]. \tag{4.3}$$

Define the projectors $\Pi^{\delta|\gamma\kappa}_{A_1^n} := \sum_{a_1^n \in B^{\delta}_{\gamma\kappa}(\Psi)} |a_1^n\rangle\langle a_1^n|_{A_1^n}$. Then, for a quantum state $\rho_{A_1^n E}$, we have that the state

$$\tilde{\rho}_{\Gamma K A_1^n E} := \sum_{\gamma\kappa} p(\gamma\kappa) |\gamma\kappa\rangle\langle\gamma\kappa|_{\Gamma K} \otimes \frac{\Pi^{\delta|\gamma\kappa}_{A_1^n} \rho_{A_1^n E} \Pi^{\delta|\gamma\kappa}_{A_1^n}}{\operatorname{tr}\left(\Pi^{\delta|\gamma\kappa}_{A_1^n} \rho_{A_1^n E}\right)} \tag{4.4}$$

is $\epsilon_{qu}^{\delta} = \sqrt{\epsilon_{cl}^{\delta}}$ *close to the state* $\rho_{\Gamma K A_1^n E} := \sum_{\gamma\kappa} p(\gamma\kappa) |\gamma\kappa\rangle\langle\gamma\kappa|_{\Gamma K} \otimes \rho_{A_1^n E}$ *in trace distance.*

If one were to measure the string in the register given by A_γ of the state $\tilde{\rho}$ defined above, then the rest of the registers $A_{\bar{\gamma}}$ of $\tilde{\rho}$ would lie in a subspace, which has relative weight δ-close to $f(\gamma, a_\gamma, \kappa)$.

4.3. Security proof for BB84 with source correlations

We consider the BB84-QKD protocol described in Protocol 4.1 (based on [DFR20] and [MR22]). This is slightly different from the protocol in Protocol 2.1 to ensure compatibility with the entropy accumulation theorem, which we use for the security proof. In Table 4.1, we list all the variables we use for our proof along with their definitions.

At the beginning of every round of the QKD protocol, Alice prepares the classical registers X_i and Θ_i, and the corresponding qubit in the register A_i. If Alice's quantum source were perfect, she would produce the following state during each round of the protocol

$$\hat{\rho}_{X\Theta A} := \sum_{x\in\mathcal{X},\theta\in\Theta} p(x,\theta) |x,\theta\rangle\langle x,\theta|_{X\Theta} \otimes H^\theta |x\rangle\langle x|_A H^\theta \tag{4.5}$$

where H is the Hadamard gate and

$$p(x,\theta) = \begin{cases} \frac{1-\mu}{|\mathcal{X}|} & \text{if } \theta = 0 \\ \frac{\mu}{|\mathcal{X}|} & \text{if } \theta = 1. \end{cases}$$

Consider the case, where Alice only has access to an imperfect quantum source to prepare qubits for the QKD protocol above. We will assume here that the classical randomness used by Alice is perfect. We do not place any assumptions on the performance of the source. Suppose Alice and Bob use n rounds for the BB84 protocol. In order, to perform the QKD protocol with the imperfect source, we require that Alice uses her source to first perform the source test given in Protocol 4.2 with $(n+m)$-total rounds. This test randomly selects m rounds of the source output, measures the qubit A_i in the basis given by Θ_i and compares the result with the encoded bit X_i for these rounds. The source passes the test if the fraction of errors is less than ϵ, which is a source error threshold chosen by Alice. Subsequently, Alice uses the n remaining rounds produced by the source for the BB84 protocol provided the source test does not abort. The complete protocol is depicted in Fig. 4.2. It should be noted that Alice can actually run Protocol 4.2 concurrently with the BB84 protocol. She does not need to create all the $(n+m)$-rounds at once and store them in a memory in order to carry out this protocol. She can classically sample a random set Γ of size m at the start of the BB84 protocol and for every round i, she can use the round as source test round if $i \in \Gamma$ or forward the state produced to Bob if $i \notin \Gamma$. For theoretical purposes, this concurrent

BB84 QKD protocol

Parameters:

- n is the number of qubits sent by Alice.
- $\mu \in (0,1)$ is the probability of both encoding and measuring in the X basis $\{|+\rangle, |-\rangle\}$.
- $e \in (0, \frac{1}{2})$ is the maximum error tolerated.
- $r \in (0,1)$ is the key rate of the protocol.

Protocol:

(1) For every $1 \leq i \leq n$ perform the following steps:
   (a) Alice chooses a random bit $X_i \in_R \{0,1\}$ and with probability $1-\mu$ encodes it in the Z basis and with probability μ in the X basis.
   (b) Alice sends her encoded qubit to Bob.
   (c) Bob measures the qubit in the Z basis with probability $1-\mu$ and in the X basis with probability μ. He records the output as Y_i.

(2) **Sifting:** Alice and Bob share their choice of bases for all the rounds and discard the rounds where their choices are different. We denote the remaining rounds by the set S.

(3) **Information reconciliation:** Alice and Bob use an information reconciliation procedure, which lets Bob obtain a guess $\hat{X}_S$ for Alice's raw key X_S.

(4) **Raw key validation:** Alice selects a random hash function from a 2-universal family and sends it along with $\text{hash}(X_S)$ to Bob. If $\text{hash}(X_S) \neq \text{hash}(\hat{X}_S)$ Bob aborts the protocol.

(5) **Parameter estimation:** Let S_X be the set of rounds where Alice prepared the qubit in the X basis and Bob measured the qubit in X basis. Bob aborts if $|\{i \in S_X : \hat{X}_i \neq Y_i\}| > e\mu^2 n$.

(6) **Privacy Amplification:** Alice chooses a random function F from a set of 2-universal hash functions from $|S|$ bits to $\lfloor rn \rfloor$ bits and announces it Bob. Alice and Bob compute the final key as $F(X_S)$ and $F(\hat{X}_S)$ respectively.

Protocol 4.1

approach is equivalent to one where Alice begins by using her source to produce all the $(n+m)$-rounds and for our arguments we assume this is the case. In this section, we assume that the measurements used in the source test are perfect; we will lift this assumption in Section 4.4.

Variable	Definition
$\mathcal{X}$	The set {0,1}; alphabet for Alice's random string.
Θ	The set {0,1}; alphabet for the basis string.
X_1^n	The random string chosen by Alice at the beginning of the protocol.
Θ_1^n	Alice's choice of randomly chosen basis. $\Theta_i = 0$ if Alice chooses Z basis and $\Theta_i = 1$ if she chooses X basis.
A_1^n	The quantum registers sent by Alice to Bob.
$\hat{\Theta}_1^n$	Bob's choice of randomly chosen basis. $\hat{\Theta}_i = 0$ if Bob chooses Z basis and $\hat{\Theta}_i = 1$ if he chooses X basis.
Y_1^n	Bob's outcomes of measuring A_1^n in $\hat{\Theta}_1^n$ basis.
S	The set $\{i \in [n] : \hat{\Theta}_i = \Theta_i\}$.
$\hat{X}_S$	Bob's guess of X_S, produced at the end of the information reconciliation step.
T	Transcript for information reconciliation and raw key validation.
$\bar{X}_1^n$	For $i \in [n]$, $\bar{X}_i = X_i$ if $\Theta_i = \hat{\Theta}_i$ else $\bar{X}_i = \perp$.
$\bar{Y}_1^n$	For $i \in [n]$, $\bar{Y}_i = Y_i$ if $\Theta_i = \hat{\Theta}_i = 1$ else $\bar{Y}_i = \perp$.
C_1^n	For $i \in [n]$, $C_i = X_i \oplus Y_i$ if $\Theta_i = \hat{\Theta}_i = 1$ else $C_i = \perp$.
$\hat{C}_1^n$	For $i \in [n]$, $\hat{C}_i = \hat{X}_i \oplus Y_i$ if $\Theta_i = \hat{\Theta}_i = 1$ else $\hat{C}_i = \perp$.
E	Eve's register created after Eve processes and forwards the states A_1^n to Bob.
Υ	The event that the protocol does not abort, i.e., $\lvert\{i \in S_X : \hat{C}_i = 1\}\rvert \leq e\mu^2 n$ and $\text{hash}(X_S) = \text{hash}(\hat{X}_S)$.
Υ'	The event that $X_S = \hat{X}_S$.
Υ''	The event that $\lvert\{i \in S_X : C_i = 1\}\rvert \leq e\mu^2 n$.

Table 4.1. Definition of variables for QKD

Let $\rho_{X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$ be the state produced by the imperfect source, Ω denote the event that the source test (Protocol 4.2) does not abort and let the output of the source test protocol conditioned on Ω be the state $\bar{\rho}_{X_1^n\Theta_1^n A_1^n|\Omega}$ (or the subnormalised state $\bar{\rho}_{X_1^n\Theta_1^n A_1^n \wedge \Omega}$ depending on the context).

In the following lemma, we prove that $\bar{\rho}_{X_1^n\Theta_1^n A_1^n|\Omega}$ has a relatively small smooth max-relative entropy with a depolarised version of the perfect source using Theorem 4.1. We then use the entropic triangle inequality to reduce the security of the BB84 protocol with the imperfect source to that of a BB84 protocol, which uses this state as its source state.

Source test

Parameters:

- ϵ is the source error tolerated.
- m is the number of rounds on which the source is tested.
- n is the number of rounds produced by the source for use in the subsequent protocol.

Protocol:

(1) The source produces the state $\rho_{X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$.

(2) Choose a random subset $\Gamma \subseteq [n+m]$ of size m, measure the quantum registers A_i in the basis given by Θ_i and let the result be $\hat{X}_i$.

(3) Abort the protocol (and any subsequent protocols) if the observed error $\frac{1}{m}|\{k \in \Gamma : \hat{X}_k \neq X_k\}| > \epsilon$.

(4) Relabel the remaining registers from 1 to n and use them as the n registers for the subsequent protocol.

Protocol 4.2

Lemma 4.2. *Let ϵ be the threshold of the source test, $\delta \in (0,1)$ a small parameter, and let $\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)} := (1-2(\epsilon+\delta))\hat{\rho}_{X\Theta A} + 2(\epsilon+\delta)\hat{\rho}_{X\Theta} \otimes \tau_A$ where τ_A is the completely mixed state on the register A. For the state $\bar{\rho}_{X_1^n\Theta_1^n A_1^n|\Omega}$ produced by the source test conditioned on passing, we have that*

$$D_{\max}^{\epsilon_f}(\bar{\rho}_{X_1^n\Theta_1^n A_1^n|\Omega} \| \left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)}\right)^{\otimes n}) \leq nh(\epsilon+\delta) + \log \frac{1}{\Pr_\rho(\Omega) - \epsilon_{qu}^\delta} \tag{4.6}$$

where $\Pr_\rho(\Omega)$ is the probability of the event Ω when the testing procedure is applied to the state ρ, $h(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function and $\epsilon_f = 2\sqrt{\frac{\epsilon_{qu}^\delta}{\Pr_\rho(\Omega)}}$ for $\epsilon_{qu}^\delta = \sqrt{2}\exp\left(-\frac{n\delta^2}{2(n+2)}m\right)$.

Proof. Define the unitaries,

$$V_A^{x,\theta} := H^\theta X^x \tag{4.7}$$

$$V_{X\Theta A} := \sum_{x,\theta} |x,\theta\rangle\langle x,\theta|_{X\Theta} \otimes V_A^{x,\theta} \tag{4.8}$$

so that $V_{X\Theta A}|x,\theta\rangle|0\rangle$ gives the perfect encoding of the BB84 state given x and θ. We also define the state

$$\nu_{X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} := \bigotimes_{i=1}^{n+m} V_{X_i\Theta_i A_i}^\dagger \; \rho_{X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} \; \bigotimes_{i=1}^{n+m} V_{X_i\Theta_i A_i}. \tag{4.9}$$

Note that if ρ were perfectly encoded, then ν would be the state $\rho_{X_1^{n+m}\Theta_1^{n+m}} \otimes (|0\rangle\langle 0|)^{\otimes(n+m)}$. Let the register Γ represent the choice of the random subset for sampling following the notation in Theorem 4.1. The state produced by measuring the subset γ of the A registers of ν in the computational $(\{|0\rangle, |1\rangle\})$ basis can equivalently be produced by measuring the subset γ of the A registers of ρ in the basis given by the corresponding Θ registers, adding (mod 2) the corresponding X register to the result and applying the unitaries $V_{X\Theta A}$ on the remaining indices. Conditioning on the sampled qubits of ρ being incorrectly encoded at most an ϵ fraction of the rounds is equivalent to measuring the corresponding random subset of the qubits of ν in the computational basis and conditioning on the relative weight of the result being less than ϵ (up to unitaries on the remaining registers; formal expression is given in Eq. 4.19). Given this equivalence, we can simply work with the state ν and transform the results back to the state ρ at the end.

Using Theorem 4.1, we have that for every $x_1^{n+m}, \theta_1^{n+m}$ there exists $\eta_{\Gamma A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}}$ such that

$$\frac{1}{2}\left\|\nu_{\Gamma A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} - \eta_{\Gamma A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}}\right\|_1 \leq \epsilon_{\text{qu}}^{\delta} \tag{4.10}$$

and

$$\eta_{\Gamma A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} := \sum_{\gamma} p(\gamma)\,|\gamma\rangle\langle\gamma| \otimes \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} \tag{4.11}$$

where $p(\gamma)$ is the uniform distribution over all size m subsets of $[n+m]$, and the state $\eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}}$ satisfies

$$\eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} = \Pi^{\delta|\gamma}_{A_1^{n+m}}\eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}}\Pi^{\delta|\gamma}_{A_1^{n+m}} \tag{4.12}$$

for the projectors $\Pi^{\delta|\gamma}_{A_1^{n+m}}$ defined as in Theorem 4.1 (our sampling procedure does not require a random seed κ, so we omit it in our analysis). Note that using Hoeffding's bound the classical error probability for our sampling strategy is $2\exp(-\frac{n\delta^2}{n+2}m)$, which implies that $\epsilon_{\text{qu}}^{\delta} = \sqrt{2}\exp(-\frac{n\delta^2}{2(n+2)}m)$ for this strategy. We can also define the extended state $\eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$ as

$$\begin{aligned}\eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} := \sum_{\gamma,x_1^{n+m},\theta_1^{n+m}} & p(\gamma)p(x_1^{n+m},\theta_1^{n+m}) \\ & |\gamma, x_1^{n+m}, \theta_1^{n+m}\rangle\langle\gamma, x_1^{n+m}, \theta_1^{n+m}| \otimes \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}}\end{aligned} \tag{4.13}$$

where $p(x_1^{n+m},\theta_1^{n+m}) = \prod_{i=1}^{n+m} p(x_i,\theta_i)$. Since, $\nu_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$ and $\eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$ have the same distributions on X_1^{n+m} and Θ_1^{n+m}, we also have that

$$\frac{1}{2}\left\|\nu_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} - \eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}\right\| \leq \epsilon_{\text{qu}}^{\delta}. \tag{4.14}$$

Define Ω' to be the event that the result produced by measuring the subset of registers A_γ in the computational basis, where γ is given by the Γ register, has a relative weight less than ϵ. Let $\bar{\nu}_{\Gamma X_1^n \Theta_1^n A_1^n \wedge \Omega'}$ be the subnormalised state produced when the relative weight of the registers A_γ of $\nu_{\Gamma X_1^{n+m} \Theta_1^{n+m} A_1^{n+m}}$ is measured and conditioned on Ω', the registers X_γ and Θ_γ are traced over, and the remaining X, Θ and A registers are relabelled between 1 and n. Also, let $\bar{\eta}_{\Gamma X_1^n \Theta_1^n A_1^n \wedge \Omega'}$ be the subnormalised state produced when this same subnormalised channel is instead applied to $\eta_{\Gamma X_1^{n+m} \Theta_1^{n+m} A_1^{n+m}}$. Let us consider the action of this map on a general state $|\gamma\rangle\langle\gamma| \otimes \sigma^{(\gamma)}_{A_1^{n+m}}$, which satisfies the condition $\sigma^{(\gamma)}_{A_1^{n+m}} = \Pi^{\delta|\gamma}_{A_1^{n+m}} \sigma^{(\gamma)}_{A_1^{n+m}} \Pi^{\delta|\gamma}_{A_1^{n+m}}$. For such a state, we have

$$\sigma^{(\gamma)}_{A_1^{n+m}} = \sum_{a_1^{n+m}, \bar{a}_1^{n+m} \epsilon B_\gamma^\delta} \sigma^{(\gamma)}(a_1^{n+m}, \bar{a}_1^{n+m}) |a_1^{n+m}\rangle\langle \bar{a}_1^{n+m}| .$$

Let $\hat{P}_{A_1^m} := \sum_{a_1^m : \omega(a_1^m) \leq \epsilon} |a_1^m\rangle\langle a_1^m|$ be the (perfect) measurement operator for conditioning on the event Ω'. Then, the state after applying the measurement and conditioning on the Ω' is

$$\mathrm{tr}_{A_\gamma}\left(\hat{P}_{A_\gamma} \sigma^{(\gamma)}_{A_1^{n+m}}\right) = \sum_{a_\gamma : \omega(a_\gamma) \leq \epsilon} \sum_{\substack{a'_{\bar{\gamma}}, \bar{a}_{\bar{\gamma}} \epsilon \{x_1^n : \\ |\omega(x_1^n) - \omega(a_\gamma)| < \delta\}}} \sigma^{(\gamma)}(a_\gamma a'_{\bar{\gamma}}, a_\gamma \bar{a}_{\bar{\gamma}}) |a'_{\bar{\gamma}}\rangle\langle \bar{a}_{\bar{\gamma}}| .$$

We can relabel the remaining registers to get the state $\bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'}$ which can be put into the form

$$\bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'} = \sum_{a_1^n, \bar{a}_1^n : \epsilon \{x_1^n : \ \omega(x_1^n) < \epsilon + \delta\}} \bar{\sigma}^{(\gamma)}(a_1^n, \bar{a}_1^n) |a_1^n\rangle\langle \bar{a}_1^n| . \tag{4.15}$$

Let $Q^w_{A_1^n}$ be the projector on the set $\mathrm{span}\{|x_1^n\rangle : \text{ for } x_1^n \text{ such that } \omega(x_1^n) < w\}$ (note that these vectors are perpendicular). Then, we have that

$$\bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'} = Q^{\epsilon+\delta}_{A_1^n} \bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'} Q^{\epsilon+\delta}_{A_1^n} \tag{4.16}$$

which implies that $\bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'} \leq Q^{\epsilon+\delta}_{A_1^n}$, since $\bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'}$ is subnormalised.

By considering $\sigma^{(\gamma)}_{A_1^{n+m}} = \eta^{(\gamma)}_{A_1^{n+m} | x_1^{n+m} \theta_1^{n+m}}$, we see that $\bar{\eta}$ satisfies

$$\begin{aligned}
\bar{\eta}_{\Gamma X_1^n \Theta_1^n A_1^n \wedge \Omega'} &= \sum_{\gamma, x_1^n, \theta_1^n} p(\gamma) p(x_1^n \theta_1^n) |\gamma x_1^n \theta_1^n\rangle\langle \gamma x_1^n \theta_1^n| \otimes \bar{\eta}^{(\gamma)}_{A_1^n | x_1^n \theta_1^n \wedge \Omega'} \\
&\leq \sum_{\gamma, x_1^n, \theta_1^n} p(\gamma) p(x_1^n \theta_1^n) |\gamma x_1^n \theta_1^n\rangle\langle \gamma x_1^n \theta_1^n| \otimes Q^{\epsilon+\delta}_{A_1^n} \\
&= \rho_\Gamma \otimes \rho_{X\theta}^{\otimes n} \otimes Q^{\epsilon+\delta}_{A_1^n} .
\end{aligned}$$

Using the data processing inequality, we also have that

$$\frac{1}{2} \left\| \bar{\nu}_{\Gamma X_1^n \Theta_1^n A_1^n \wedge \Omega'} - \bar{\eta}_{\Gamma X_1^n \Theta_1^n A_1^n \wedge \Omega'} \right\|_1 \leq \epsilon^\delta_{\mathrm{qu}} \tag{4.17}$$

Let $\hat{\eta}_A^{(\epsilon+\delta)} := (1-\epsilon-\delta)\,|0\rangle\langle 0| + (\epsilon+\delta)\,|1\rangle\langle 1|$ or equivalently the state $\hat{\eta}_A^{(\epsilon+\delta)}$ is the classical probability distribution over $\{0,1\}$ which is 1 with probability $(\epsilon+\delta)$. For this distribution, a simple calculation shows that

$$\min_{z_1^n : \omega(z_1^n) < \epsilon+\delta} \langle z_1^n | \, (\hat{\eta}_A^{(\epsilon+\delta)})^{\otimes n} \, | z_1^n \rangle \geq e^{-nh(\epsilon+\delta)}$$

which implies that

$$Q_{A_1^n}^{\epsilon+\delta} \leq e^{nh(\epsilon+\delta)} (\hat{\eta}_A^{(\epsilon+\delta)})^{\otimes n}.$$

Thus, we have

$$\bar{\eta}_{X_1^n \Theta_1^n A_1^n \wedge \Omega'} \leq e^{nh(\epsilon+\delta)} \left(\rho_{X\Theta} \otimes \hat{\eta}_A^{(\epsilon+\delta)} \right)^{\otimes n}. \tag{4.18}$$

As noted earlier, the state produced by measuring the registers A_γ of ν in the computational basis is the same as the state produced by measuring the same registers on the real state ρ in the basis given by Θ_i, adding X_i to the result (mod 2), and transforming the remaining registers with $\bigotimes_{k=1}^n V_{X_i \Theta_i A_i}^\dagger$. Under this correspondence, we have that the state produced by the source test satisfies

$$\bar{\rho}_{X_1^n \Theta_1^n A_1^n \wedge \Omega} = \bigotimes_{i=1}^n V_{X_i \Theta_i A_i} \bar{\nu}_{X_1^n \Theta_1^n A_1^n \wedge \Omega'} \bigotimes_{i=1}^n V_{X_i \Theta_i A_i}^\dagger. \tag{4.19}$$

Further, for the state defined as

$$\begin{aligned}
\bar{\tilde{\rho}}_{X_1^n \Theta_1^n A_1^n \wedge \Omega} &:= \bigotimes_{i=1}^n V_{X_i \Theta_i A_i} \bar{\eta}_{X_1^n \Theta_1^n A_1^n \wedge \Omega'} \bigotimes_{i=1}^n V_{X_i \Theta_i A_i}^\dagger \\
&\leq e^{nh(\epsilon+\delta)} \bigotimes_{i=1}^n \left(V_{X_i \Theta_i A_i} \; \rho_{X_i \Theta_i} \otimes \hat{\eta}_{A_i}^{(\epsilon+\delta)} V_{X_i \Theta_i A_i}^\dagger \right) \\
&= e^{nh(\epsilon+\delta)} \left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)} \right)^{\otimes n}
\end{aligned} \tag{4.20}$$

where $\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)} := (1-2(\epsilon+\delta))\hat{\rho}_{X\Theta A} + 2(\epsilon+\delta)\hat{\rho}_{X\Theta} \otimes \tau_A$ for the completely mixed state τ_A on register A. Using Eq. 4.17, we also have

$$\frac{1}{2} \left\| \bar{\rho}_{X_1^n \Theta_1^n A_1^n \wedge \Omega} - \bar{\tilde{\rho}}_{X_1^n \Theta_1^n A_1^n \wedge \Omega} \right\|_1 \leq \epsilon_{\text{qu}}^\delta. \tag{4.21}$$

Following the argument in Lemma A.14, we can show that

$$\frac{1}{2} \left\| \bar{\rho}_{X_1^n \Theta_1^n A_1^n | \Omega} - \bar{\tilde{\rho}}_{X_1^n \Theta_1^n A_1^n | \Omega} \right\|_1 \leq \frac{2\epsilon_{\text{qu}}^\delta}{\Pr_\rho(\Omega)} \tag{4.22}$$

where $\Pr_\rho(\Omega) := \text{tr}\left(\bar{\rho}_{X_1^n \Theta_1^n A_1^n \wedge \Omega} \right)$ is the probability of the event Ω when the testing procedure is applied to the state ρ, and

$$\begin{aligned}
\bar{\tilde{\rho}}_{X_1^n \Theta_1^n A_1^n | \Omega} &\leq \frac{e^{nh(\epsilon+\delta)}}{\Pr_{\tilde{\rho}}(\Omega)} \left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)} \right)^{\otimes n} \\
&\leq \frac{e^{nh(\epsilon+\delta)}}{\Pr_\rho(\Omega) - \epsilon_{\text{qu}}^\delta} \left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)} \right)^{\otimes n}
\end{aligned} \tag{4.23}$$

where $\Pr_{\bar{\rho}}(\Omega) := \mathrm{tr}\left(\bar{\bar{\rho}}_{X_1^n\Theta_1^n A_1^n \wedge \Omega}\right)$ is defined similar to $\Pr_\rho(\Omega)$. Together these imply that

$$D^{\epsilon_f}_{\max}(\bar{\rho}_{X_1^n\Theta_1^n A_1^n|\Omega}\|\left(\hat{\rho}^{(\epsilon+\delta)}_{X\Theta A}\right)^{\otimes n}) \leq nh(\epsilon+\delta) + \log\frac{1}{\Pr_\rho(\Omega)-\epsilon^\delta_{\mathrm{qu}}} \tag{4.24}$$

where $\epsilon_f = 2\sqrt{\frac{\epsilon^\delta_{\mathrm{qu}}}{\Pr_\rho(\Omega)}}$. □

We now give an outline for bounding the smooth min-entropy for a BB84-QKD protocol, which uses an imperfect source. We give a complete formal proof in Appendix B.1. Let Φ_{QKD} be the CPTP map denoting the action of the entire QKD protocol on the source states produced by Alice. In order to prove security for QKD, informally speaking, it is sufficient to prove a linear lower bound for[(4)]

$$H^{\epsilon_f+\epsilon'}_{\min}(X_S|ET\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho}_{|\Omega})}.$$

Let us define the virtual state $\sigma_{X_1^n\Theta_1^n A_1^n} := \left(\hat{\rho}^{(\epsilon+\delta)}_{X\Theta A}\right)^{\otimes n}$. This state can be viewed as the state produced when each of the qubits produced by Alice is passed through a depolarising channel. Using Lemma 3.5, for an arbitrary $\epsilon' > 0$, we have

$$\begin{aligned} H^{\epsilon_f+\epsilon'}_{\min}(X_S|ET\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho}_{|\Omega})} &\geq \tilde{H}^\uparrow_\alpha(X_S|ET\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\sigma)} \\ &\quad - \frac{\alpha}{\alpha-1}D^{\epsilon_f}_{\max}(\Phi_{\mathrm{QKD}}(\bar{\rho}_{|\Omega})\|\Phi_{\mathrm{QKD}}(\sigma)) - \frac{g_1(\epsilon',\epsilon_f)}{\alpha-1} \\ &\geq \tilde{H}^\uparrow_\alpha(X_S|ET\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\sigma)} - \frac{\alpha}{\alpha-1}nh(\epsilon+\delta) - \frac{O(1)}{\alpha-1}. \end{aligned}$$

Thus, it is sufficient to bound the α-Rényi conditional entropy $\tilde{H}^\uparrow_\alpha(X_S|ET\Theta_1^n\hat{\Theta}_1^n)$ for the QKD protocol running on a noisy version of the perfect source. We can now simply use standard techniques developed for the security proofs of QKD to show a linear lower bound for this conditional entropy. In particular, source purification can be used for the source state σ. In Appendix B.1, we show how one can modify the security proof for BB84 based on entropy accumulation to get the following bound.

Theorem 4.3. *Suppose Alice uses the output of the source test (Protocol 4.2), with error threshold ϵ and any imperfect source as its input, as her source for the BB84 protocol. Let $\delta > 0$ and assume that $h(\epsilon+\delta) < \frac{1}{\sqrt{2}}$. Then, for*

$$\epsilon^\delta_{qu} = \sqrt{2}\exp\left(-\frac{n\delta^2}{2(n+2)}m\right) \tag{4.25}$$

$$\epsilon_{pa} = 2\left(\frac{2\epsilon^\delta_{qu}}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')}\right)^{1/2} \tag{4.26}$$

[(4)] We also need to condition on the QKD protocol not aborting. We do this in Appendix B.1

and $\epsilon' > 0$, we have the following lower bound on the smooth min-entropy for the raw key produced during the BB84 protocol

$$
\begin{aligned}
&H_{\min}^{\epsilon_{pa}+\epsilon'}(X_S|E\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{QKD}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\geq n((1-2\mu)\log(2) - h(e) - \mu^2(1-\log(2)) - V\sqrt{2h(\epsilon+\delta)}) \\
&- \sqrt{n}\left(\mu^2\ln(2) + 2\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + g_0\left(\frac{\epsilon'}{8}\right)\right) - \frac{V}{\sqrt{2h(\epsilon+\delta)}}\left(\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'') - 2\epsilon_{qu}^{\delta}} + 1\right) \\
&- \frac{g_1(\frac{\epsilon'}{2},\epsilon_{pa})}{2\sqrt{2h(\epsilon+\delta)}}V - \log|T| - 3g_0\left(\frac{\epsilon'}{8}\right)
\end{aligned} \tag{4.27}
$$

where $V := \frac{2}{\mu^2}\log\frac{1-e}{e} + 2\log(1+2|\mathcal{X}|^2)$, $\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')$ is the probability of the event $\Omega\wedge\Upsilon''$ for the state $\Phi_{QKD}(\bar{\rho})$ and it is assumed that $\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'') > 2\epsilon_{qu}^{\delta}$ [(5)], $g_0(x) = -\log(1-\sqrt{1-x^2})$ and $g_1(x,y) = -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$.

According to the Theorem above, the asymptotic key rate for the BB84 protocol using an imperfect source is $V\sqrt{2h(\epsilon+\delta)}$ lesser than a protocol, which uses a perfect source. There is a lot of room to improve the analysis used for the Theorem above (given in Appendix B.1). We use the simplest possible techniques to demonstrate a complete security proof.

4.4. Imperfect measurements

In our analysis above, we assumed that the measurements used in the source test are perfect. It should be noted that if the source produces states at a rate r_s, then the measurement device is only used at an average rate $\frac{m}{n+m}r_s$, which is much smaller than r_s. So, the measurement devices have a much longer relaxation time than the source. As such, it should be easier to create almost "perfect" measurement devices than it is to create perfect sources.

In this section, we will show how measurement imperfections can also be incorporated in our analysis. Let $\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma}$ be the POVM element associated with the source test passing, i.e., with measuring a relative weight less than ϵ with respect to the encoded random bits given the choice of random subset γ, encoded random bits x_γ, and basis choice θ_γ. Informally speaking, in this subsection, we assume that this measurement measures the relative weight with an error at most ϵ_m with high probability. To formally state our

[(5)] If $\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'') \leq 2\epsilon_{\text{qu}}^{\delta}$, one can easily show that the secrecy condition for QKD is satisfied for a security parameter greater than $2\epsilon_{\text{qu}}^{\delta}$ since this condition is weighted by the abort probability (Eq. 2.60).

assumption, define

$$\hat{P}_{A_\gamma}^{x_\gamma,\theta_\gamma} := \bigotimes_{i\in\gamma} V_{A_i}^{x_i,\theta_i} \left(\sum_{a_\gamma:\omega(a_\gamma)\leq\epsilon+\epsilon_m} |a_\gamma\rangle\langle a_\gamma|_{A_\gamma} \right) \bigotimes_{i\in\gamma} \left(V_{A_i}^{x_i,\theta_i}\right)^\dagger \tag{4.28}$$

$$\hat{P}_{A_\gamma}^{\perp|x_\gamma,\theta_\gamma} := \mathbb{1}_{A_1^n} - \hat{P}_{A_\gamma}^{x_\gamma,\theta_\gamma} \tag{4.29}$$

to be the projectors on the subspace with relative weight at most $\epsilon + \epsilon_m$, and at least $\epsilon + \epsilon_m$ with respect to x_γ in the basis θ_γ. Here the parameter ϵ is the same as the source error threshold in the previous section and $\epsilon_m > 0$ is a small parameter quantifying the measurement device error. The projector $\hat{P}_{A_\gamma}^{x_\gamma,\theta_\gamma}$ is the rotated version of projector $\hat{P}$, which was used for the measurement map in the previous section. In this section, we need to use the rotated version because the real measurements in an implementation will depend on the inputs γ, x_γ and θ_γ.

We assume that for some fixed small $\xi > 0$ the measurement elements $\{\Lambda\left(\leq \epsilon|\gamma, x_\gamma, \theta_\gamma\right)_{A_\gamma}\}_{\gamma,x_\gamma,\theta_\gamma}$ satisfy the following for every collection of states $\{\sigma_{A_\gamma|x_\gamma\theta_\gamma}^{(\gamma)}\}_{\gamma,x_\gamma,\theta_\gamma}$:

$$\sum_\gamma p(\gamma) \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}\left(\Lambda\left(\leq \epsilon|\gamma, x_\gamma, \theta_\gamma\right)_{A_\gamma} \hat{P}_{A_\gamma}^{\perp|x_\gamma,\theta_\gamma} \sigma_{A_\gamma|x_\gamma\theta_\gamma}^{(\gamma)} \hat{P}_{A_\gamma}^{\perp|x_\gamma,\theta_\gamma}\right) \leq \xi. \tag{4.30}$$

Stated in words, we require that for any collection of states $\{\sigma_{A_\gamma|x_\gamma\theta_\gamma}^{(\gamma)}\}_{\gamma,x_\gamma,\theta_\gamma}$ with a relative weight larger than $\epsilon + \epsilon_m$ (lying in the subspace corresponding to the projector $\hat{P}_{A_\gamma}^{\perp|x_\gamma,\theta_\gamma}$), the probability that a weight lesser than ϵ is measured is smaller than ξ when averaged over the choice of the random set γ and x_γ, θ_γ. Using this assumption on the measurements, in Lemma 4.4 we will derive a smooth max-relative entropy bound similar to the one in the previous section. The smoothing parameter of the relative entropy in this bound, however, will depend on ξ, which in turn implies that the privacy amplification error of the subsequent QKD protocol will be lower bounded by a function of ξ. It does not seem that this dependence of the smoothing parameter on ξ can be avoided. For example, if the measurements measure a small weight for a set of large weight states and the source emits those states, then they can be exploited by Eve to extract additional information during the QKD protocol. It also seems that we cannot use some kind of joint test for the source and measurement device (similar to Protocol 4.2) without an additional assumption to ensure that the weight measured by the measurement device is almost correct, since the source can always embed its information using an arbitrary unitary and the measurement can always decode that information using the same unitary.

I.I.D. measurements with error ϵ'_m or more generally measurements, which are guaranteed to measure each input qubit correctly with probability at least $(1-\epsilon'_m)$ independent of the previous rounds (both these examples consider measurements which measure the qubits A_γ in the provided basis θ_γ to produce the results $\hat{x}_\gamma$ and then use these results to test if $\omega(x_\gamma \oplus \hat{x}_\gamma) \leq \epsilon$ or not), satisfy the above assumption for the choice of some $\delta' > 0$, $\epsilon_m = \epsilon'_m + \delta'$ and $\xi = e^{-2m\delta'^2}$ (using the Chernoff-Hoeffding bound).

Additionally, since we average over the random set γ as well, it is possible to guarantee with high probability that for most test measurements the relaxation time of the measurement device is large. This should enable us to model a large and practical class of measurements using these assumptions. We leave the details for the specific measurement model for future work.

We will show that for measurements, which satisfy the above assumption the following lemma holds. One can use this bound in place Lemma 4.2 to prove a smooth min-entropy lower bound for the QKD protocol, similar to the previous section. Note that the following proof builds on the proof of Lemma 4.2 and makes use of definitions used in that proof.

Lemma 4.4. *Suppose that the measurements used for the source test $\{\Lambda(\leq \epsilon|\gamma, x_\gamma, \theta_\gamma)_{A_\gamma}\}_{\gamma,x_\gamma,\theta_\gamma}$ satisfy the assumption in Eq. 4.30 with parameters ϵ_m and ξ. Let ϵ be the threshold of the source test, $\delta \in (0,1)$ a small parameter, and let $\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A} := (1-2(\epsilon+\epsilon_m+\delta))\hat{\rho}_{X\Theta A} + 2(\epsilon+\epsilon_m+\delta)\hat{\rho}_{X\Theta}\otimes\tau_A$ where τ_A is the completely mixed state on the register A. Let the event Ω_{im} denote that the source test using the imperfect measurements succeeds and let state $\rho'_{X_1^n\Theta_1^n A_1^n|\Omega_{im}}$ denote the state produced by the source test conditioned on passing. For this state, we have that*

$$D^{\epsilon_f}_{\max}(\rho'_{X_1^n\Theta_1^n A_1^n|\Omega_{im}} \| \left(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A}\right)^{\otimes n}) \leq nh(\epsilon+\epsilon_m+\delta) + 2 + \log\frac{1}{\Pr_\rho(\Omega_{im}) - \epsilon^\delta_{qu}} + \log\frac{1}{4\xi(\Pr_\rho(\Omega_{im}) - \epsilon^\delta_{qu} - 4\xi)} \tag{4.31}$$

where $\Pr_\rho(\Omega_{im})$ is the probability of the event Ω_{im} when the testing procedure is applied to the state ρ, $h(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy function and $\epsilon_f := \frac{2\xi^{1/2}}{\sqrt{\Pr_\rho(\Omega_{im})-\epsilon^\delta_{qu}}} + 2\sqrt{\frac{\epsilon^\delta_{qu}}{\Pr_\rho(\Omega_{im})}}$ for $\epsilon^\delta_{qu} = \sqrt{2}\exp\left(-\frac{n\delta^2}{2(n+2)}m\right)$

Proof. For every x_1^{n+m} and θ_1^{n+m}, we define the following appropriately rotated versions of the projector $\Pi^{\delta|\gamma}_{A_1^{n+m}}$ given by Theorem 4.1, so that we can compare the relative weight with

the string x_1^{n+m} in the basis given by θ_1^{n+m}.

$$\bar{\Pi}^{\delta|\gamma,x_1^{n+m},\theta_1^{n+m}}_{A_1^{n+m}} := \bigotimes_{i=1}^{n+m} V_{A_i}^{x_i,\theta_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{i=1}^{n+m} (V_{A_i}^{x_i,\theta_i})^\dagger \tag{4.32}$$

where the unitaries $V_{A_i}^{x_i,\theta_i}$ are defined in Eq. 4.8. We use the state η from the previous section (Eq. 4.13) to define the state

$$\tilde{\rho}_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} := \bigotimes_{i=1}^{n+m} V_{X_i\Theta_i A_i} \eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} \bigotimes_{i=1}^{n+m} V^\dagger_{X_i\Theta_i A_i}. \tag{4.33}$$

Using the distance bound proven in Eq. 4.14 and the definition of ν in Eq. 4.9, we have

$$\frac{1}{2} \left\| \rho_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} - \tilde{\rho}_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} \right\|_1 \leq \epsilon_{\text{qu}}^\delta$$

The conditional states $\tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}}$ of the state $\tilde{\rho}$ above satisfy

$$\begin{aligned}
&\bar{\Pi}^{\delta|\gamma,x_1^{n+m},\theta_1^{n+m}}_{A_1^{n+m}} \tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \bar{\Pi}^{\delta|\gamma,x_1^{n+m},\theta_1^{n+m}}_{A_1^{n+m}} \\
&= \bigotimes_{i=1}^{n+m} V_{A_i}^{x_i,\theta_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{i=1}^{n+m} (V_{A_i}^{x_i,\theta_i})^\dagger \tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \bigotimes_{i=1}^{n+m} V_{A_i}^{x_i,\theta_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{i=1}^{n+m} (V_{A_i}^{x_i,\theta_i})^\dagger \\
&= \bigotimes_{i=1}^{n+m} V_{A_i}^{x_i,\theta_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{i=1}^{n+m} (V_{A_i}^{x_i,\theta_i})^\dagger \\
&= \bigotimes_{i=1}^{n+m} V_{A_i}^{x_i,\theta_i} \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \bigotimes_{j=1}^{n+m} (V_{A_i}^{x_i,\theta_i})^\dagger \\
&= \tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}}
\end{aligned}$$

where we have used the definition of $\tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}}$ (Eq. 4.33) in the second equality, and Eq. 4.12 for the fourth line.

We call the subnormalised state produced after performing the (imperfect) measurements on the states $\rho_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$, conditioning on the event Ω_{im} and tracing over the registers X_Γ and Θ_Γ as $\rho'_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega_{\text{im}}}$. Similarly, we let $\tilde{\rho}'_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega_{\text{im}}}$ denote the subnormalised state produced when this subnormalised map is applied to $\tilde{\rho}_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$. We have that

$$\begin{aligned}
\tilde{\rho}'_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega_{\text{im}}} = &\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle\left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}}\right) \\
\leq 2 &\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle\left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\left[\sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}} \hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma}\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad + \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\Lambda\left(\leq \epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right)\Bigg] \\
&\leq 2 \sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle\left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\Lambda\left(\leq \epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}} \hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma}\right) \\
&\qquad + 2\xi \mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}}
\end{aligned}
$$

where we have used the pinching inequality (see, for example [Tom16, Section 2.6.3]) in the second line, defined the state $\mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}}$ as the normalization of the state

$$
\begin{aligned}
&\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle\left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\Lambda\left(\leq \epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right)
\end{aligned}
$$

and used

$$
\begin{aligned}
&\operatorname{tr}\Bigg(\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle\left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\Lambda\left(\leq \epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right)\Bigg) \\
&= \sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}\left(\Lambda\left(\leq \epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right) \\
&= \sum_\gamma p(\gamma) \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}\left(\Lambda\left(\leq \epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\left(\sum_{x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \tilde{\rho}^{(\gamma)}_{A_\gamma|x_1^{n+m}\theta_1^{n+m}}\right) \hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right) \\
&\leq \xi,
\end{aligned}
$$

which follows from our assumption about the measurements (Eq. 4.30). Therefore, we have

$$
\begin{aligned}
\tilde{\rho}'_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega_{\text{im}}} &\leq 2 \sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle\left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\Lambda\left(\leq \epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma} \hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}} \hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma}\right) \\
&\qquad + 2\xi \mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}} \\
&\leq 2 \sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle\left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma) \operatorname{tr}_{A_\gamma}\left(\hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma} \tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}}\right) + 2\xi \mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}} \\
&\leq 2\tilde{\bar{\rho}}^{(\epsilon+\epsilon_m)}_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega} + 2\xi \mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}}
\end{aligned}
$$

where the state $\bar{\tilde{\rho}}^{(\epsilon+\epsilon_m)}_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega}$ is the state produced when the perfect measurement is used to measure A_γ and condition the state $\tilde{\rho}_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$ on the event that the relative weight of the measured results is lesser than $\epsilon+\epsilon_m$ from the string contained in X_γ. This is the state, which was used in the previous section to derive the smooth max-relative entropy bound. The only difference being that the threshold for the relative weight of the perfect measurement in the last section was ϵ. Thus, we can use the previously derived bound in Eq. 4.20 for this state by simply replacing ϵ with $\epsilon+\epsilon_m$. Relabelling the remaining registers between 1 and n, tracing over the Γ register and using the Eq. 4.20, we get

$$\begin{aligned}\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}} &\leq 2\bar{\tilde{\rho}}^{(\epsilon+\epsilon_m)}_{X_1^n\Theta_1^nA_1^n\wedge\Omega} + 2\xi\mu_{X_1^n\Theta_1^nA_1^n} \\ &\leq e^{nh(\epsilon+\epsilon_m+\delta)+1}\left(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A}\right)^{\otimes n} + 2\xi\mu_{X_1^n\Theta_1^nA_1^n} \end{aligned} \tag{4.34}$$

where $\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A} := (1-2(\epsilon+\epsilon_m+\delta))\hat{\rho}_{X\Theta A} + 2(\epsilon+\epsilon_m+\delta)\hat{\rho}_{X\Theta}\otimes\tau_A$. As before using the data processing inequality, we have

$$\frac{1}{2}\left\|\rho'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}} - \tilde{\rho}'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}}\right\|_1 \leq \epsilon^\delta_{\text{qu}}. \tag{4.35}$$

Once again following the argument in Lemma A.14, the conditional states satisfy

$$\frac{1}{2}\left\|\rho'_{X_1^n\Theta_1^nA_1^n|\Omega_{\text{im}}} - \tilde{\rho}'_{X_1^n\Theta_1^nA_1^n|\Omega_{\text{im}}}\right\|_1 \leq \frac{2\epsilon^\delta_{\text{qu}}}{\Pr_\rho(\Omega_{\text{im}})} \tag{4.36}$$

for $\Pr_\rho(\Omega_{\text{im}}) := \text{tr}\left(\rho'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}}\right)$, defined as the probability that the Protocol 4.2 does not abort with the imperfect measurements and

$$\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n|\Omega_{\text{im}}} \leq \frac{e^{nh(\epsilon+\epsilon_m+\delta)+1}}{\Pr_{\tilde{\rho}}(\Omega_{\text{im}})}\left(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A}\right)^{\otimes n} + \frac{4\xi}{\Pr_{\tilde{\rho}}(\Omega_{\text{im}})}\frac{\mu_{X_1^n\Theta_1^nA_1^n}}{2}. \tag{4.37}$$

where $\Pr_{\tilde{\rho}}(\Omega_{\text{im}}) := \text{tr}\left(\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}}\right)$. For $0<\mu<1$, the hypothesis testing relative entropy [**WR12**] is defined as

$$D_h^\mu(\rho\|\sigma) := -\inf\left\{\log\text{tr}(\sigma Q) : 0\leq\mu Q\leq\mathbb{1}\text{ , and }\text{tr}(\rho Q)\geq 1\right\}. \tag{4.38}$$

Equivalently, using semidefinite programming duality (see [**Wat20**]) it can be shown that

$$\begin{aligned} D_h^\mu(\rho\|\sigma) &= -\sup\left\{\log(\lambda-\text{tr}(Y)) : Y\geq 0, \lambda\geq 0, \text{ and } \lambda\rho\leq\sigma+\mu Y\right\} &(4.39)\\ &= \inf\left\{\log\lambda' - \log(1-\text{tr}(Z)) : Z\geq 0, \lambda'\geq 0, \text{ and } \rho\leq\lambda'\sigma+\mu Z\right\}. &(4.40)\end{aligned}$$

Thus, Eq. 4.37 implies

$$D_h^\mu(\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n|\Omega_{\text{im}}}\|(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A})^{\otimes n}) \leq nh(\epsilon+\epsilon_m+\delta) + 2 + \log\frac{1}{\Pr_{\tilde{\rho}}(\Omega_{\text{im}})} \tag{4.41}$$

for $\mu := \frac{4\xi}{\Pr_{\tilde{\rho}}(\Omega_{\text{im}})}$. Using [Wat20, Theorem 5.11] (originally proven in [ABJT19]), this implies that[6]

$$D_{\max}^{\sqrt{\mu}}(\tilde{\rho}'_{X_1^n\Theta_1^n A_1^n|\Omega_{\text{im}}}\|(\hat{\rho}_{X\Theta A}^{(\epsilon+\epsilon_m+\delta)})^{\otimes n}) \leq nh(\epsilon+\epsilon_m+\delta) + 2 + \log\frac{1}{\Pr_{\tilde{\rho}}(\Omega_{\text{im}})} + \log\frac{1}{\mu(1-\mu)} \tag{4.42}$$

Using the triangle inequality, we can state this in terms of the real state $\tilde{\rho}'_{X_1^n\Theta_1^n A_1^n|\Omega_{\text{im}}}$

$$\begin{aligned} D_{\max}^{\epsilon_f}(\rho'_{X_1^n\Theta_1^n A_1^n|\Omega_{\text{im}}}\|\left(\hat{\rho}_{X\Theta A}^{(\epsilon+\epsilon_m+\delta)}\right)^{\otimes n}) \leq nh(\epsilon+\epsilon_m+\delta) + 2 + \log\frac{1}{\Pr_\rho(\Omega_{\text{im}}) - \epsilon_{\text{qu}}^\delta} \\ + \log\frac{1}{4\xi(\Pr_\rho(\Omega_{\text{im}}) - \epsilon_{\text{qu}}^\delta - 4\xi)} \end{aligned} \tag{4.43}$$

for $\epsilon_f := \frac{2\xi^{1/2}}{\sqrt{\Pr_\rho(\Omega_{\text{im}})-\epsilon_{\text{qu}}^\delta}} + 2\sqrt{\frac{\epsilon_{\text{qu}}^\delta}{\Pr_\rho(\Omega_{\text{im}})}}$. Note that if $\xi = \exp(-\Omega(m))$, then the last term in the bound above adds $O(m)$ to the smooth max-relative entropy, so it cannot be chosen to be too small (This seems to be an artifact of the bound in [Wat20, Theorem 5.11], and it should be possible to improve this dependence). □

4.5. Discussion and future work

We demonstrated a general method to reduce the security of the BB84 protocol with an imperfect source with source correlations to that of the BB84 protocol with an almost perfect source. In order to minimise the rate loss and privacy amplification error, we used a source test to test the output of the imperfect source before using it for the QKD protocol. Theorem 4.3 gives a simple bound on the smooth min-entropy for the BB84 protocol which uses the output of the source test. According to this bound, for a source error of ϵ, the rate of the QKD protocol decreases by $O((\epsilon\log\frac{1}{\epsilon})^{1/2})$ and the privacy amplification error can be made arbitrarily small assuming perfect measurements are used for the source test. With imperfect measurements, satisfying a very broad assumption, we showed that the rate decrease is similar to the perfect case and the privacy amplification error depends on an error parameter of the measurements. This error parameter too can be made arbitrarily small under further reasonable physical assumptions, like independence of the measurement errors or almost perfect behaviour given a sufficient relaxation time. We leave the details of such a physical model and its relation to our assumption on the measurements for future work. It should be noted that one could also place physical assumptions on the source, which would guarantee that it passes the source test and hence imply security for the protocol. Further, if the source can be guaranteed to pass the source test with a high probability (which can

[6] The smoothing for $D_{\max}^{\epsilon}(\rho\|\sigma)$ in [Wat20] is defined using the trace distance instead of purified distance, which we use here. It can, however, be verified that the proof there also works with purified distance.

be made arbitrarily close to 1), say $1 - \epsilon_s$, then the source test need not even be performed before the QKD protocol. The error ϵ_s can simply be added to the QKD security parameter.

Chapter 5

Universal chain rules

5.1. Introduction

As mentioned in Chapter 1, we can decompose the von Neumann entropy of a large system A_1^n given B into a sum of the entropies of its parts as:

$$H(A_1^n|B)_\rho = \sum_{k=1}^{n} H(A_k|A_1^{k-1}B)_\rho. \tag{5.1}$$

However, it is easy to demonstrate that a smooth min-entropy counterpart for this relation cannot hold true. Mathematically, a relation of the following form cannot be valid for ϵ in a neighborhood of 0:

$$H_{\min}^{g_1(\epsilon)}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\epsilon}(A_k|A_1^{k-1}B)_\rho - ng_2(\epsilon) - k(\epsilon), \tag{5.2}$$

where the functions g_1, g_2, and k are dependent solely on ϵ and $|A|$ (the dimension of the A_k registers, assumed to be constant in n) and are independent of n. Furthermore, $g_1(\epsilon)$ and $g_2(\epsilon)$ are required to be *small* functions of ϵ, meaning they are continuous and approach 0 as ϵ tends to 0.

The impossibility of a bound of the form in Eq. 5.2 implies that $H_{\min}^{\epsilon}(A_1^n|B)_\rho$ cannot be lower bounded meaningfully in terms of the min-entropies of the approximation chain states of ρ, since these approximation chain states can simply be states satisfying $H_{\min}(A_k|A_1^{k-1}B)_{\sigma^{(k)}} = H_{\min}^{\epsilon}(A_k|A_1^{k-1}B)_\rho$ for every k. On the other hand, the von Neumann entropy of a state can easily be bounded in terms of its approximation chain by using the continuity of the conditional von Neumann entropy [**AF04**, **Win16**] to modify Eq. 5.1 and derive:

$$H(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H(A_k|A_1^{k-1}B)_{\sigma^{(k)}} - nf(\epsilon) \tag{5.3}$$

where $f(\epsilon) = O\left(\epsilon \log \frac{|A|}{\epsilon}\right)$. The absence of a comparable bound for the smooth min-entropy severely limits us. As we will see in the next chapter, it is often useful to initially prove novel results using von Neumann entropies before translating them into the corresponding one-shot entropies. This approach separates the complexity of the problem into two distinct phases. The impossibility of Eq. 5.2 prevents us from porting arguments based on the von Neumann entropy into smooth min-entropy arguments.

There are multiple alternative definitions of the smooth min-entropy, which are equal to the one we defined above up to a constant [**Ren06**, **TRSS10**, **ABJT20**]. One of these, is the $H_{\min}^{\downarrow}$ min-entropy and its smoothed variant $H_{\min}^{\downarrow,\epsilon}$, defined as:

$$H_{\min}^{\downarrow}(A|B)_\rho := \sup\left\{\lambda \in \mathbb{R} : \rho_{AB} \leq e^{-\lambda}\, \mathbb{1}_A \otimes \rho_B\right\} \tag{5.4}$$

$$H_{\min}^{\downarrow,\epsilon}(A|B)_\rho := \sup_{\tilde{\rho}} H_{\min}^{\downarrow}(A|B)_{\tilde{\rho}} \tag{5.5}$$

where the supremum is over all subnormalised states $\tilde{\rho}_{AB}$ which are ϵ-close to the state ρ in the purified distance. [**TRSS10**, Lemma 20] showed that this smooth min-entropy is equal to $H_{\min}^{\epsilon}$ up to a constant:

$$H_{\min}^{\epsilon/2}(A_1^n|B)_\rho - O\left(\log \frac{1}{\epsilon}\right) \leq H_{\min}^{\downarrow,\epsilon}(A_1^n|B)_\rho \leq H_{\min}^{\epsilon}(A_1^n|B)_\rho. \tag{5.6}$$

One can now ask whether $H_{\min}^{\downarrow,\epsilon}$ satisfies a chain rule like Eq. 5.2, that is, does

$$H_{\min}^{\downarrow,g_1(\epsilon)}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho - n g_2(\epsilon) - k(\epsilon), \tag{5.7}$$

hold true for some g_1, g_2 and k as in Eq. 5.2? Remarkably, we prove that this is indeed the case. Establishing this in Theorem 5.7 is the first main result of this chapter. We call this a *universal chain rule* for the smooth min-entropy to emphasise the fact that it is true and meaningful for a constant $\epsilon \in (0,1)$ and an arbitrary $n \in \mathbb{N}$. This universal chain rule can be viewed as a smoothed generalisation of the chain rule for $H_{\min}^{\downarrow}$:

$$H_{\min}^{\downarrow}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow}(A_k|A_1^{k-1}B)_\rho, \tag{5.8}$$

which is well known and fairly simple to prove [**Tom16**, Proposition 5.5]. The universal chain rule in Eq. 5.7 is particularly interesting because it can be used to decompose the smooth min-entropy (both $H_{\min}^{\epsilon}$ and $H_{\min}^{\downarrow,\epsilon}$) into a sum of conditional entropies, which are equally strong.

We provide two proofs for this chain rule. We briefly describe the first proof technique here, since it is simpler and we use it to prove the unstructured approximate entropy accumulation theorem as well. For this we use the simple entropic triangle inequality (Lemma 3.6)

which allows us to bound the smooth min-entropy of the state ρ in Eq. 5.7 with the min-entropy of an auxiliary state σ as

$$H_{\min}^{\delta}(A_1^n|B)_\rho \geq H_{\min}(A_1^n|B)_\sigma - D_{\max}^{\delta}(\rho\|\sigma). \tag{5.9}$$

Using the Generalised Golden-Thompson inequality [SBT17], we identify a state $\sigma_{A_1^n B}$ satisfying

$$D_m(\rho_{A_1^n B}\|\sigma_{A_1^n B}) \leq ng(\epsilon) \tag{5.10}$$

$$H_{\min}(A_1^n|B)_\sigma \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho \tag{5.11}$$

where D_m is the measured relative entropy and $g(\epsilon)$ is a small function of ϵ. Selecting an appropriate state σ is the most non-trivial part of the proof. The bound in Eq. 5.10 can further be transformed into a smooth max-relative entropy bound using the substate theorem (Theorem 2.26). Putting these bounds together into the triangle inequality yields the universal chain rule.

The second major result in this chapter is an *unstructured* approximate entropy accumulation theorem (Theorem 5.8). The entropy accumulation theorem (EAT) can only be used for states $\rho_{A_1^n B_1^n E}$ satisfying the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$ for every $k \in [n]$, which are produced by applying maps $\mathcal{M}_k : R_{k-1} \to A_k B_k R_k$ sequentially so that $\rho_{A_1^n B_1^n E} = \mathrm{tr}_{R_n} \circ \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})$. In Theorem 5.8, we significantly relax the conditions on the structure of the state required to obtain an EAT like bound. We show that for any state $\rho_{A_1^n B_1^n E}$ with an approximation chain $(\sigma_{A_1^k B_1^k E}^{(k)})_{k=1}^n$ such that for every $1 \leq k \leq n$:

$$\sigma_{A_1^k B_1^k E}^{(k)} = \mathcal{N}_k\left(\tilde{\sigma}_{A_1^{k-1} B_1^{k-1} E R_k}^{(k)}\right) \tag{5.12}$$

for some state $\tilde{\sigma}_{A_1^{k-1} B_1^{k-1} E R_k}^{(k)}$ and a channel $\mathcal{N}_k : R_k \to A_k B_k$ which samples B_k independent of the previous registers[(1)], we have the bound

$$H_{\min}^{\tilde{O}(\epsilon^{1/6})}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^{n} \inf_{\omega_{R_k \tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k(\omega)} - n\tilde{O}(\epsilon^{1/12}) - \tilde{O}\left(\frac{1}{\epsilon^{5/12}}\right) \tag{5.13}$$

where the infimum is over all states $\omega_{R_k\tilde{R}_k}$ and $\tilde{O}$ hides logarithmic factors in $1/\epsilon$. This bound holds irrespective of the process which produces the state. In fact, the central motivation behind proving this theorem was using it to prove the security of parallel DIQKD. We do this in the next chapter. The proof approach for this theorem is very similar to that of Theorem 5.7.

[(1)] Two points are worth noting. First, this condition is typically satisfied in applications of EAT. Second, if one strengthens the Markov chain condition for EAT, such that the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$ holds for all inputs to the channels $\mathcal{M}_k$, then this seems to reduce to the independence condition used here.

5.1.1. Key lemmas and theorems

As stated above, the entropic triangle inequalities proven in Lemma 3.5 and 3.6 serve as key tools for our proofs. The quantum substate theorem stated in Sec. 2.5 is one of the major results used to bound the smooth max-relative entropy in this chapter. We use the following form of the theorem here:

$$D_{\max}^{\sqrt{\epsilon}}(\rho\|\sigma) \leq \frac{D_m(\rho\|\sigma)+1}{\epsilon} + \log\frac{1}{1-\epsilon}. \tag{5.14}$$

where ρ and σ are normalised states and $\epsilon \in (0,1)$. Usually, this bound is stated with the relative entropy $D(\rho\|\sigma)$ on the right-hand side. However, as we discuss in Sec. 2.5, we can strengthen this to D_m.

In addition to the quantum substate theorem, the following generalisation of the Golden-Thompson (GT) inequality is another major result enabling the proof approach used in this chapter.

Theorem 5.1 (Generalised Golden-Thompson (GT) Inequality [SBT17])**.** *For a collection of Hermitian matrices $\{H_k\}_{k=1}^n$, we have*

$$\operatorname{tr}\exp\left(\sum_{k=1}^n H_k\right) \leq \int_{-\infty}^{\infty} dt\beta_0(t)\operatorname{tr}\left(e^{H_n}e^{\frac{1-it}{2}H_{n-1}}\cdots e^{\frac{1-it}{2}H_2}e^{H_1}e^{\frac{1+it}{2}H_2}\cdots e^{\frac{1+it}{2}H_{n-1}}\right) \tag{5.15}$$

where $\beta_0(t) := \frac{\pi}{2}(\cosh(\pi t)+1)^{-1}$ is a probability density function.

Proof. Using [SBT17, Corollary 3.3], we have

$$\log\left\|\exp\left(\sum_{k=1}^n \frac{1}{2}H_k\right)\right\|_2 \leq \int_{-\infty}^{\infty} dt\beta_0(t)\log\left\|\prod_{k=1}^n \exp\left(\frac{1+it}{2}H_k\right)\right\|_2. \tag{5.16}$$

Expanding the norm gives

$$\frac{1}{2}\log\operatorname{tr}\left(\exp\left(\sum_{k=1}^n H_k\right)\right) \leq \int_{-\infty}^{\infty} dt\beta_0(t)\frac{1}{2}\log\operatorname{tr}\left(e^{H_n}e^{\frac{1-it}{2}H_{n-1}}\cdots e^{\frac{1-it}{2}H_2}e^{H_1}e^{\frac{1+it}{2}H_2}\cdots e^{\frac{1+it}{2}H_{n-1}}\right). \tag{5.17}$$

Using the concavity and monotonicity of log, we get the statement in the Theorem. □

The GT inequality above is often used in conjunction with the following variational expressions for the relative entropies. This is also the case in this chapter.

Lemma 5.2 ([Pet88, BFT17]). *For a normalised state ρ and a positive operator Q, the following variational forms hold true*[(2)]*:*

$$D(\rho\|Q) = \sup_{\omega>0} \{\mathrm{tr}(\rho \log \omega) + 1 - \mathrm{tr}\exp(\log Q + \log \omega)\} \tag{5.18}$$

$$D_m(\rho\|Q) = \sup_{\omega>0} \{\mathrm{tr}(\rho \log \omega) + 1 - \mathrm{tr}(Q\omega)\}. \tag{5.19}$$

We will also use the following lemma, which is a quantum generalisation of the fact that if two probability distributions p_{AB} and q_{AB} are close to each other, then the probability distributions p_{AB} and $p_B q_{A|B}$ are also close to each other.

Lemma 5.3. *For a normalised state ρ_{AB} and a subnormalised state $\tilde{\rho}_{AB}$ such that $P(\rho_{AB}, \tilde{\rho}_{AB}) \leq \epsilon$, the state $\eta_{AB} := \rho_B^{1/2}\tilde{\rho}_B^{-1/2}\tilde{\rho}_{AB}\tilde{\rho}_B^{-1/2}\rho_B^{1/2}$ ($\tilde{\rho}_B^{-1/2}$ is the Moore-Penrose pseudo-inverse) satisfies $P(\rho_{AB}, \eta_{AB}) \leq (\sqrt{2}+1)\epsilon$. Note that if $\tilde{\rho}_B$ is full rank, then $\eta_B = \rho_B$.*

Proof. Note that since ρ_{AB} is normalised, we have $F(\tilde{\rho}_{AB}, \rho_{AB}) \geq 1 - \epsilon^2$. Let $|\tilde{\rho}\rangle_{ABR}$ be an arbitrary purification of $\tilde{\rho}_{AB}$. Observe that the pure state $|\eta\rangle := \rho_B^{1/2}\tilde{\rho}_B^{-1/2}|\tilde{\rho}\rangle_{ABR}$ is a purification of η_{AB}.

Let $|\tilde{\rho}\rangle_{ABR} = \sum_i \sqrt{\tilde{p}_i}|u_i\rangle_B \otimes |v_i\rangle_{AR}$ where all $p_i > 0$ be the Schmidt decomposition of $\tilde{\rho}_{ABR}$. This implies $\tilde{\rho}_B = \sum_i \tilde{p}_i |u_i\rangle\langle u_i|_B$. Then, using Uhlmann's theorem [Wat18, Theorem 3.22], we have

$$\begin{aligned}
F(\tilde{\rho}_{AB}, \eta_{AB}) &\geq |\langle\tilde{\rho}|\eta\rangle|^2 \\
&= \left|\langle\tilde{\rho}|\rho_B^{1/2}\tilde{\rho}_B^{-1/2}|\tilde{\rho}\rangle\right|^2 \\
&= \left|\mathrm{tr}\left(\rho_B^{1/2}\tilde{\rho}_B^{-1/2}\tilde{\rho}_{ABR}\right)\right|^2 \\
&= \left|\mathrm{tr}\left(\rho_B^{1/2}\tilde{\rho}_B^{1/2}\right)\right|^2 \\
&\geq F(\tilde{\rho}_B, \rho_B)^2 \\
&\geq (1-\epsilon^2)^2 \\
&\geq 1 - 2\epsilon^2
\end{aligned}$$

where we have used the relation between the *pretty good* fidelity and fidelity [IRS17, Eq. 44] for the first inequality and the fact that $F(\tilde{\rho}_B, \rho_B) \geq F(\tilde{\rho}_{AB}, \rho_{AB})$. Further, we have

$$P(\tilde{\rho}_{AB}, \eta_{AB}) = \sqrt{1 - F_*(\tilde{\rho}_{AB}, \eta_{AB})}$$

[(2)] log and exp are required to have base e for this lemma.

$$\leq \sqrt{1 - F(\tilde{\rho}_{AB}, \eta_{AB})}$$
$$\leq \sqrt{2}\epsilon.$$

Using the triangle inequality, we get

$$P(\rho_{AB}, \eta_{AB}) \leq P(\rho_{AB}, \tilde{\rho}_{AB}) + P(\tilde{\rho}_{AB}, \eta_{AB})$$
$$\leq (\sqrt{2} + 1)\epsilon.$$

□

5.2. A universal chain rule for smooth min-entropy

In this section, we will prove the universal chain rule for $H_{\min}^{\downarrow,\epsilon}$. Specifically, we will show that for a state $\rho_{A_1^n B}$,

$$H_{\min}^{\downarrow,g_1(\epsilon)}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho - ng_2(\epsilon) - k(\epsilon), \tag{5.20}$$

where g_1 and g_2 are small functions of ϵ (g_1 and g_2 are continuous and tend to 0 as $\epsilon \to 0$) and k is a general function of ϵ. It should be noted that these functions may depend on $|A|$, which is the size of the individual registers A_k. We begin by sketching the proof of such a chain rule in the classical case first. This will be beneficial for understanding the challenges that need to be solved in order to prove the statement in the quantum case. We will then generalise this proof to the quantum case. In Sec. 5.4, we provide an alternate proof for this chain rule.

5.2.1. Proof sketch for classical distributions

Consider the probability distribution $p_{A_1^n B}$. We will sketch a proof for a chain rule of the form in Eq. 5.20 for $\rho = p$. Let us first broadly describe the proof strategy. We will identify an auxiliary distribution $p^{(\text{aux})}_{A_1^n B}$, such that

$$D_{\max}^{g_1(\epsilon)}(p_{A_1^n B}\|p^{(\text{aux})}_{A_1^n B}) \leq ng_2(\epsilon) + k(\epsilon) \text{ and} \tag{5.21}$$

$$H_{\min}(A_1^n|B)_{p^{(\text{aux})}} \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_p \tag{5.22}$$

where g_1, g_2 and k are as in Eq. 5.20. Then, we can simply use the entropic triangle inequality to show that

$$H_{\min}^{g_1(\epsilon)}(A_1^n|B)_p \geq H_{\min}(A_1^n|B)_{p^{(\text{aux})}} - D_{\max}^{g_1(\epsilon)}(p_{A_1^n B}\|p^{(\text{aux})}_{A_1^n B}) \tag{5.23}$$

$$\geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_p - ng_2(\epsilon) - k(\epsilon). \tag{5.24}$$

$p^{(\text{aux})}$ will be defined using the distributions which achieve $H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_p$. For $k \in [n]$, let $q^{(k)}_{A_1^k B}$ be the distribution, such that

$$\frac{1}{2}\left\| p_{A_1^k B} - q^{(k)}_{A_1^k B} \right\|_1 \leq \epsilon \tag{5.25}$$

$$H_{\min}^{\epsilon,\downarrow}(A_k|A_1^{k-1}B)_p = H_{\min}^{\downarrow}(A_k|A_1^{k-1}B)_{q^{(k)}} \tag{5.26}$$

It is easy to show that we also have

$$\frac{1}{2}\left\| p_{A_1^k B} - p_{A_1^{k-1}B} q^{(k)}_{A_k|A_1^{k-1}B} \right\|_1 \leq 2\epsilon. \tag{5.27}$$

We can define an auxiliary distribution as $q_{A_1^n B} := p_B \prod_{k=1}^n q^{(k)}_{A_k|A_1^{k-1}B}$. For this distribution, the conditional min-entropy satisfies (Eq. 2.35),

$$H_{\min}^{\downarrow}(A_k|A_1^{k-1}B)_q = H_{\min}^{\downarrow}(A_k|A_1^{k-1}B)_{q^{(k)}} = H_{\min}^{\epsilon,\downarrow}(A_k|A_1^{k-1}B)_p \tag{5.28}$$

for every k since $q_{A_k|A_1^{k-1}B} = q^{(k)}_{A_k|A_1^{k-1}B}$. Further, we can bound the min-entropy for this distribution as desired in Eq. 5.22 as

$$H_{\min}(A_1^n|B)_q \geq H_{\min}^{\downarrow}(A_1^n|B)_q \geq \sum_{k=1}^n H_{\min}^{\downarrow}(A_k|A_1^{k-1}B)_q \tag{5.29}$$

Next, we need to ensure that the smooth max-divergence between p and q is relatively small. Our strategy will be to bound the relative entropy between these two distributions and use the substate theorem to convert that to a bound for the smooth max-divergence. We expect to be able to use the following chain rule for the relative entropy

$$D(p_{A_1^k B}\|q_{A_1^k B}) = D(p_{A_1^{k-1}B}\|q_{A_1^{k-1}B}) + D(p_{A_1^k B}\|p_{A_1^{k-1}B} q_{A_k|A_1^{k-1}B}) \tag{5.30}$$

to prove that the relative entropy distance between $p_{A_1^n B}$ and $q_{A_1^n B}$ is small. Using this repeatedly along with the fact that $q_{A_k|A_1^{k-1}B} = q^{(k)}_{A_k|A_1^{k-1}B}$, gives us

$$D(p_{A_1^n B}\|q_{A_1^n B}) = \sum_{k=1}^n D(p_{A_1^k B}\|p_{A_1^{k-1}B} q^{(k)}_{A_k|A_1^{k-1}B}). \tag{5.31}$$

If we could now show that each term inside the summation is some small function in ϵ, $g(\epsilon)$, we could show that the $D(p_{A_1^n B}\|q_{A_1^n B})$ is bounded by $ng(\epsilon)$. Eq. 5.27 shows that the two distributions in each of the terms are close to each other. However, the relative entropy between them need not be small or even bounded, since it could be that $p_{A_1^k B} \not\ll p_{A_1^{k-1}B} q^{(k)}_{A_k|A_1^{k-1}B}$. To circumvent this technicality, we need to *massage* the distributions $q^{(k)}_{A_k|A_1^{k-1}B}$ by a small amount. This is done by mixing these distributions with the uniform distribution u_{A_k} to produce the distributions

$$r^{(k)}_{A_k|A_1^{k-1}B} = (1-\delta) q^{(k)}_{A_k|A_1^{k-1}B} + \delta u_{A_k}. \tag{5.32}$$

For these distributions, we can show that (Lemma C.3)

$$\frac{1}{2}\left\|p_{A_1^k B} - p_{A_1^{k-1}B} r^{(k)}_{A_k|A_1^{k-1}B}\right\|_1 \leq 2\epsilon + \delta \tag{5.33}$$

and

$$D(p_{A_1^k B} \| p_{A_1^{k-1}B} r^{(k)}_{A_k|A_1^{k-1}B}) \leq z(\epsilon, \delta) \tag{5.34}$$

where $z(\epsilon,\delta) = O\left((\epsilon+\delta)\log\frac{|A|}{\epsilon\delta}\right)$. Note that $z(\epsilon,\delta)$ can be made $O\left(\epsilon \log \frac{|A|}{\epsilon}\right)$ for $\delta = \epsilon$. Instead of using $q_{A_1^n B}$ as the auxiliary distribution, we can use $r_{A_1^n B} \coloneqq p_B \prod_{k=1}^n r^{(k)}_{A_k|A_1^{k-1}B}$. We can use the same argument as above to bound the relative entropy between $p_{A_1^n B}$ and $r_{A_1^n B}$. This gives us

$$\begin{aligned} D(p_{A_1^n B} \| r_{A_1^n B}) &= \sum_{k=1}^n D(p_{A_1^k B} \| p_{A_1^{k-1}B} r^{(k)}_{A_k|A_1^{k-1}B}) \\ &\leq n z(\epsilon,\delta). \end{aligned} \tag{5.35}$$

Let $\mu \coloneqq z(\epsilon,\delta)^{1/3}$, then using the substate theorem, we have

$$D^\mu_{\max}(p_{A_1^n B} \| r_{A_1^n B}) \leq n\mu + \frac{1}{\mu^2} + \log\frac{1}{1-\mu^2}.$$

Using the entropic triangle inequality, we get

$$\begin{aligned} H^\mu_{\min}(A_1^n|B)_p &\geq H_{\min}(A_1^n|B)_r - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \\ &\geq \sum_{k=1}^n H^\downarrow_{\min}(A_k|A_1^{k-1}B)_r - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \\ &\geq \sum_{k=1}^n H^\downarrow_{\min}(A_k|A_1^{k-1}B)_{q^{(k)}} - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \\ &= \sum_{k=1}^n H^{\downarrow,\epsilon}_{\min}(A_k|A_1^{k-1}B)_p - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \end{aligned}$$

where in the third line, we use the quasi-concavity of $H^\downarrow_{\min}$ [Tom16, Pg 73].

Thus, we have proven a chain rule for $H^{\downarrow,\epsilon}_{\min}$ of the desired form in the classical case. The primary challenge in generalizing this approach to the quantum setting lies in defining the auxiliary state. In the classical case, we could use the product of (suitably massaged) conditional distributions $q^{(k)}_{A_k|A_1^{k-1}B}$ to define the auxiliary distribution $r_{A_1^n B}$. However, the quantum generalisation of such a distribution is not unique. Moreover, such generalisations are all quite difficult to manipulate. A closely related problem is that we need to be able to prove a relative entropy bound similar to that in Eq. 5.35 for the auxiliary state. This also turns out to be quite challenging, especially because the quantum chain rules [SBT17, Theorem 4.1 and Proposition F.1] for the relative entropy do not yield something as simple and convenient

as Eq. 5.30. We solve both of these problems indirectly. We will use a simple relative entropy bound between the state $\rho_{A_1^n B}$ and an unwieldy, yet simple exponential generalisation of the auxiliary probability distribution used above. Then, in order to convert this bound to a (measured) relative entropy bound between the state $\rho_{A_1^n B}$ and a simpler and more convenient auxiliary state, we employ the Generalised GT inequality (Lemma 5.4). Through this approach, we are able to delegate the pesky question of the correct generalisation of the auxiliary state to the GT inequality.

5.2.2. Proof for the quantum case

The following lemma uses the GT inequality as described above to create a generalisation of the auxiliary state used in the classical proof and also prove a measured relative entropy bound between the original state and this auxiliary state.

Lemma 5.4. *Suppose $\rho_{A_1^n B}$ is a normalised state and for $1 \leq k \leq n$ the normalised states $\bar{\rho}^{(k)}_{A_1^k B}$ are full rank and satisfy*

$$D\left(\rho_{A_1^k B} \| \bar{\rho}^{(k)}_{A_1^k B}\right) \leq \epsilon \tag{5.36}$$

for some $\epsilon > 0$. Let $\bar{\rho}^{(0)}_B := \rho_B$ for notational simplicity. Then, the subnormalised state[3]

$$\sigma_{A_1^n B} := \int_{-\infty}^{\infty} dt \beta_0(t) \prod_{k=0}^{n-1} \left[\left(\bar{\rho}^{(k)}_{A_1^k B}\right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(k+1)}_{A_1^k B}\right)^{-\frac{1-it}{2}} \right] \cdot \bar{\rho}^{(n)}_{A_1^n B} \cdot \prod_{k=n-1}^{0} \left[\left(\bar{\rho}^{(k+1)}_{A_1^k B}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(k)}_{A_1^k B}\right)^{\frac{1+it}{2}} \right] \tag{5.37}$$

$$= \int_{-\infty}^{\infty} dt \beta_0(t) \rho_B^{\frac{1-it}{2}} \left(\bar{\rho}^{(1)}_B\right)^{-\frac{1-it}{2}} \left(\bar{\rho}^{(1)}_{A_1 B}\right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(2)}_{A_1 B}\right)^{-\frac{1-it}{2}} \left(\bar{\rho}^{(2)}_{A_1^2 B}\right)^{\frac{1-it}{2}} \cdots \left(\bar{\rho}^{(n)}_{A_1^{n-1} B}\right)^{-\frac{1-it}{2}}$$
$$\bar{\rho}^{(n)}_{A_1^n B} \cdot \left(\bar{\rho}^{(n)}_{A_1^{n-1} B}\right)^{-\frac{1+it}{2}} \cdots \left(\bar{\rho}^{(2)}_{A_1^2 B}\right)^{\frac{1+it}{2}} \left(\bar{\rho}^{(2)}_{A_1 B}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(1)}_{A_1 B}\right)^{\frac{1+it}{2}} \left(\bar{\rho}^{(1)}_B\right)^{-\frac{1+it}{2}} \rho_B^{\frac{1+it}{2}}. \tag{5.38}$$

is such that

$$D_m(\rho_{A_1^n B} \| \sigma_{A_1^n B}) \leq n\epsilon. \tag{5.39}$$

Further, the partial states of σ are

$$\sigma_{A_1^k B} = \int_{-\infty}^{\infty} dt \beta_0(t) \prod_{j=0}^{k-1} \left[\left(\bar{\rho}^{(j)}_{A_1^j B}\right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(j+1)}_{A_1^j B}\right)^{-\frac{1-it}{2}} \right] \cdot \bar{\rho}^{(k)}_{A_1^k B} \cdot \prod_{j=k-1}^{0} \left[\left(\bar{\rho}^{(j+1)}_{A_1^j B}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(j)}_{A_1^j B}\right)^{\frac{1+it}{2}} \right] \tag{5.40}$$

$$= \int_{-\infty}^{\infty} dt \beta_0(t) \rho_B^{\frac{1-it}{2}} \left(\bar{\rho}^{(1)}_B\right)^{-\frac{1-it}{2}} \left(\bar{\rho}^{(1)}_{A_1 B}\right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(2)}_{A_1 B}\right)^{-\frac{1-it}{2}} \left(\bar{\rho}^{(2)}_{A_1^2 B}\right)^{\frac{1-it}{2}} \cdots \left(\bar{\rho}^{(k)}_{A_1^{k-1} B}\right)^{-\frac{1-it}{2}}$$

[3] In the following and throughout this chapter, the product notation $\Pi_{j=0}^{k} M_j$ represents the operator product $M_0 M_1 \cdots M_k$ and the notation $\Pi_{j=k}^{0} M_j$ represents the operator product $M_k M_{k-1} \cdots M_0$.

$$\bar{\rho}^{(k)}_{A_1^k B} \cdot \left(\bar{\rho}^{(k)}_{A_1^{k-1}B}\right)^{-\frac{1+it}{2}} \cdots \left(\bar{\rho}^{(2)}_{A_1^2 B}\right)^{\frac{1+it}{2}} \left(\bar{\rho}^{(2)}_{A_1 B}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(1)}_{A_1 B}\right)^{\frac{1+it}{2}} \left(\bar{\rho}^{(1)}_{B}\right)^{-\frac{1+it}{2}} \rho_B^{\frac{1+it}{2}}. \tag{5.41}$$

Hence, $\sigma_B = \rho_B$, and σ is normalised.

Proof. Our approach here is broadly based on the proof of [SBT17, Theorem 4.1]. Define the positive operator[4]

$$Q_{A_1^n B} := \exp\left(\sum_{k=1}^{n} \left(\log \bar{\rho}^{(k)}_{A_1^k B} - \log \bar{\rho}^{(k)}_{A_1^{k-1}B}\right) + \log \rho_B\right). \tag{5.42}$$

Note that if all the operators above commuted, $Q_{A_1^n B}$ (and also $\sigma_{A_1^n B}$) would be the probability distribution $\bar{\rho}^{(n)}_{A_n|A_1^{n-1}B}\bar{\rho}^{(n-1)}_{A_{n-1}|A_1^{n-2}B} \cdots \bar{\rho}^{(1)}_{A_1|B}\rho_B$. As is often times the case with the quantum relative entropy, the best quantum generalisation of the above conditional distributions turns out to be an exponential[5]. For this operator, we have

$$\begin{aligned}
D(\rho_{A_1^n B}\|Q_{A_1^n B}) &= \operatorname{tr}\left(\rho_{A_1^n B}\left(\log \rho_{A_1^n B} - \log Q_{A_1^n B}\right)\right) \\
&= \operatorname{tr} \rho_{A_1^n B}\left(\log \rho_{A_1^n B} - \sum_{k=1}^{n}\left(\log \bar{\rho}^{(k)}_{A_1^k B} - \log \bar{\rho}^{(k)}_{A_1^{k-1}B}\right) - \log \rho_B\right) \\
&= \sum_{k=1}^{n} \operatorname{tr}\left(\rho_{A_1^n B}\left(\log \rho_{A_1^k B} - \log \bar{\rho}^{(k)}_{A_1^k B}\right)\right) - \sum_{k=1}^{n} \operatorname{tr}\left(\rho_{A_1^n B}\left(\log \rho_{A_1^{k-1} B} - \log \bar{\rho}^{(k)}_{A_1^{k-1} B}\right)\right) \\
&= \sum_{k=1}^{n}\left(D(\rho_{A_1^k B}\|\bar{\rho}^{(k)}_{A_1^k B}) - D(\rho_{A_1^{k-1} B}\|\bar{\rho}^{(k)}_{A_1^{k-1} B})\right) \\
&\leq \sum_{k=1}^{n} D(\rho_{A_1^k B}\|\bar{\rho}^{(k)}_{A_1^k B}) \\
&\leq n\epsilon
\end{aligned} \tag{5.43}$$

where in the second last line, we have used the fact that the relative entropy of two normalised states is positive and in the last line, we have used the bound in Eq. 5.36. Next, by using the variational expression for the relative entropy (Eq. 5.18), we see that

$$\begin{aligned}
n\epsilon &\geq D(\rho_{A_1^n B}\|Q_{A_1^n B}) \\
&= \sup_{\omega_{A_1^n B} > 0} \left\{\operatorname{tr}(\rho_{A_1^n B} \log \omega_{A_1^n B}) + 1 - \operatorname{tr} \exp\left(\sum_{k=1}^{n}\left(\log \bar{\rho}^{(k)}_{A_1^k B} - \log \bar{\rho}^{(k)}_{A_1^{k-1}B}\right) + \log \rho_B + \log \omega_{A_1^n B}\right)\right\}.
\end{aligned} \tag{5.44}$$

[4] By restricting ourselves to strictly positive operators $\bar{\rho}^{(k)}_{A_1^k B}$, we ensure that this expression is always well-defined.

[5] Classically, we have that $D(P_{AB}\|Q_{AB}) - D(P_A\|Q_A) = D(P_{AB}\|P_A Q_{B|A})$. Quantumly, this equation is directly best generalised as $D(\rho_{AB}\|\sigma_{AB}) - D(\rho_A\|\sigma_A) = D(\rho_{AB}\|\exp(\log\sigma_{AB} - \log\sigma_A + \log\rho_A))$. Though, this is not very useful on its own.

The trace exponential above can be bounded using the Generalised GT inequality (Theorem 5.1) as

$$\begin{aligned}
&\operatorname{tr}\exp\left(\sum_{k=1}^{n}\left(\log\bar{\rho}^{(k)}_{A_1^kB}-\log\bar{\rho}^{(k)}_{A_1^{k-1}B}\right)+\log\rho_B+\log\omega_{A_1^nB}\right)\\
&\leq\int_{-\infty}^{\infty}dt\beta_0(t)\operatorname{tr}\left(\omega_{A_1^nB}\,\rho_B^{\frac{1-it}{2}}\left(\bar{\rho}^{(1)}_B\right)^{-\frac{1-it}{2}}\left(\bar{\rho}^{(1)}_{A_1B}\right)^{\frac{1-it}{2}}\left(\bar{\rho}^{(2)}_{A_1B}\right)^{-\frac{1-it}{2}}\left(\bar{\rho}^{(2)}_{A_1^2B}\right)^{\frac{1-it}{2}}\cdots\left(\bar{\rho}^{(n)}_{A_1^{n-1}B}\right)^{-\frac{1-it}{2}}\cdot\right.\\
&\qquad\left.\bar{\rho}^{(n)}_{A_1^nB}\cdot\left(\bar{\rho}^{(n)}_{A_1^{n-1}B}\right)^{-\frac{1+it}{2}}\cdots\left(\bar{\rho}^{(2)}_{A_1^2B}\right)^{\frac{1+it}{2}}\left(\bar{\rho}^{(2)}_{A_1B}\right)^{-\frac{1+it}{2}}\left(\bar{\rho}^{(1)}_{A_1B}\right)^{\frac{1+it}{2}}\left(\bar{\rho}^{(1)}_{B}\right)^{-\frac{1+it}{2}}\rho_B^{\frac{1+it}{2}}\right)\\
&=\int_{-\infty}^{\infty}dt\beta_0(t)\operatorname{tr}\left(\omega_{A_1^nB}\prod_{k=0}^{n-1}\left[\left(\bar{\rho}^{(k)}_{A_1^kB}\right)^{\frac{1-it}{2}}\left(\bar{\rho}^{(k+1)}_{A_1^kB}\right)^{-\frac{1-it}{2}}\right]\cdot\bar{\rho}^{(n)}_{A_1^nB}\cdot\prod_{k=n-1}^{0}\left[\left(\bar{\rho}^{(k+1)}_{A_1^kB}\right)^{-\frac{1+it}{2}}\left(\bar{\rho}^{(k)}_{A_1^kB}\right)^{\frac{1+it}{2}}\right]\right)\\
&=\operatorname{tr}\left(\omega_{A_1^nB}\sigma_{A_1^nB}\right)
\end{aligned}\tag{5.45}$$

for the state σ defined in the statement of the lemma. Plugging the bound above into Eq. 5.44, we get

$$\begin{aligned}
n\epsilon&\geq D(\rho_{A_1^nB}\|Q_{A_1^nB})\\
&\geq\sup_{\omega_{A_1^nB}>0}\left\{\operatorname{tr}(\rho_{A_1^nB}\log\omega_{A_1^nB})+1-\operatorname{tr}\left(\omega_{A_1^nB}\sigma_{A_1^nB}\right)\right\}\\
&=D_m(\rho_{A_1^nB}\|\sigma_{A_1^nB}).
\end{aligned}\tag{5.46}$$

□

We use the following theorem and its corollary to bound the relative entropy between two states which are close to each other and have a bounded max-relative entropy.

Theorem 5.5 ([Aud14, Theorem 4]). *For two normalised quantum states ρ and σ, and $\delta\in(0,1]$, we have*

$$D(\rho\|(1-\delta)\sigma+\delta\rho)\leq\frac{1}{2}\|\rho-\sigma\|_1\log\frac{1}{\delta}.\tag{5.47}$$

Corollary 5.6. *Suppose two normalised quantum states ρ_{AB} and σ_{AB} are such that*

$$\frac{1}{2}\|\rho_{AB}-\sigma_{AB}\|_1\leq\epsilon\tag{5.48}$$

for $\epsilon\in[0,1]$ and

$$\delta\tau_A\otimes\rho_B\leq\sigma_{AB}\tag{5.49}$$

where $\delta\in(0,1)$ and τ_A is the completely mixed state on register A. Then,

$$D(\rho_{AB}\|\sigma_{AB})\leq\frac{\epsilon}{1-\delta/|A|^2}\log\frac{|A|^2}{\delta}.\tag{5.50}$$

Proof. Note that

$$\frac{\delta}{|A|^2}\rho_{AB} \leq \delta\tau_A \otimes \rho_B \leq \sigma_{AB}.$$

So, we can write σ as

$$\sigma_{AB} = \left(1 - \frac{\delta}{|A|^2}\right)\sigma'_{AB} + \frac{\delta}{|A|^2}\rho_{AB} \tag{5.51}$$

for some normalised quantum state σ'_{AB}. For this state, we have

$$\begin{aligned}
\frac{1}{2}\left\|\rho_{AB} - \sigma'_{AB}\right\|_1 &= \frac{1}{2}\left\|\rho_{AB} - \frac{1}{1-\delta/|A|^2}(\sigma_{AB} - \frac{\delta}{|A|^2}\rho_{AB})\right\|_1 \\
&= \frac{1}{2}\frac{1}{1-\delta/|A|^2}\left\|\rho_{AB} - \frac{\delta}{|A|^2}\rho_{AB} - (\sigma_{AB} - \frac{\delta}{|A|^2}\rho_{AB})\right\|_1 \\
&\leq \frac{\epsilon}{1-\delta/|A|^2}.
\end{aligned}$$

Using Theorem 5.5, we have that

$$\begin{aligned}
D(\rho_{AB}\|\sigma_{AB}) &\leq \frac{1}{2}\left\|\rho_{AB} - \sigma'_{AB}\right\|_1 \log\frac{|A|^2}{\delta} \\
&\leq \frac{\epsilon}{1-\delta/|A|^2}\log\frac{|A|^2}{\delta}.
\end{aligned}$$

□

Now we are ready to state and prove the universal chain rule for the smooth min-entropy for quantum states.

Theorem 5.7. *For a normalised quantum state $\rho_{A_1^n B}$ such that for all $k \in [n]$ the dimension $|A_k| = |A|$, and $\epsilon \in (0,1)$ such that $\mu = \left(\frac{2\epsilon}{1-\epsilon/|A|^2}\log\frac{|A|^2}{\epsilon}\right)^{1/3} = O\left(\left(\epsilon\log\frac{|A|}{\epsilon}\right)^{1/3}\right)$ lies in $(0,1)$, we have the chain rule*

$$H_{\min}^{\downarrow,2\mu+\epsilon/2}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon/2}(A_k|A_1^{k-1}B)_\rho - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} - \log\left(\frac{2}{\mu^2} + \frac{1}{1-\mu}\right) \tag{5.52}$$

Proof. **Case 1:** Let's first consider states $\rho_{A_1^n B}$, which are full rank.

For every $k \in [n]$, define $\lambda_k := H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho$ and let the state $\tilde{\rho}^{(k)}_{A_1^k B}$ be such that[6]

$$P(\rho_{A_1^k B}, \tilde{\rho}^{(k)}_{A_1^k B}) \leq \epsilon \tag{5.53}$$

$$\tilde{\rho}^{(k)}_{A_1^k B} \leq e^{-\lambda_k}\,\mathbb{1}_{A_k} \otimes \tilde{\rho}^{(k)}_{A_1^{k-1}B}. \tag{5.54}$$

[6]Since $H_{\min}^{\downarrow}$ is discontinuous, strictly speaking, states achieving $H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho$ may not exist. To take this into account, we can consider λ_k to be arbitrarily close to $H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho$. However, for simplicity we assume that such states exist throughout this paper.

Without loss of generality, we will choose the states $\tilde{\rho}^{(k)}_{A_1^k B}$ to be normalised states, since dividing $\tilde{\rho}^{(k)}_{A_1^k B}$ by a normalising factor $N \leq 1$ only decreases its purified distance from ρ and leaves the operator inequality above invariant. Further, let $\delta \in (0,1)$ be a small parameter. For every $k \in [n]$, we define the normalised states,

$$\bar{\rho}^{(k)}_{A_1^k B} := (1-\delta)\tilde{\rho}^{(k)}_{A_1^k B} + \delta \tau_{A_k} \otimes \rho_{A_1^{k-1} B} \tag{5.55}$$

where τ_{A_k} is the completely mixed state on register A_k. Also, define $\bar{\rho}^{(0)}_B := \rho_B$ for notational convenience. Since, $\rho_{A_1^{k-1}B}$ is full rank, $\bar{\rho}^{(k)}_{A_1^k B}$ are also full rank for all k.

For each $k \in [n]$, these states satisfy

$$\begin{aligned} \frac{1}{2}\left\|\rho_{A_1^k B} - \bar{\rho}^{(k)}_{A_1^k B}\right\|_1 &\leq \frac{1-\delta}{2}\left\|\rho_{A_1^k B} - \tilde{\rho}^{(k)}_{A_1^k B}\right\|_1 + \frac{\delta}{2}\left\|\rho_{A_1^k B} - \tau_{A_k} \otimes \rho_{A_1^{k-1} B}\right\|_1 \\ &\leq \epsilon + \delta \end{aligned} \tag{5.56}$$

and

$$D(\rho_{A_1^k B} \| \bar{\rho}^{(k)}_{A_1^k B}) \leq \frac{\epsilon + \delta}{1 - \delta/|A|^2} \log \frac{|A|^2}{\delta}. \tag{5.57}$$

using Corollary 5.6. Let's define the right-hand side above as

$$z(\epsilon, \delta) := \frac{\epsilon + \delta}{1 - \delta/|A|^2} \log \frac{|A|^2}{\delta}. \tag{5.58}$$

Note that this can be made small, say $O(\epsilon \log 1/\epsilon)$, by choosing $\delta = \epsilon$ for example.

Using Lemma 5.4, we have that the state σ defined as

$$\sigma_{A_1^n B} := \int_{-\infty}^{\infty} dt \beta_0(t) \prod_{k=0}^{n-1} \left[\left(\bar{\rho}^{(k)}_{A_1^k B}\right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(k+1)}_{A_1^k B}\right)^{-\frac{1-it}{2}} \right] \cdot \bar{\rho}^{(n)}_{A_1^n B} \cdot \prod_{k=n-1}^{0} \left[\left(\bar{\rho}^{(k+1)}_{A_1^k B}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(k)}_{A_1^k B}\right)^{\frac{1+it}{2}} \right] \tag{5.59}$$

$$\begin{aligned} = \int_{-\infty}^{\infty} dt \beta_0(t) \rho_B^{\frac{1-it}{2}} \left(\bar{\rho}^{(1)}_B\right)^{-\frac{1-it}{2}} \left(\bar{\rho}^{(1)}_{A_1 B}\right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(2)}_{A_1 B}\right)^{-\frac{1-it}{2}} \left(\bar{\rho}^{(2)}_{A_1^2 B}\right)^{\frac{1-it}{2}} \cdots \left(\bar{\rho}^{(n)}_{A_1^{n-1} B}\right)^{-\frac{1-it}{2}} \\ \bar{\rho}^{(n)}_{A_1^n B} \cdot \left(\bar{\rho}^{(n)}_{A_1^{n-1} B}\right)^{-\frac{1+it}{2}} \cdots \left(\bar{\rho}^{(2)}_{A_1^2 B}\right)^{\frac{1+it}{2}} \left(\bar{\rho}^{(2)}_{A_1 B}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(1)}_{A_1 B}\right)^{\frac{1+it}{2}} \left(\bar{\rho}^{(1)}_B\right)^{-\frac{1+it}{2}} \rho_B^{\frac{1+it}{2}} \end{aligned} \tag{5.60}$$

is normalised and satisfies

$$D_m(\rho_{A_1^n B} \| \sigma_{A_1^n B}) \leq n z(\epsilon, \delta). \tag{5.61}$$

Further,

$$\sigma_{A_1^k B} = \int_{-\infty}^{\infty} dt \beta_0(t) \prod_{j=0}^{k-1} \left[\left(\bar{\rho}^{(j)}_{A_1^j B}\right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(j+1)}_{A_1^j B}\right)^{-\frac{1-it}{2}} \right] \cdot \bar{\rho}^{(k)}_{A_1^k B} \cdot \prod_{j=k-1}^{0} \left[\left(\bar{\rho}^{(j+1)}_{A_1^j B}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(j)}_{A_1^j B}\right)^{\frac{1+it}{2}} \right] \tag{5.62}$$

for $k \in [n]$. Using the substate theorem[7], we have that for $\mu := z(\epsilon,\delta)^{1/3}$ (we omit the dependence of μ on ϵ and δ for clarity)

$$D^{\mu}_{\max}(\rho_{A_1^nB}\|\sigma_{A_1^nB}) \leq \frac{D_m(\rho_{A_1^nB}\|\sigma_{A_1^nB})+1}{\mu^2} + \log\frac{1}{1-\mu^2} \tag{5.63}$$

$$\leq n\mu + \frac{1}{\mu^2} + \log\frac{1}{1-\mu^2}. \tag{5.64}$$

A simple application of the entropic triangle inequality in Lemma 3.6 gives us that

$$\begin{aligned} H^{\mu}_{\min}(A_1^n|B)_\rho &\geq H_{\min}(A_1^n|B)_\sigma - D^{\mu}_{\max}(\rho_{A_1^nB}\|\sigma_{A_1^nB}) \\ &\geq H_{\min}(A_1^n|B)_\sigma - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \\ &\geq H^{\downarrow}_{\min}(A_1^n|B)_\sigma - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \\ &\geq \sum_{k=1}^{n} H^{\downarrow}_{\min}(A_k|A_1^{k-1}B)_\sigma - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \end{aligned} \tag{5.65}$$

where we have used the chain rule for $H^{\downarrow}_{\min}$ (Eq. 5.8). We will now show that σ is such that $H^{\downarrow}_{\min}(A_k|A_1^{k-1}B)_\sigma \geq H^{\downarrow,\epsilon}_{\min}(A_k|A_1^{k-1}B)_\rho$.

Using the quasi-concavity of $H^{\downarrow}_{\min}$ for $\bar{\rho}^{(k)}$ [**Tom16**, Pg 73], we have

$$\begin{aligned} H^{\downarrow}_{\min}(A_k|A_1^{k-1}B)_{\bar{\rho}^{(k)}} &\geq \min\left\{H^{\downarrow}_{\min}(A_k|A_1^{k-1}B)_{\tilde{\rho}^{(k)}}, H^{\downarrow}_{\min}(A_k|A_1^{k-1}B)_{\tau_{A_k}\otimes\rho_{A_1^{k-1}B}}\right\} \\ &\geq \min\left\{H^{\downarrow}_{\min}(A_k|A_1^{k-1}B)_{\tilde{\rho}^{(k)}}, \log|A|\right\} \\ &\geq H^{\downarrow}_{\min}(A_k|A_1^{k-1}B)_{\tilde{\rho}^{(k)}}. \end{aligned}$$

Therefore, we have that

$$\bar{\rho}^{(k)}_{A_1^kB} \leq e^{-\lambda_k}\,\mathbb{1}_{A_k}\otimes\bar{\rho}^{(k)}_{A_1^{k-1}B}. \tag{5.66}$$

This implies that

$$\begin{aligned} \sigma_{A_1^kB} &= \int_{-\infty}^{\infty} dt\beta_0(t)\prod_{j=0}^{k-1}\left[\left(\bar{\rho}^{(j)}_{A_1^jB}\right)^{\frac{1-it}{2}}\left(\bar{\rho}^{(j+1)}_{A_1^jB}\right)^{-\frac{1-it}{2}}\right]\cdot\bar{\rho}^{(k)}_{A_1^kB}\cdot\prod_{j=k-1}^{0}\left[\left(\bar{\rho}^{(j+1)}_{A_1^jB}\right)^{-\frac{1+it}{2}}\left(\bar{\rho}^{(j)}_{A_1^jB}\right)^{\frac{1+it}{2}}\right] \\ &\leq e^{-\lambda_k}\int_{-\infty}^{\infty} dt\beta_0(t)\prod_{j=0}^{k-1}\left[\left(\bar{\rho}^{(j)}_{A_1^jB}\right)^{\frac{1-it}{2}}\left(\bar{\rho}^{(j+1)}_{A_1^jB}\right)^{-\frac{1-it}{2}}\right]\cdot\mathbb{1}_{A_k}\otimes\bar{\rho}^{(k)}_{A_1^{k-1}B}\cdot\prod_{j=k-1}^{0}\left[\left(\bar{\rho}^{(j+1)}_{A_1^jB}\right)^{-\frac{1+it}{2}}\left(\bar{\rho}^{(j)}_{A_1^jB}\right)^{\frac{1+it}{2}}\right] \\ &\leq e^{-\lambda_k}\,\mathbb{1}_{A_k}\otimes\int_{-\infty}^{\infty} dt\beta_0(t)\prod_{j=0}^{k-2}\left[\left(\bar{\rho}^{(j)}_{A_1^jB}\right)^{\frac{1-it}{2}}\left(\bar{\rho}^{(j+1)}_{A_1^jB}\right)^{-\frac{1-it}{2}}\right]\cdot\bar{\rho}^{(k-1)}_{A_1^{k-1}B}\cdot\prod_{j=k-2}^{0}\left[\left(\bar{\rho}^{(j+1)}_{A_1^jB}\right)^{-\frac{1+it}{2}}\left(\bar{\rho}^{(j)}_{A_1^jB}\right)^{\frac{1+it}{2}}\right] \end{aligned}$$

[7]It is quite remarkable that the substate theorem works with a D_m bound and the generalised GT inequality only yields a D_m bound. All our proofs exploit this fact. Whether it is an incredible coincidence or an indication of the tightness of these bounds: you decide.

$= e^{-\lambda_k} \mathbb{1}_{A_k} \otimes \sigma_{A_1^{k-1}B}$.

Or, equivalently that $H_{\min}^{\downarrow}(A_k|A_1^{k-1}B)_\sigma \geq \lambda_k = H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho$. Plugging this bound in Eq. 5.65, we get

$$H_{\min}^{\mu}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2}. \tag{5.67}$$

Using Eq. 2.42, we can upper bound the left-hand side with $H_{\min}^{\downarrow,2\mu}(A_1^n|B)$ to derive

$$H_{\min}^{\downarrow,2\mu}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_\rho - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} - \log\left(\frac{2}{\mu^2} + \frac{1}{1-\mu}\right). \tag{5.68}$$

Case 2: The above proves the Theorem for the case when ρ was full rank. If ρ was not full rank, then for an arbitrary $\nu \in (0,1)$, the state $\rho'_{A_1^nB} := (1-\nu)\rho_{A_1^nB} + \nu\tau_{A_1^nB}$, which has full support, satisfies

$$H_{\min}^{\downarrow,2\mu}(A_1^n|B)_{\rho'} \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_{\rho'} - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} - \log\left(\frac{2}{\mu^2} + \frac{1}{1-\mu}\right) \tag{5.69}$$

which implies that

$$H_{\min}^{\downarrow,2\mu+\sqrt{2\nu}}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon-\sqrt{2\nu}}(A_k|A_1^{k-1}B)_\rho - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} - \log\left(\frac{2}{\mu^2} + \frac{1}{1-\mu}\right) \tag{5.70}$$

for every $\nu \in (0,1)$. Unfortunately, since $H_{\min}^{\downarrow}$ is not continuous, we cannot use continuity to claim the above for ρ itself. We can, however, simply choose $\nu = \frac{\epsilon^2}{8}$, which gives us

$$H_{\min}^{\downarrow,2\mu+\epsilon/2}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon/2}(A_k|A_1^{k-1}B)_{\rho'} - n\mu - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} - \log\left(\frac{2}{\mu^2} + \frac{1}{1-\mu}\right) \tag{5.71}$$

where $\mu = z(\epsilon,\delta)^{1/3} = \left(\frac{\epsilon+\delta}{1-\delta/|A|^2}\log\frac{|A|^2}{\delta}\right)^{1/3}$. To derive the bound in the Theorem, we make the concrete choice $\delta = \epsilon$. □

5.3. Unstructured approximate entropy accumulation

In this section, we develop another approximate version of the entropy accumulation theorem (EAT) in Theorem 5.8. We show that for any state $\rho_{A_1^nB_1^nE}$ whose partial states $\rho_{A_1^kB_1^kE}$ can be ϵ-approximated as the output of channels $\mathcal{M}_k$, which sample the side information B_k independent of the previous registers $A_1^{k-1}B_1^{k-1}E$ we can recover a statement similar to EAT[(8)]. Crucially, Theorem 5.8 does not require $\rho_{A_1^nB_1^nE}$ be produced by a structured process like EAT (Fig. 2.1); the state may even be produced by a completely parallel process. In fact, our central motivation to develop this theorem was to prove the security of parallel device-independent QKD. We do this in the next chapter. For this

[(8)]This seems to be equivalent to requiring that all outputs of the channel $\mathcal{M}_k$ satisfy $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$ (see Appendix C.3).

reason, we call this theorem the *unstructured* approximate EAT.

We also proved an approximate entropy accumulation theorem in Theorem 3.12. The key difference between these two theorems lies in their applicability: Theorem 3.12 applies only to states ρ produced by a sequential process of the same structure as the original EAT (Fig. 3.2), while Theorem 5.8 can be applied to states generated by a completely unstructured or parallel process. Furthermore, Theorem 3.12 considers approximation at the level of channels. Specifically, it considers the diamond norm approximation of the channels $\mathcal{M}_k$ producing the output state ρ in Fig. 3.2 by *nicer* channels $\mathcal{M}'_k$, which satisfy certain Markov chain conditions. This is a strong approximation condition, since it requires the outputs of $\mathcal{M}_k$ and $\mathcal{M}'_k$ to be approximately equal for all input states. In contrast, in Theorem 5.8, we only need the trace distance between $\rho_{A_1^k B_1^k E}$ and the output of $\mathcal{M}_k$ to be small for a single input state. We pay the price for these weaker conditions in terms of a much stronger condition on the side information produced by the approximation channels. Additionally, the smoothing parameter in Theorem 5.8 depends on the approximation parameter ϵ and cannot be made arbitrarily small. In comparison, the smoothing parameter for the smooth min-entropy in Theorem 3.12 is independent of the approximation parameter and can be chosen arbitrarily small. Consequently, in its current form, Theorem 5.8 may have limited utility in cryptographic scenarios where experimental noise or imperfections need to be accounted for, as cryptographic protocols typically require the smoothing parameter to be independent of the noise parameters. Under an analysis using Theorem 5.8, however, the noise parameters would determine the approximation parameter and hence the smoothing parameter. We discuss the possibility of improving this theorem to decouple these parameters in Sec. 5.3.2. Nonetheless, Theorem 5.8 is a valuable theoretical tool in scenarios where the approximation parameter can be made arbitrarily small, as we will demonstrate in the security analysis of parallel DIQKD in the next chapter. We state the unstructured approximate EAT as the following theorem.

Theorem 5.8. *Let $\epsilon \in (0,1)$ and for every $k \in [n]$ the registers A_k and B_k be such that $|A_k| = |A|$ and $|B_k| = |B|$. Suppose, the state $\rho_{A_1^n B_1^n E}$ is such that for every $k \in [n]$, there exists a channel $\mathcal{M}_k : R_k \to A_k B_k$ such that for all inputs X_{R_k},*

$$\mathrm{tr}_{A_k} \circ \mathcal{M}_k (X_{R_k}) = \mathrm{tr}(X) \theta^{(k)}_{B_k} \tag{5.72}$$

for some state $\theta^{(k)}_{B_k}$[9], *and a state* $\tilde{\rho}^{(k,0)}_{A_1^{k-1}B_1^{k-1}ER_k}$ *for which*

$$\frac{1}{2}\left\|\rho_{A_1^kB_1^kE} - \mathcal{M}_k(\tilde{\rho}^{(k,0)}_{A_1^{k-1}B_1^{k-1}ER_k})\right\|_1 \leq \epsilon. \tag{5.73}$$

Then, as long as $\mu := \left(\frac{4\sqrt{\epsilon}+\epsilon}{1-\epsilon/(|A||B|)^2}\log\frac{|A|^2|B|^2}{\epsilon}\right)^{1/3} = O\left(\epsilon^{1/6}\left(\log\frac{|A|^2|B|^2}{\epsilon}\right)^{1/3}\right)$ *lies in* $(0,1)$*, we have the following lower bound for the smooth min-entropy of* ρ*:*

$$\begin{aligned} H_{\min}^{\mu+\epsilon'}(A_1^n|B_1^nE)_\rho \geq \sum_{k=1}^n \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k(\omega)} - 3n\sqrt{\mu}\log(1+2|A|) \\ - \frac{\log(1+2|A|)}{\sqrt{\mu}}\left(\frac{2}{\mu^2} + 2\log\frac{1}{1-\mu^2} + g_1(\epsilon',\mu)\right) \end{aligned} \tag{5.74}$$

where the infimum is over all the input states $\omega_{R_k\tilde{R}_k}$ *to the channel* $\mathcal{M}_k$ *and* $g_1(x,y) := -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$*.*

The proof of this theorem follows almost the same approach as that of Theorem 5.7. The major difference being that here we use the stronger triangle inequality (Lemma 3.5) to bound the smooth min-entropy with the α-Rényi conditional entropy of an auxiliary state. We then use the chain rule for these entropies along with the independence conditions on B_k to derive the lower bound above similar to [**DFR20**].

Proof. **Case 1:** First let's consider states $\rho_{A_1^nB_1^nE}$, which have full rank. Let $\nu \in (0,1)$ be an arbitrarily chosen small parameter. For every $k \in [n]$, define the states

$$\tilde{\tilde{\rho}}^{(k,0)}_{A_1^{k-1}B_1^{k-1}ER_k} := (1-\nu)\tilde{\rho}^{(k,0)}_{A_1^{k-1}B_1^{k-1}ER_k} + \nu\tau_{A_1^{k-1}B_1^{k-1}ER_k} \tag{5.75}$$

$$\tilde{\tilde{\rho}}^{(k)}_{A_1^kB_1^kE} := \mathcal{M}_k\left(\tilde{\tilde{\rho}}^{(k,0)}_{A_1^{k-1}B_1^{k-1}ER_k}\right) \tag{5.76}$$

We make this modification to $\tilde{\rho}^{(k)}$ states so that $\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}$ is full rank. For each k these states satisfy

$$\begin{aligned} &\frac{1}{2}\left\|\rho_{A_1^kB_1^kE} - \tilde{\tilde{\rho}}^{(k)}_{A_1^kB_1^kE}\right\|_1 \\ &= \frac{1}{2}\left\|\rho_{A_1^kB_1^kE} - (1-\nu)\,\mathcal{M}_k\left(\tilde{\rho}^{(k,0)}_{A_1^{k-1}B_1^{k-1}ER_k}\right) - \nu\,\mathcal{M}_k\left(\tau_{A_1^{k-1}B_1^{k-1}ER_k}\right)\right\|_1 \\ &\leq \frac{1-\nu}{2}\left\|\rho_{A_1^kB_1^kE} - \mathcal{M}_k\left(\tilde{\rho}^{(k)}_{A_1^{k-1}B_1^{k-1}ER_k}\right)\right\|_1 + \nu \\ &\leq \epsilon + \nu. \end{aligned} \tag{5.77}$$

Now, for each k, we define the states

$$\omega^{(k,0)}_{A_1^{k-1}B_1^{k-1}R_kE} := \rho^{1/2}_{A_1^{k-1}B_1^{k-1}E}\left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}\right)^{-1/2}\tilde{\tilde{\rho}}^{(k,0)}_{A_1^{k-1}B_1^{k-1}R_kE}\left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}\right)^{-1/2}\rho^{1/2}_{A_1^{k-1}B_1^{k-1}E} \tag{5.78}$$

[9] This condition can be relaxed to requiring channels $\mathcal{M}_k$ such that for all input states $\sigma_{A_1^{k-1}B_1^{k-1}ER_k}$, the output state $\mathcal{M}_k(\sigma)$ satisfies the Markov chain condition $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$. This condition seems to imply the independence condition used in the theorem (see Appendix C.3).

$$\omega^{(k)}_{A_1^k B_1^k E} := \rho^{1/2}_{A_1^{k-1}B_1^{k-1}E}\left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}\right)^{-1/2}\tilde{\tilde{\rho}}^{(k)}_{A_1^k B_1^k E}\left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}\right)^{-1/2}\rho^{1/2}_{A_1^{k-1}B_1^{k-1}E} \tag{5.79}$$

$$= \mathcal{M}_k\left(\omega^{(k,0)}_{A_1^{k-1}B_1^{k-1}R_kE}\right). \tag{5.80}$$

Since, we defined $\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}$ to be full rank, we have that

$$\omega^{(k)}_{A_1^{k-1}B_1^{k-1}E} = \rho_{A_1^{k-1}B_1^{k-1}E}. \tag{5.81}$$

Using Lemma 5.3, we have that

$$\begin{aligned}\frac{1}{2}\left\|\rho_{A_1^kB_1^kE} - \omega^{(k)}_{A_1^kB_1^kE}\right\|_1 &\leq (\sqrt{2}+1)P(\rho_{A_1^kB_1^kE}, \tilde{\tilde{\rho}}_{A_1^kB_1^kE}) \\ &\leq (\sqrt{2}+1)\sqrt{2(\epsilon+\nu)} \\ &\leq 4\sqrt{\epsilon+\nu}.\end{aligned} \tag{5.82}$$

Let $\delta \in (0,1)$ be a small parameter (to be set equal to ϵ later). Finally, for every $k \in [n]$, we define the states

$$\bar{\rho}^{(k)}_{A_1^kB_1^kE} :=(1-\delta)\omega^{(k)}_{A_1^kB_1^kE} + \delta\tau_{A_kB_k}\otimes\rho_{A_1^{k-1}B_1^{k-1}E} \tag{5.83}$$

$$=(1-\delta)\omega^{(k)}_{A_1^kB_1^kE} + \delta\tau_{A_kB_k}\otimes\omega^{(k)}_{A_1^{k-1}B_1^{k-1}E} \tag{5.84}$$

where $\tau_{A_kB_k}$ is the completely mixed state on the registers A_k and B_k. Also, define $\bar{\rho}^{(0)}_E := \rho_E$. Let $\Delta_k : R_k \to A_kB_k$ be the map which traces out the register R_k and simply outputs $\tau_{A_kB_k}$ and let $\mathcal{M}^\delta_k := (1-\delta)\mathcal{M}_k + \delta\Delta_k$. Then, we have that

$$\begin{aligned}\bar{\rho}^{(k)}_{A_1^kB_1^kE} &= (1-\delta)\omega^{(k)}_{A_1^kB_1^kE} + \delta\tau_{A_kB_k}\otimes\omega^{(k)}_{A_1^{k-1}B_1^{k-1}E} \\ &= ((1-\delta)\mathcal{M}_k + \delta\Delta_k)\left(\omega^{(k,0)}_{A_1^{k-1}B_1^{k-1}R_kE}\right) \\ &= \mathcal{M}^\delta_k\left(\omega^{(k,0)}_{A_1^{k-1}B_1^{k-1}R_kE}\right).\end{aligned} \tag{5.85}$$

We also have that

$$\frac{1}{2}\left\|\bar{\rho}^{(k)}_{A_1^kB_1^kE} - \rho_{A_1^kB_1^kE}\right\|_1 \leq 4\sqrt{\epsilon+\nu} + \delta. \tag{5.86}$$

Using Corollary 5.6, this gives us

$$D(\rho_{A_1^kB_1^kE}\|\bar{\rho}^{(k)}_{A_1^kB_1^kE}) \leq \frac{4\sqrt{\epsilon+\nu}+\delta}{1-\delta/(|A||B|)^2}\log\frac{|A|^2|B|^2}{\delta}. \tag{5.87}$$

We define the above bound as $z(\epsilon+\nu,\delta)$. Once again, using Lemma 5.4 (for $A_k \leftarrow A_kB_k$, $B \leftarrow E$ and $\bar{\rho}^{(k)}_{A_1^kB} \leftarrow \bar{\rho}^{(k)}_{A_1^kB_1^kE}$) we have that auxiliary state

$$\begin{aligned}&\sigma_{A_1^nB_1^nE} \\ &:= \int_{-\infty}^{\infty} dt\beta_0(t)\prod_{k=0}^{n-1}\left[\left(\bar{\rho}^{(k)}_{A_1^kB_1^kE}\right)^{\frac{1-it}{2}}\left(\bar{\rho}^{(k+1)}_{A_1^kB_1^kE}\right)^{-\frac{1-it}{2}}\right]\cdot\bar{\rho}^{(n)}_{A_1^nB_1^nE}\cdot\prod_{k=n-1}^{0}\left[\left(\bar{\rho}^{(k+1)}_{A_1^kB_1^kE}\right)^{-\frac{1+it}{2}}\left(\bar{\rho}^{(k)}_{A_1^kB_1^kE}\right)^{\frac{1+it}{2}}\right]\end{aligned} \tag{5.88}$$

$$= \int_{-\infty}^{\infty} dt \beta_0(t) \rho_E^{\frac{1-it}{2}} \left(\bar{\rho}_E^{(1)}\right)^{-\frac{1-it}{2}} \left(\bar{\rho}_{A_1B_1E}^{(1)}\right)^{\frac{1-it}{2}} \left(\bar{\rho}_{A_1B_1E}^{(2)}\right)^{-\frac{1-it}{2}} \left(\bar{\rho}_{A_1^2B_1^2E}^{(2)}\right)^{\frac{1-it}{2}} \cdots \left(\bar{\rho}_{A_1^{n-1}B_1^{n-1}E}^{(n)}\right)^{-\frac{1-it}{2}} \cdot \bar{\rho}_{A_1^nB_1^nE}^{(n)} \cdot \left(\bar{\rho}_{A_1^{n-1}B_1^{n-1}E}^{(n)}\right)^{-\frac{1+it}{2}} \cdots \left(\bar{\rho}_{A_1^2B_1^2E}^{(2)}\right)^{\frac{1+it}{2}} \left(\bar{\rho}_{A_1B_1E}^{(2)}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}_{A_1B_1E}^{(1)}\right)^{\frac{1+it}{2}} \left(\bar{\rho}_E^{(1)}\right)^{-\frac{1+it}{2}} \rho_E^{\frac{1+it}{2}} \tag{5.89}$$

is a normalised state satisfying

$$D_m(\rho_{A_1^nB_1^nE} \| \sigma_{A_1^nB_1^nE}) \leq nz(\epsilon+\nu,\delta) \tag{5.90}$$

and

$$\sigma_{A_1^kB_1^kE}$$

$$= \int_{-\infty}^{\infty} dt \beta_0(t) \prod_{j=0}^{k-1} \left[\left(\bar{\rho}_{A_1^jB_1^jE}^{(j)}\right)^{\frac{1-it}{2}} \left(\bar{\rho}_{A_1^jB_1^jE}^{(j+1)}\right)^{-\frac{1-it}{2}} \right] \cdot \bar{\rho}_{A_1^kB_1^kE}^{(k)} \cdot \prod_{j=k-1}^{0} \left[\left(\bar{\rho}_{A_1^jB_1^jE}^{(j+1)}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}_{A_1^jB_1^jE}^{(j)}\right)^{\frac{1+it}{2}} \right] \tag{5.91}$$

$$= \mathcal{M}_k^{\delta} \Bigg(\int_{-\infty}^{\infty} dt \beta_0(t) \prod_{j=0}^{k-1} \left[\left(\bar{\rho}_{A_1^jB_1^jE}^{(j)}\right)^{\frac{1-it}{2}} \left(\bar{\rho}_{A_1^jB_1^jE}^{(j+1)}\right)^{-\frac{1-it}{2}} \right] \cdot \omega_{A_1^{k-1}B_1^{k-1}R_kE}^{(k,0)} \cdot \prod_{j=k-1}^{0} \left[\left(\bar{\rho}_{A_1^jB_1^jE}^{(j+1)}\right)^{-\frac{1+it}{2}} \left(\bar{\rho}_{A_1^jB_1^jE}^{(j)}\right)^{\frac{1+it}{2}} \right] \Bigg). \tag{5.92}$$

Let $\sigma_{A_1^{k-1}B_1^{k-1}R_kE}^{(k,0)}$ be the input state for $\mathcal{M}_k^{\delta}$ above, so that

$$\sigma_{A_1^kB_1^kE} = \mathcal{M}_k^{\delta}\left(\sigma_{A_1^{k-1}B_1^{k-1}R_kE}^{(k,0)}\right) \tag{5.93}$$

Let's define $\mu := z(\epsilon+\nu,\delta)^{1/3}$. Using the substate theorem (Theorem 2.26), we get the following bound from the above relative entropy bound

$$D_{\max}^{\mu}(\rho_{A_1^nB_1^nE} \| \sigma_{A_1^nB_1^nE}) \leq n\mu + \frac{1}{\mu^2} + \log\frac{1}{1-\mu^2}. \tag{5.94}$$

Let $\epsilon' \in (0,1)$ be an arbitrary small parameter such that $\epsilon' + \mu < 1$ and let $\alpha \in (1,2]$. We can now use the entropic triangle inequality in Lemma 3.5 to derive

$$H_{\min}^{\mu+\epsilon'}(A_1^n|B_1^nE)_\rho \geq \tilde{H}_\alpha^{\uparrow}(A_1^n|B_1^nE)_\sigma - \frac{\alpha}{\alpha-1}n\mu - \frac{1}{\alpha-1}\left(\frac{\alpha}{\mu^2} + \alpha\log\frac{1}{1-\mu^2} + g_1(\epsilon',\mu)\right) \tag{5.95}$$

Moreover, using Eq. 5.92, we can show that B_k is independent of $A_1^{k-1}B_1^{k-1}E$ in σ. For $\sigma_{A_1^{k-1}B_1^kE} = \mathrm{tr}_{A_k} \circ \mathcal{M}_k^{\delta}(\sigma_{A_1^{k-1}B_1^{k-1}R_kE}^{(k,0)})$, we have

$$\sigma_{A_1^{k-1}B_1^kE} = (1-\delta)\,\mathrm{tr}_{A_k} \circ \mathcal{M}_k(\sigma_{A_1^{k-1}B_1^{k-1}R_kE}^{(k,0)}) + \delta\tau_{B_k} \otimes \sigma_{A_1^{k-1}B_1^{k-1}E}^{(k,0)} \tag{5.96}$$

$$= \left((1-\delta)\theta_{B_k}^{(k)} + \delta\tau_{B_k}\right) \otimes \sigma_{A_1^{k-1}B_1^{k-1}E}. \tag{5.97}$$

In particular, we have that σ satisfies the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$. So, we can use [DFR20, Corollary 3.5] to show that for every $k \in [n]$

$$\begin{aligned}\tilde{H}^{\downarrow}_{\alpha}(A_1^k|B_1^kE)_{\sigma} &\geq \tilde{H}^{\downarrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\sigma} + \inf_{\omega_{R_k\tilde{R}_k}} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k^{\delta}(\omega)} \\ &\geq \tilde{H}^{\downarrow}_{\alpha}(A_1^{k-1}|B_1^{k-1}E)_{\sigma} + \inf_{\omega_{R_k\tilde{R}_k}} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k(\omega)} \end{aligned} \tag{5.98}$$

where we use the quasi-concavity of $\tilde{H}^{\downarrow}_{\alpha}$ [Tom16, Pg 73] in the second line. Consecutively using this bound in Eq. 5.95 gives us

$$\begin{aligned}H^{\mu+\epsilon'}_{\min}(A_1^n|B_1^nE)_{\rho} &\geq \sum_{k=1}^{n} \inf_{\omega_{R_k\tilde{R}_k}} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k(\omega)} - \frac{\alpha}{\alpha-1}n\mu - \frac{1}{\alpha-1}\left(\frac{\alpha}{\mu^2} + \alpha\log\frac{1}{1-\mu^2} + g_1(\epsilon',\mu)\right) \\ &\geq \sum_{k=1}^{n} \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k(\omega)} - n(\alpha-1)\log^2(1+2|A|) - \frac{\alpha}{\alpha-1}n\mu \\ &\quad - \frac{1}{\alpha-1}\left(\frac{\alpha}{\mu^2} + \alpha\log\frac{1}{1-\mu^2} + g_1(\epsilon',\mu)\right)\end{aligned}$$

where in the second line we have used [DFR20, Lemma B.9] which is valid as long as $\alpha < 1 + 1/\log(1+2|A|)$. Lastly, we choose $\alpha = 1 + \frac{\sqrt{\mu}}{\log(1+2|A|)}$ and use $\alpha < 2$ as an upper bound to derive

$$\begin{aligned}H^{\mu+\epsilon'}_{\min}(A_1^n|B_1^nE)_{\rho} &\geq \sum_{k=1}^{n} \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k(\omega)} - 3n\sqrt{\mu}\log(1+2|A|) \\ &\quad - \frac{\log(1+2|A|)}{\sqrt{\mu}}\left(\frac{2}{\mu^2} + 2\log\frac{1}{1-\mu^2} + g_1(\epsilon',\mu)\right)\end{aligned} \tag{5.99}$$

where $\mu = z(\epsilon+\nu,\delta)^{1/3}$. The above bound holds true for all $\nu > 0$. Therefore, it also holds for $\nu \to 0$. To derive the bound in the theorem statement, we fix $\delta = \epsilon$.

Case 2: If $\rho_{A_1^nB_1^nE}$ were not full rank then we can always select a full rank state $\rho'_{A_1^nB_1^nE}$ ε-close to ρ, which would satisfy Eq. 5.73 and hence the above inequality with $\epsilon \to \epsilon + \varepsilon$. Now we can take $\varepsilon \to 0$ and use the continuity of $H_{\min}$ to arrive at the above bound for such ρ. □

5.3.1. Testing for unstructured approximate EAT

We incorporate testing into the approximate EAT proven above. We begin by defining the testing channels $\mathcal{T}_k$. These channels measure the outputs A_k and B_k of the state ρ and output a result X_k based on these measurements. Concretely, for every $k \in [n]$ the channel $\mathcal{T}_k : A_kB_k \to A_kB_kX_k$ is of the form

$$\mathcal{T}_k(\omega_{A_kB_k}) = \sum_{a,b} \Pi^{(a)}_{A_k} \otimes \Pi^{(b)}_{B_k} \omega_{A_kB_k} \Pi^{(a)}_{A_k} \otimes \Pi^{(b)}_{B_k} \otimes |x(a,b)\rangle\langle x(a,b)|_{X_k} \tag{5.100}$$

where $\{\Pi_{A_k}^{(a)}\}_a$ and $\{\Pi_{B_k}^{(b)}\}_b$ are orthogonal projectors and $x(\cdot)$ is some deterministic function which uses the measurements a and b to create the output register X_k.

Next we define the min-tradeoff functions. We consider n channels $\mathcal{N}_k$ for $k \in [n]$ and assume that the registers X_k, which consist of the result of the testing channels $\mathcal{T}_k$ are isomorphic, that is, $X_k \equiv X$ for all k. Let $\mathbb{P}$ be the set of probability distributions over the alphabet of X registers. Let R be any register isomorphic to R_k. For a probability $q \in \mathbb{P}$ and a channel $\mathcal{N}_k : R_k \to A_k B_k$, we define the set

$$\Sigma(q|\mathcal{N}_k) := \{\nu_{A_k B_k X_k R} = \mathcal{T}_k \circ \mathcal{N}_k(\omega_{R_k R}) : \text{ for a state } \omega_{R_{k-1} R} \text{ such that } \nu_{X_k} = q\}. \tag{5.101}$$

Definition 5.9. *A function $f : \mathbb{P} \to \mathbb{R}$ is called a min-tradeoff function for the channels $\{\mathcal{N}_k\}_{k=1}^n$ if for every k and $q \in \mathbb{P}$, it satisfies*

$$f(q) \leq \inf_{\nu \in \Sigma(q|\mathcal{N}_k)} H(A_k|B_k R)_\nu. \tag{5.102}$$

We now state the unstructured approximate EAT with testing.

Theorem 5.10. *Let $\epsilon \in (0,1)$ and for every $k \in [n]$, the registers A_k and B_k be such that $|A_k| = |A|$ and $|B_k| = |B|$. Suppose, the state $\rho_{A_1^n B_1^n X_1^n E}$ is such that*

(1) The registers X_1^n can be recreated by applying the testing maps to the registers A_1^n and B_1^n, that is,

$$\rho_{A_1^n B_1^n X_1^n E} = \mathcal{T}_n \circ \cdots \circ \mathcal{T}_1(\rho_{A_1^n B_1^n E}) \tag{5.103}$$

(2) For every $k \in [n]$, there exists a channel $\mathcal{M}_k : R_k \to A_k B_k$ and a state $\theta_{B_k}^{(k)}$ such that

$$\mathrm{tr}_{X_k} \circ \mathcal{T}_k \circ \mathcal{M}_k = \mathcal{M}_k \tag{5.104}$$

$$\mathrm{tr}_{A_k} \circ \mathcal{M}_k(X_{R_k}) = \mathrm{tr}(X)\theta_{B_k}^{(k)} \quad \text{for all operators } X_{R_k} \tag{5.105}$$

and a state $\tilde{\rho}_{A_1^{k-1} B_1^{k-1} E R_k}^{(k,0)}$ for which

$$\frac{1}{2}\left\| \rho_{A_1^k B_1^k E} - \mathcal{M}_k(\tilde{\rho}_{A_1^{k-1} B_1^{k-1} E R_k}^{(k,0)}) \right\|_1 \leq \epsilon. \tag{5.106}$$

Then, for an event Ω defined using X_1^n, an affine min-tradeoff function f for $\{\mathcal{M}_k\}_{k=1}^n$ such that for every $x_1^n \in \Omega$, $f(\mathrm{freq}(x_1^n)) \geq h$, we have

$$\begin{aligned} H_{\min}^{\mu'+\epsilon'}(A_1^n|B_1^n E)_{\rho|\Omega} \geq n(h - V(3\sqrt{\mu} + 4\epsilon) - g_2(2\epsilon)) \\ - \frac{V}{\sqrt{\mu}}\left(2\log\frac{1}{\Pr_\rho(\Omega) - \mu} + \frac{2}{\mu^2} + 2\log\frac{1}{1-\mu^2} + g_1(\epsilon', \mu')\right) \end{aligned} \tag{5.107}$$

where

$$\mu := \left(\frac{8\sqrt{\epsilon} + 2\epsilon}{1 - \epsilon^2/(|A||B|)^2} \log \frac{|A||B|}{\epsilon}\right)^{1/3} \tag{5.108}$$

$$\mu' := 2\sqrt{\frac{\mu}{\Pr_\rho(\Omega)}} \tag{5.109}$$

$$V := \log(1 + 2|A|) + 2\lceil \|\nabla f\|_\infty \rceil \tag{5.110}$$

$$g_1(x,y) := -\log(1 - \sqrt{1 - x^2}) - \log(1 - y^2) \tag{5.111}$$

$$g_2(x) := x \log \frac{1}{x} + (1 + x)\log(1 + x) \tag{5.112}$$

and $\epsilon' \in (0,1)$ *such that* $\mu' + \epsilon' < 1$.

The proof for this is similar to that of Theorem 5.8. We provide it in Appendix C.4

5.3.2. Dependence of smoothing parameter on the approximation parameter

It is evident that the smoothing parameter must depend on the approximation parameter, ϵ, for the unstructured approximate EAT in its current form. To illustrate this, consider a distribution $p_{A_1^n B}$ where $B = 1$ with probability $(1 - \epsilon)$ and $B = 0$ otherwise. In this distribution, A_1^n is sampled uniformly at random from $\{0,1\}^n$ if $B = 1$, and otherwise set to a constant string. For every k, this distribution satisfies

$$p_{A_1^k B} \approx_{O(\epsilon)} \Delta(p_{A_1^{k-1} B}) \tag{5.113}$$

where Δ is a channel that disregards its input and uniformly samples a bit A_k. Thus, this distribution meets the requirements for the unstructured approximate EAT. However, for this distribution, $H_{\min}^{\epsilon}(A_1^n|B)_p = n$, but for any $\epsilon' \leq \epsilon/2$, we have $H_{\min}^{\epsilon'}(A_1^n|B)_p = O(\log 1/\epsilon)$. This example demonstrates the necessity of the smoothing parameter's dependence on ϵ.

Nevertheless, it appears possible to decouple this dependence in certain interesting cases, such as sequential DIQKD with leakage and parallel DIQKD. It seems that the smoothing parameter's dependence on ϵ is required to remove the case of "correlated failure". In the aforementioned example, it is necessary to exclude the case where $B = 0$ and A_1^n are perfectly determined (highly correlated). Consider, however, the state $\rho_{X_1^n Y_1^n A_1^n B_1^n E}$ produced in a DIQKD protocol with imperfections or leakages (X_1^n and Y_1^n represent Alice and Bob's questions, A_1^n and B_1^n their answers, and E the adversary's register), where for each k

$$\rho_{X_1^k Y_1^k A_1^k B_1^k E} \approx_\epsilon \mathcal{M}_k\left(\rho^{(k,0)}_{X_1^{k-1} Y_1^{k-1} A_1^{k-1} B_1^{k-1} R_k E}\right) \tag{5.114}$$

for a channel $\mathcal{M}_k$, which plays the CHSH game between Alice and Bob, as one would expect in DIQKD with no leakage. We can use the unstructured EAT to prove that the entropy of the answers with respect to the questions and E is large for such a state. Drawing an analogy with the example distribution p above, we can deduce that the smoothing parameter must depend on ϵ to remove the case of correlated failure, whereby all the games fail together. However, this can also be achieved through testing, similar to the approach adopted in [MD24a]. By measuring the winning probability of the CHSH game on a random sample of games, we can determine if it is sufficiently high. If so, we can conclude that the CHSH game must have been played using "good" entangled quantum states between the two parties, and an event of correlated failure must not have occurred. While some additional assumptions may be required regarding the side information, this approach could potentially allow for an arbitrary smoothing parameter in such cases. This is an interesting line of research to pursue in the future.

5.4. Alternative proof for the universal chain rule

The proofs so far in this chapter have followed the approach of creating an auxiliary state using certain conditional states, proving that this state has a small relative entropy from the original state and reducing the original problem to a simpler one in terms of this auxiliary state. This is also the general approach we followed in Chapter 3. In Sec. 3.3 we provide a simple and intuitive alternate proof technique that lower bounds the smooth min-entropy for the classical approximately independent registers problem in an *elementwise* fashion. We are now ready to generalise this proof. Here we use it to provide an alternative proof of the universal chain rule for the smooth min-entropy.

5.4.1. Classical proof

We begin by once again first sketching the proof in the classical case. Recall that we would like to prove that for a probability distribution $p_{A_1^n B}$,

$$H_{\min}^{g_1(\epsilon)}(A_1^n|B)_p \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_p - n g_2(\epsilon) - k(\epsilon). \tag{5.115}$$

Let $\lambda_k := H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_p$. Following Sec. 5.2.1, for $k \in [n]$, let $q^{(k)}_{A_1^k B}$ be the distribution, such that

$$\frac{1}{2}\left\| p_{A_1^k B} - q^{(k)}_{A_1^k B} \right\|_1 \leq \epsilon \tag{5.116}$$

$$H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B)_p = \lambda_k = H_{\min}^{\downarrow}(A_k|A_1^{k-1}B)_{q^{(k)}}. \tag{5.117}$$

Similar to Eq. 5.27, we also have that

$$\frac{1}{2}\left\|p_{A_1^k B} - p_{A_1^{k-1}B} q^{(k)}_{A_k|A_1^{k-1}B}\right\|_1 \leq 2\epsilon. \tag{5.118}$$

Using Lemma 3.8, for every $k \in [n]$, we know that the set

$$\begin{aligned} B_k &:= \left\{(a_1^n, b) : p(a_1^k b) > (1+\sqrt{2\epsilon}) p(a_1^{k-1} b) q^{(k)}(a_k|a_1^{k-1}b)\right\} \\ &= \left\{(a_1^n, b) : p(a_k|a_1^{k-1} b) > (1+\sqrt{2\epsilon}) q^{(k)}(a_k|a_1^{k-1}b)\right\} \end{aligned}$$

satisfies $\Pr_p(B_k) \leq 3\sqrt{\epsilon}$. We can now define $L = \sum_{k=1}^n \chi_{B_k}$, which is a random variable that simply counts the number of bad sets B_k an element (a_1^n, b) belongs to. Using the Markov inequality, we have

$$\Pr_p\left[L > n\epsilon^{\frac{1}{4}}\right] \leq \frac{\mathbb{E}_p[L]}{n\epsilon^{\frac{1}{4}}} \leq 3\epsilon^{\frac{1}{4}}.$$

We can define the bad set $\mathcal{B} := \left\{(a_1^n, b) : L(a_1^n, b) > n\epsilon^{\frac{1}{4}}\right\}$. Then for the subnormalised distribution $\tilde{p}_{A_1^n B}$ defined as

$$\tilde{p}_{A_1^n B}(a_1^n, b) = \begin{cases} p_{A_1^n B}(a_1^n, b) & (a_1^n, b) \notin \mathcal{B} \\ 0 & \text{else} \end{cases},$$

we have $P(\tilde{p}_{A_1^n B}, p_{A_1^n B}) \leq \sqrt{6}\epsilon^{1/8}$. Further, note that for every $(a_1^n, b) \notin \mathcal{B}$, we have

$$\begin{aligned} p(a_1^n|b) &= \prod_{k=1}^n p(a_k|a_1^{k-1}, b) \\ &= \prod_{k:(a_1^n,b)\notin B_k} p(a_k|a_1^{k-1}, b) \prod_{k:(a_1^n,b)\in B_k} p(a_k|a_1^{k-1}, b) \\ &\leq (1+\sqrt{2\epsilon})^n \prod_{k:(a_1^n,b)\notin B_k} q^{(k)}(a_k|a_1^{k-1}, b) \prod_{k:(a_1^n,b)\in B_k} e^{\log|A|-\lambda_k} \\ &\leq (1+\sqrt{2\epsilon})^n \prod_{k:(a_1^n,b)\notin B_k} e^{-\lambda_k} \prod_{k:(a_1^n,b)\in B_k} e^{\log|A|-\lambda_k} \\ &\leq (1+\sqrt{2\epsilon})^n \; e^{n\epsilon^{1/4}\log|A|} \prod_{k=1}^n e^{-\lambda_k} \end{aligned}$$

where in the third line we have used the fact that if $(a_1^n, b) \notin B_k$, then $p(a_k|a_1^{k-1}b) \leq (1+\sqrt{\epsilon}) q^{(k)}(a_k|a_1^{k-1}b)$ and if $(a_1^n, b) \in B_k$ then $p(a_k|a_1^{k-1}b) \leq 1 \leq \exp(\log|A| - \lambda_k)$ since $\lambda_k \leq \log|A|$. In the last line we have used the fact that for $(a_1^k, b) \notin \mathcal{B}$, we have $|\{k \in [n] : (a_1^n, b) \in B_k\}| = L(a_1^n, b) \leq n\epsilon^{\frac{1}{4}}$. This proves the following lower bound for the smooth min-entropy of ρ

$$H_{\min}^{\downarrow, \sqrt{6}\epsilon^{1/8}}(A_1^n|B) \geq \sum_{k=1}^n H_{\min}^{\downarrow,\epsilon}(A_k|A_1^{k-1}B) - n\epsilon^{1/4}\log|A| - n\log(1+\sqrt{\epsilon}). \tag{5.119}$$

There are many hurdles to generalising this proof to the quantum setting. Firstly, the correct generalisation of Lemma 3.8 is quite difficult. Secondly, the proof above bounds the conditional probability $p(a_1^n|b)$ for every element $(a_1^n, b) \notin \mathcal{B}$ differently depending on which

bad sets $(\mathcal{B}_k)_k$ it belongs in. It is not obvious how such an argument can be carried out for quantum states. For example, one cannot simply use the eigenvalues of $\rho_{A_1^n B}$ instead of $p(a_1^n b)$ since the eigenvectors for each of the partial states $\rho_{A_1^k B}$ differ. Lastly, in the quantum case it does not seem that the Markov inequality can be used to identify the bad sets and the smoothed state as we did above.

The correct generalisation of Lemma 3.8 was proven by [**fu23**]. We state it in Lemma 5.13. To address the second and third hurdles mentioned above, we modify the classical proof before proceeding to the quantum proof.

Instead of figuring out the bad set $\mathcal{B}$ and eliminating the elements belonging to this set to construct the smoothed distribution, we use the substate theorem and the entropic triangle inequality. Define the auxiliary subnormalised distribution

$$q(a_1^n b) := \delta^{L(a_1^n b)} p(a_1^n b) \tag{5.120}$$

for some small $\delta \in (0, \frac{1}{|A|})$. The $\delta^{L(a_1^n b)}$ factor in q simply guarantees that elements which have a large L (are *bad*) are significantly damped. This ensures that for all a_1^n, b, we have that

$$\begin{aligned} q(a_1^n b) &= \delta^{L(a_1^n b)} p(a_1^n b) \\ &= p(b) \prod_{k=1}^{n} \delta^{\chi_{B_k}(a_1^n b)} p(a_k | a_1^{k-1} b) \\ &\leq p(b) \prod_{k=1}^{n} (1 + \sqrt{2\epsilon}) e^{-\lambda_k}. \end{aligned} \tag{5.121}$$

The last inequality above is guaranteed both when $a_1^n, b \notin \mathcal{B}_k$ and when $a_1^n, b \in \mathcal{B}_k$ since we chose $\delta < \frac{1}{|A|}$. This gives us a lower bound the min-entropy of q. Moreover, we can easily show that the relative entropy between p and q is small:

$$\begin{aligned} D(p_{A_1^n B} \| q_{A_1^n B}) &= \mathbb{E}_p \left[\log \frac{p(A_1^n B)}{q(A_1^n B)} \right] \\ &= \mathbb{E}_p \left[L \log \frac{1}{\delta} \right] \\ &= 3n\sqrt{\epsilon} \log \frac{1}{\delta}. \end{aligned} \tag{5.122}$$

Now, we can bound the smooth min-entropy of p in terms of the min-entropy of q by using the substate theorem and the entropic triangle inequality. We will generalise this proof to the quantum case.

5.4.2. Lemmas

In this section, we will primarily use the following variant of the smooth min-entropy, which has previously appeared in [ABJT20]:

$$\bar{H}^{\epsilon}_{\min}(A|B)_\rho := \sup\left\{\lambda \in \mathbb{R} : \text{ for } \tilde{\rho}_{AB} \in B_\epsilon(\rho_{AB}) \text{ such that } \tilde{\rho}_{AB} \leq e^{-\lambda}\mathbb{1}_A \otimes \rho_B\right\}. \tag{5.123}$$

Note that $\bar{H}^{\epsilon}_{\min}(A|B)_\rho \geq -\log|A|$ because $\rho_{AB} \leq |A|\,\mathbb{1}_A \otimes \rho_B$. Also, $\bar{H}^{\epsilon}_{\min}(A|B)_\rho \leq \log \frac{|A|}{1-\epsilon^2}$. To see this note that for any $\tilde{\rho}_{AB}$ and λ, such that

$$\begin{aligned} &\tilde{\rho}_{AB} \leq e^{-\lambda}\mathbb{1}_A \otimes \rho_B \\ \Rightarrow &\operatorname{tr}\tilde{\rho}_{AB} \leq e^{-\lambda}|A| \\ \Rightarrow &\lambda \leq \log \frac{|A|}{\operatorname{tr}\tilde{\rho}_{AB}} \end{aligned} \tag{5.124}$$

where we have simply taken a trace in the second line. Finally, use the fact that $\operatorname{tr}\tilde{\rho}_{AB} \geq 1-\epsilon^2$ for all $\tilde{\rho}_{AB} \in B_\epsilon(\rho_{AB})$.

This smooth min-entropy can be lower bounded using the $H^{\downarrow,\epsilon}_{\min}$ min-entropy as the following lemma shows.

Lemma 5.11. *For a normalized quantum state ρ_{AB} and $\epsilon \geq 0$, we have*

$$\bar{H}^{2\epsilon}_{\min}(A|B)_\rho \geq H^{\downarrow,\epsilon}_{\min}(A|B)_\rho. \tag{5.125}$$

Proof. Let $\lambda = H^{\downarrow,\epsilon}_{\min}(A|B)_\rho$ and the state $\tilde{\rho}_{AB} \in B_\epsilon(\rho_{AB})$ be such that

$$\tilde{\rho}_{AB} \leq e^{-\lambda}\mathbb{1}_A \otimes \tilde{\rho}_B. \tag{5.126}$$

By Lemma C.1, we have that $\eta_{AB} := \rho_B^{1/2} U_B \tilde{\rho}_B^{-1/2} \tilde{\rho}_{AB} \tilde{\rho}_B^{-1/2} U_B^\dagger \rho_B^{1/2}$ satisfies $P(\rho_{AB}, \eta_{AB}) \leq 2\epsilon$. Clearly, this state also satisfies

$$\eta_{AB} \leq e^{-\lambda}\mathbb{1}_A \otimes \rho_B. \tag{5.127}$$

□

We also require the following operator inequality relating an operator and its compressions.

Lemma 5.12. *For a positive operator $X \geq 0$ and orthogonal projectors Π and $\Pi_\perp = \mathbb{1} - \Pi$, we have*

$$\Pi_\perp X \Pi_\perp \leq 2X + 2\Pi X \Pi. \tag{5.128}$$

Proof. We will write the operator X as the block matrix

$$X = \begin{pmatrix} X_1 & X_2 \\ X_2^* & X_3 \end{pmatrix} \tag{5.129}$$

where the blocks are partitioned according to the direct sum $\mathrm{im}(\Pi) \oplus \mathrm{im}(\Pi_\perp)$. The statement in the Lemma is now equivalent to proving that

$$0 \leq \begin{pmatrix} 4X_1 & 2X_2 \\ 2X_2^* & X_3 \end{pmatrix}. \tag{5.130}$$

Call the matrix above X'. Note that using the Schur's complement [**Bha07**, Exercise 1.3.5] $X \geq 0$ implies that $X_3 \geq 0$, $\mathrm{im}(X_2^*) \subseteq \mathrm{im}(X_3)$ and

$$X_1 \geq X_2 X_3^{-1} X_2^*. \tag{5.131}$$

Using Schur's complement characterization again for X', we see that $X' \geq 0$ is equivalent to $X_3 \geq 0$, $\mathrm{im}(X_2^*) \subseteq \mathrm{im}(X_3)$ and

$$4X_1 \geq (2X_2) X_3^{-1} (2X_2)^*$$

which are all true because of the corresponding relations for X itself. □

The following lemma correctly generalises Lemma 3.8. It was proven by user:fedja in response to a question by user:noel (pseudonym used by AM) on MathOverflow [**fu23**]. We reproduce the proof in Appendix C.5 for completeness.

Lemma 5.13 ([**fu23**]). *For $\epsilon \in [0,1]$, and subnormalized quantum states ρ and σ on the finite dimensional Hilbert space $\mathcal{X}$ such that $\frac{1}{2}\|\rho - \sigma\|_1 \leq \epsilon$, there exists an orthogonal projector Π such that*

$$\Pi\rho\Pi \leq (1 + g_1(\epsilon))\sigma \tag{5.132}$$

and

$$\mathrm{tr}((\mathbb{1} - \Pi)\rho) \leq g_2(\epsilon) \tag{5.133}$$

for the small functions

$$g_1(\epsilon) := \frac{8}{3}(1 + \epsilon^{1/3})\epsilon^{1/3} \log \frac{1}{\epsilon} + (1 + \epsilon^{1/3} + \epsilon^{2/3})\epsilon^{1/3} = O\left(\epsilon^{1/3} \log \frac{1}{\epsilon}\right) \tag{5.134}$$

$$g_2(\epsilon) := 4(1 + \epsilon^{1/3})\epsilon^{1/3} + 2\epsilon = O(\epsilon^{1/3}) \tag{5.135}$$

5.4.3. Quantum proof

In this section, we will generalise the classical proof presented earlier to the quantum case. We will prove a statement of the following form for all normalised quantum states $\rho_{A_1^n B}$:

$$H_{\min}^{g_1'(\epsilon)}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} \bar{H}_{\min}^{\epsilon}(A_k|A_1^{k-1}B)_\rho - n g_2'(\epsilon) - k(\epsilon) \tag{5.136}$$

where g_1' and g_2' are small functions of ϵ and k is a general function of ϵ. This will be sufficient to prove the universal chain rule.

For every k, let $\lambda_k := \bar{H}_{\min}^{\epsilon}(A_k|A_1^{k-1}B)_\rho$ and $\tilde{\rho}_{A_1^k B}^{(k)}$ be the subnormalised state such that

$$P(\rho_{A_1^k B}, \tilde{\rho}_{A_1^k B}^{(k)}) \leq \epsilon \tag{5.137}$$

$$\tilde{\rho}_{A_1^k B}^{(k)} \leq e^{-\lambda_k} \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B}. \tag{5.138}$$

Using the Fuchs-van de Graaf inequality, for each k, we also have

$$\frac{1}{2} \left\| \rho_{A_1^k B} - \tilde{\rho}_{A_1^k B}^{(k)} \right\|_1 \leq \epsilon. \tag{5.139}$$

We will now define the projectors $P_{A_1^k B}^{(k,l)}$ for every $k \in [n]$ and the label $l \in \{g,b\}$ ({good, bad}). We define the *good* projector $P_{A_1^k B}^{(k,g)}$ to be the projector given by Lemma 5.13 when applied to the states $\rho_{A_1^k B}$ and $\tilde{\rho}_{A_1^k B}^{(k)}$, so that it satisfies

$$P_{A_1^k B}^{(k,g)} \rho_{A_1^k B} P_{A_1^k B}^{(k,g)} \leq (1 + g_1(\epsilon)) \tilde{\rho}_{A_1^k B}^{(k)} \tag{5.140}$$

and

$$\operatorname{tr}(P_{A_1^k B}^{(k,g)} \rho_{A_1^k B}) \geq 1 - g_2(\epsilon) \tag{5.141}$$

for g_1 and g_2 as defined in Lemma 5.13. Further, define its orthogonal complement as the *bad* projector,

$$P_{A_1^k B}^{(k,b)} := \mathbb{1}_{A_1^k B} - P_{A_1^k B}^{(k,g)}. \tag{5.142}$$

The label l in $P_{A_1^k B}^{(k,l)}$ will allow us to succinctly refer to these two projectors together. Define the subnormalised state

$$\eta_{A_1^n B} := P_{A_1 B}^{(1,g)} P_{A_1^2 B}^{(2,g)} \cdots P_{A_1^n B}^{(n,g)} \ \rho_{A_1^n B} \ P_{A_1^n B}^{(n,g)} \cdots P_{A_1^2 B}^{(2,g)} P_{A_1 B}^{(1,g)}. \tag{5.143}$$

For this state, observe that

$$\begin{aligned} &P_{A_1 B}^{(1,g)} P_{A_1^2 B}^{(2,g)} \cdots P_{A_1^n B}^{(n,g)} \ \rho_{A_1^n B} \ P_{A_1^n B}^{(n,g)} \cdots P_{A_1^2 B}^{(2,g)} P_{A_1 B}^{(1,g)} \\ &\quad \leq (1 + g_1(\epsilon)) P_{A_1 B}^{(1,g)} P_{A_1^2 B}^{(2,g)} \cdots P_{A_1^{n-1} B}^{(n-1,g)} \ \tilde{\rho}_{A_1^n B}^{(n)} \ P_{A_1^{n-1} B}^{(n-1,g)} \cdots P_{A_1^2 B}^{(2,g)} P_{A_1 B}^{(1,g)} \end{aligned}$$

$$
\begin{aligned}
&\leq e^{-\lambda_n}(1+g_1(\epsilon))P^{(1,g)}_{A_1B}P^{(2,g)}_{A_1^2B}\cdots P^{(n-1,g)}_{A_1^{n-1}B}\ \mathbb{1}_{A_n}\otimes\rho_{A_1^{n-1}B}\ P^{(n-1,g)}_{A_1^{n-1}B}\cdots P^{(2,g)}_{A_1^2B}P^{(1,g)}_{A_1B}\\
&= e^{-\lambda_n}(1+g_1(\epsilon))\,\mathbb{1}_{A_n}\otimes P^{(1,g)}_{A_1B}P^{(2,g)}_{A_1^2B}\cdots P^{(n-1,g)}_{A_1^{n-1}B}\ \rho_{A_1^{n-1}B}\ P^{(n-1,g)}_{A_1^{n-1}B}\cdots P^{(2,g)}_{A_1^2B}P^{(1,g)}_{A_1B}\\
&\leq \cdots\\
&\leq e^{-\sum_{k=1}^n\lambda_k}(1+g_1(\epsilon))^n\,\mathbb{1}_{A_1^n}\otimes\rho_B
\end{aligned}
$$

which implies that $H_{\min}(A_1^n|B)_\eta \geq \sum_{k=1}^n \bar{H}^\epsilon_{\min}(A_k|A_1^{k-1}B)_\rho - n\log(1+g_1(\epsilon))$. Moreover, the distance between η and ρ is

$$
\begin{aligned}
P(\rho_{A_1^nB},\,\eta_{A_1^nB}) &= P(\rho_{A_1^nB},\ P^{(1,g)}_{A_1B}P^{(2,g)}_{A_1^2B}\cdots P^{(n,g)}_{A_1^nB}\ \rho_{A_1^nB}\ P^{(n)}_{A_1^nB}\cdots P^{(2,g)}_{A_1^2B}P^{(1,g)}_{A_1B})\\
&\leq P(\rho_{A_1^nB},\ P^{(1,g)}_{A_1B}\ \rho_{A_1^nB}\ P^{(1,g)}_{A_1B})\\
&\qquad + P(P^{(1,g)}_{A_1B}\ \rho_{A_1^nB}\ P^{(1,g)}_{A_1B},\ P^{(1,g)}_{A_1B}P^{(2,g)}_{A_1^2B}\cdots P^{(n,g)}_{A_1^nB}\ \rho_{A_1^nB}\ P^{(n,g)}_{A_1^nB}\cdots P^{(2,g)}_{A_1^2B}P^{(1,g)}_{A_1B})\\
&\leq P(\rho_{A_1^nB}, P^{(1,g)}_{A_1B}\ \rho_{A_1^nB}\ P^{(1,g)}_{A_1B}) + P(\rho_{A_1^nB},\ P^{(2,g)}_{A_1^2B}\cdots P^{(n,g)}_{A_1^nB}\ \rho_{A_1^nB}\ P^{(n,g)}_{A_1^nB}\cdots P^{(2,g)}_{A_1^2B})\\
&\leq \cdots\\
&\leq \sum_{k=1}^n P(\rho_{A_1^nB},\ P^{(k,g)}_{A_1^kB}\ \rho_{A_1^nB}\ P^{(k,g)}_{A_1^kB})\\
&\leq n\sqrt{2g_2(\epsilon)}
\end{aligned}
$$

where we have used the gentle measurement lemma [**Wat18**, Proposition 3.14] for the last line. The distance between η and ρ grows like n times a small function and the min-entropy of η satisfies $H_{\min}(A_1^n|B)_\eta \geq \sum_{k=1}^n \bar{H}^\epsilon_{\min}(A_k|A_1^{k-1}B)_\rho - n\log(1+g_1(\epsilon))$. If we were able to show that the relative entropy distance between these two states also grew like n times a small function, similar to the purified distance then we could use the substate theorem and the entropic triangle inequality to prove a bound of the required form. However, we can't directly prove that the relative entropy between these two states is small because in general the projectors sandwiching the state ρ in η could lead to a situation, where $\rho \not\ll \eta$, which would cause their relative entropy diverge. To remedy this, one could imagine adding a small amount of the complement projector to these projectors, so that we sandwich with $P^{(k,g)}_{A_1^kB} + \delta P^{(k,b)}_{A_1^kB}$ instead of simply $P^{(k)}_{A_1^kB}$. Under no further assumptions on ρ, the question of whether the divergence of ρ and η is finite or not in this case reduces to the question: given a state σ, a projector Π and its complement $\Pi_\perp := \mathbb{1} - \Pi$, is

$$
\sigma \ll (\Pi + \delta\Pi_\perp)\sigma(\Pi + \delta\Pi_\perp)? \tag{5.144}
$$

The answer to this is easily seen to be negative when one consider the pure state $\sigma = |u\rangle\langle u|$. In this case, the above is equivalent to asking if $|u\rangle\langle u| \ll |v\rangle\langle v|$ for some vector $|v\rangle \neq |u\rangle$, which is not true.

The next simple remedy that comes to mind is to use a "pinching type" disturbance to ensure that the divergence is finite. For $\delta \in (0,1)$, a projector Π and its complement $\Pi_\perp$, define the CP map $\mathcal{P}_\delta[\Pi]$ as

$$\mathcal{P}_\delta[\Pi](\sigma) := \Pi\sigma\Pi + \delta\Pi_\perp\sigma\Pi_\perp. \tag{5.145}$$

Then, by the pinching inequality, for every Π, we have

$$\begin{aligned}\sigma &\leq \frac{2}{\delta}\left(\delta\Pi\sigma\Pi + \delta\Pi_\perp\sigma\Pi_\perp\right)\\ &\leq \frac{2}{\delta}\mathcal{P}_\delta[\Pi](\sigma)\end{aligned}$$

which implies $D_{\max}(\sigma\|\mathcal{P}_\delta[\Pi](\sigma)) \leq O(\log 1/\delta)$– a bound that should be sufficient for our purposes. Therefore, one can try proving the chain rule using the subnormalised state

$$\begin{aligned}&\mathcal{P}_\delta[P^{(1,g)}_{A_1B}] \circ \mathcal{P}_\delta[P^{(2,g)}_{A_1^2B}] \circ \cdots \circ \mathcal{P}_\delta[P^{(n,g)}_{A_1^nB}](\rho_{A_1^nB})\\ &\qquad = \sum_{l_1^n \in \{g,b\}^n} \delta^{\omega_b(l_1^n)} P^{(1,l_1)}_{A_1B} P^{(2,l_2)}_{A_1^2B} \cdots P^{(n,l_n)}_{A_1^nB}\ \rho_{A_1^nB}\ P^{(n,l_n)}_{A_1^nB} \cdots P^{(2,l_2)}_{A_1^2B} P^{(1,l_1)}_{A_1B}\end{aligned} \tag{5.146}$$

where $\omega_b(l_1^n) := |\{i : l_i = b\}|$ is the weight of b labels in the string l_1^n. In our proof below, we do not attempt to directly identify the correct modified state ourselves. Similar to the proofs before, we instead leave this question for the generalised GT inequality. We will use an exponential generalisation of the distribution q in Eq. 5.120. Variants of both of the above remedies make an appearance during the following proof.

Lemma 5.14. *For a full rank state $\rho_{A_1^nB}$ and $\epsilon \in (0,1)$, we have the chain rule*

$$\begin{aligned}H^\mu_{\min}(A_1^n|B)_\rho \geq \sum_{k=1}^n \bar{H}^\epsilon_{\min}(A_k|A_1^{k-1}B)_\rho &- n\log\left(1 + 4|A|^2\frac{\epsilon}{1-\epsilon^2}\right) - n\log(1+g_1(\epsilon))\\ &- n\log(1+\epsilon) - n(\mu+\mu^3) - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2}\end{aligned} \tag{5.147}$$

where

$$\mu := \left(8(1+\epsilon^{1/3})\log\frac{1}{\epsilon}\right)^{1/3}\epsilon^{1/9} = O\left(\epsilon^{1/9}\left(\log\frac{1}{\epsilon}\right)^{1/3}\right) \tag{5.148}$$

$$g_1(\epsilon) := \frac{8}{3}(1+\epsilon^{1/3})\epsilon^{1/3}\log\frac{1}{\epsilon} + (1+\epsilon^{1/3}+\epsilon^{2/3})\epsilon^{1/3} = O\left(\epsilon^{1/3}\log\frac{1}{\epsilon}\right) \tag{5.149}$$

as long as $\mu \in (0,1)$.

Proof. We retain the definitions of $\tilde{\rho}_{A_1^kB}$, λ_k, and $P^{(k,l)}_{A_1^kB}$ from the discussion above. Following the classical proof in Sec. 5.4.1, define $L_{A_1^nB} := \sum_{k=1}^n P^{(k,b)}_{A_1^kB}$ as the sum of the "bad" projectors. Then, we have that

$$\operatorname{tr}(L_{A_1^nB}\ \rho_{A_1^nB}) = \sum_{k=1}^n \operatorname{tr}(P^{(k,b)}_{A_1^kB}\ \rho_{A_1^kB})$$

$$\leq n g_2(\epsilon). \tag{5.150}$$

Let $\delta \in (0,1)$ be a parameter to be chosen later. We have that

$$\begin{aligned}
&n g_2(\epsilon) \log \frac{1}{\delta} \\
&\geq D\left(\rho_{A_1^n B} \| \exp\left(\log \rho_{A_1^n B} + \log(\delta) L_{A_1^n B}\right)\right) \\
&= \sup_{\omega_{A_1^n B} > 0} \left\{ \operatorname{tr}(\rho_{A_1^n B} \ \log \omega_{A_1^n B}) + 1 - \operatorname{tr} \exp\left(\log \omega_{A_1^n B} + \log(\delta) \sum_{k=1}^{n} P^{(k,b)}_{A_1^k B} + \log \rho_{A_1^n B} \right) \right\}
\end{aligned} \tag{5.151}$$

where we have used Eq. 5.150 in the first line and the variational expression in Eq. 5.18 in the second line. For the trace of exponential term in the last expression above, we can use the generalised Golden-Thompson inequality (Theorem 5.1) as

$$\begin{aligned}
&\operatorname{tr} \exp\left(\log \omega_{A_1^n B} + \log(\delta) \sum_{k=1}^{n} P^{(k,b)}_{A_1^k B} + \log \rho_{A_1^n B} \right) \\
&\leq \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}\left(\omega_{A_1^n B} \ e^{\frac{1-it}{2} \log(\delta) P^{(1,b)}_{A_1 B}} \cdots \ e^{\frac{1-it}{2} \log(\delta) P^{(n,b)}_{A_1^n B}} \ \rho_{A_1^n B} \ e^{\frac{1+it}{2} \log(\delta) P^{(n,b)}_{A_1^n B}} \cdots \ e^{\frac{1+it}{2} \log(\delta) P^{(1,b)}_{A_1 B}} \right) \\
&= \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}\big(\omega_{A_1^n B} \ (\delta^{\frac{1-it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B}) \cdots (\delta^{\frac{1-it}{2}} P^{(n,b)}_{A_1^n B} + P^{(n,g)}_{A_1^n B}) \ \rho_{A_1^n B} \\
&\qquad (\delta^{\frac{1+it}{2}} P^{(n,b)}_{A_1^n B} + P^{(n,g)}_{A_1^n B}) \cdots (\delta^{\frac{1+it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B}) \big),
\end{aligned} \tag{5.152}$$

where we have used the fact that for a projector P and $\beta \in \mathbb{C}$, $\exp(\beta P) = e^{\beta} P + (I - P)$ in the last line. Define the subnormalised state

$$\begin{aligned}
\eta_{A_1^n B} \coloneqq \int_{-\infty}^{\infty} dt \beta_0(t) &(\delta^{\frac{1-it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B}) \cdots (\delta^{\frac{1-it}{2}} P^{(n,b)}_{A_1^n B} + P^{(n,g)}_{A_1^n B}) \ \rho_{A_1^n B} \\
&(\delta^{\frac{1+it}{2}} P^{(n,b)}_{A_1^n B} + P^{(n,g)}_{A_1^n B}) \cdots (\delta^{\frac{1+it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B})
\end{aligned} \tag{5.153}$$

Observe that this is just a clever (and correct) way of implementing the first remedy we discussed in the motivation. From Eq. 5.152, we have

$$\operatorname{tr} \exp\left(\log \rho_{A_1^n B} + \log(\delta) \sum_{k=1}^{n} P^{(k,b)}_{A_1^k B} + \log \omega_{A_1^n B} \right) \leq \operatorname{tr}(\omega_{A_1^n B} \eta_{A_1^n B}) \tag{5.154}$$

Plugging this in Eq. 5.151, we get

$$\begin{aligned}
n g_2(\epsilon) \log \frac{1}{\delta} &\geq \sup_{\omega_{A_1^n B} > 0} \left\{ \operatorname{tr}(\rho_{A_1^n B} \ \log \omega_{A_1^n B}) + 1 - \operatorname{tr}(\omega_{A_1^n B} \eta_{A_1^n B}) \right\} \\
&= D_m(\rho_{A_1^n B} \| \eta_{A_1^n B}).
\end{aligned} \tag{5.155}$$

We will now show that the state η has a small smooth max-relative entropy distance from ρ. We use the substate theorem for this. However, we need to be a bit careful since η is not normalised.

Claim 5.15. *For $\mu(\epsilon,\delta) \coloneqq \left(g_2(\epsilon)\log\frac{1}{\delta}\right)^{1/3}$ (we will use the shorthand μ going forth)*

$$D^{\mu}_{\max}\left(\rho_{A_1^nB}\|\eta_{A_1^nB}\right) \leq n(\mu+\mu^3) + \frac{1}{\mu^2} + \log\frac{1}{1-\mu^2}. \tag{5.156}$$

Proof. Let $Z \coloneqq \operatorname{tr}(\eta_{A_1^nB})$. Then, using the data processing inequality on the above $D_m(\rho_{A_1^nB}\|\eta_{A_1^nB})$ bound, we see that

$$\log\frac{1}{Z} \leq n g_2(\epsilon)\log\frac{1}{\delta}. \tag{5.157}$$

We can also upper bound Z as

$$\begin{aligned}
\operatorname{tr}(\eta_{A_1^nB}) &= \int_{-\infty}^{\infty} dt\beta_0(t)\operatorname{tr}\Bigg(\left(\delta^{\frac{1-it}{2}}P^{(1,b)}_{A_1B} + P^{(1,g)}_{A_1B}\right)\cdots\left(\delta^{\frac{1-it}{2}}P^{(n,b)}_{A_1^nB} + P^{(n,g)}_{A_1^nB}\right)\rho_{A_1^nB} \\
&\qquad\qquad \left(\delta^{\frac{1+it}{2}}P^{(n,b)}_{A_1^nB} + P^{(n,g)}_{A_1^nB}\right)\cdots\left(\delta^{\frac{1+it}{2}}P^{(1,b)}_{A_1B} + P^{(1,g)}_{A_1B}\right)\Bigg) \\
&\leq \int_{-\infty}^{\infty} dt\beta_0(t)\left\|\delta P^{(1,b)}_{A_1B} + P^{(1,g)}_{A_1B}\right\|_{\infty}\operatorname{tr}\Bigg(\left(\delta^{\frac{1-it}{2}}P^{(2,b)}_{A_1^2B} + P^{(2,g)}_{A_1^2B}\right)\cdots\left(\delta^{\frac{1-it}{2}}P^{(n,b)}_{A_1^nB} + P^{(n,g)}_{A_1^nB}\right)\rho_{A_1^nB} \\
&\qquad\qquad \left(\delta^{\frac{1+it}{2}}P^{(n,b)}_{A_1^nB} + P^{(n,g)}_{A_1^nB}\right)\cdots\left(\delta^{\frac{1+it}{2}}P^{(2,b)}_{A_1^2B} + P^{(2,g)}_{A_1^2B}\right)\Bigg) \\
&\leq \int_{-\infty}^{\infty} dt\beta_0(t)\operatorname{tr}\Bigg(\left(\delta^{\frac{1-it}{2}}P^{(2,b)}_{A_1^2B} + P^{(2,g)}_{A_1^2B}\right)\cdots\left(\delta^{\frac{1-it}{2}}P^{(n,b)}_{A_1^nB} + P^{(n,g)}_{A_1^nB}\right)\rho_{A_1^nB} \\
&\qquad\qquad \left(\delta^{\frac{1+it}{2}}P^{(n,b)}_{A_1^nB} + P^{(n,g)}_{A_1^nB}\right)\cdots\left(\delta^{\frac{1+it}{2}}P^{(2,b)}_{A_1^2B} + P^{(2,g)}_{A_1^2B}\right)\Bigg) \\
&\leq \cdots \\
&\leq \operatorname{tr}\rho_{A_1^nB} \\
&= 1.
\end{aligned} \tag{5.158}$$

Using the substate theorem, we have

$$\begin{aligned}
D^{\mu}_{\max}\left(\rho_{A_1^nB}\|\frac{\eta_{A_1^nB}}{Z}\right) &\leq \frac{D_m\left(\rho_{A_1^nB}\|\frac{\eta_{A_1^nB}}{Z}\right)+1}{\mu^2} + \log\frac{1}{1-\mu^2} \\
&= \frac{D_m\left(\rho_{A_1^nB}\|\eta_{A_1^nB}\right) - \log\frac{1}{Z} + 1}{\mu^2} + \log\frac{1}{1-\mu^2} \\
&\leq \frac{n\mu^3+1}{\mu^2} + \log\frac{1}{1-\mu^2} \\
&\leq n\mu + \frac{1}{\mu^2} + \log\frac{1}{1-\mu^2}
\end{aligned} \tag{5.159}$$

where we have used $Z \leq 1$ in the third line.

We can get rid of Z normalisation factor by using Eq. 5.157

$$\begin{aligned} D^{\mu}_{\max}\left(\rho_{A_1^n B}\|\eta_{A_1^n B}\right) &= D^{\mu}_{\max}\left(\rho_{A_1^n B}\|\frac{\eta_{A_1^n B}}{Z}\right) + \log\frac{1}{Z} \\ &\leq n(\mu+\mu^3) + \frac{1}{\mu^2} + \log\frac{1}{1-\mu^2} \end{aligned} \tag{5.160}$$

□

Though η has a nice form, it is still difficult to use the properties of the projectors $P^{(k,g)}_{A_1^k B}$ (Eq. 5.140) with it. We will now show how we can dominate η using a state of the form in Eq. 5.146.

Claim 5.16. *Define the subnormalised state* $\sigma_{A_1^n B} := P_{\sqrt{\delta}}[P^{(1,g)}_{A_1 B}] \circ \cdots \circ P_{\sqrt{\delta}}[P^{(n,g)}_{A_1^n B}](\rho_{A_1^n B})$ *for* $P_{\sqrt{\delta}}[\Pi](X) = \Pi X \Pi + \sqrt{\delta}\Pi_\perp X \Pi_\perp$ *as defined in Eq. 5.145. For this state, we have*

$$D^{\mu}_{\max}\left(\rho_{A_1^n B}\|\sigma_{A_1^n B}\right) \leq n(\mu+\mu^3) + n\log(1+\sqrt{\delta}) + \frac{1}{\mu^2} + \log\frac{1}{1-\mu^2}. \tag{5.161}$$

Proof. Let X be an arbitrary positive operator, and Π and $\Pi_\perp = \mathbb{1} - \Pi$ be orthogonal projectors. Asymmetric pinching (Lemma 3.13) with the projectors Π and $\Pi_\perp$, and parameter $\sqrt{\delta}$ shows that for all $t \in \mathbb{R}$

$$\begin{aligned} (\Pi + \delta^{\frac{1-it}{2}}\Pi_\perp) X (\Pi + \delta^{\frac{1+it}{2}}\Pi_\perp) &\leq (1+\sqrt{\delta})\Pi X \Pi + \left(1 + \frac{1}{\sqrt{\delta}}\right)\delta \Pi_\perp X \Pi_\perp \\ &= (1+\sqrt{\delta}) P_{\sqrt{\delta}}[\Pi](X) \end{aligned} \tag{5.162}$$

Using Eq. 5.162 repeatedly, we have that for every $t \in \mathbb{R}$

$$\begin{aligned} &(\delta^{\frac{1-it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B}) \cdots (\delta^{\frac{1-it}{2}} P^{(n,b)}_{A_1^n B} + P^{(n,g)}_{A_1^n B})\ \rho_{A_1^n B}\ (\delta^{\frac{1+it}{2}} P^{(n,b)}_{A_1^n B} + P^{(n,g)}_{A_1^n B}) \cdots (\delta^{\frac{1+it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B}) \\ &\leq (1+\sqrt{\delta})(\delta^{\frac{1-it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B}) \cdots (\delta^{\frac{1-it}{2}} P^{(n-1,b)}_{A_1^{n-1} B} + P^{(n-1,g)}_{A_1^{n-1} B})\ P_{\sqrt{\delta}}[P^{(n,g)}_{A_1^n B}](\rho_{A_1^n B}) \\ &\qquad (\delta^{\frac{1+it}{2}} P^{(n-1,b)}_{A_1^{n-1} B} + P^{(n-1,g)}_{A_1^{n-1} B}) \cdots (\delta^{\frac{1+it}{2}} P^{(1,b)}_{A_1 B} + P^{(1,g)}_{A_1 B}) \\ &\leq \cdots \\ &\leq (1+\sqrt{\delta})^n P_{\sqrt{\delta}}[P^{(1,g)}_{A_1 B}] \circ \cdots \circ P_{\sqrt{\delta}}[P^{(n,g)}_{A_1^n B}](\rho_{A_1^n B}) \end{aligned} \tag{5.163}$$

which implies that

$$\eta_{A_1^n B} \leq (1+\sqrt{\delta})^n P_{\sqrt{\delta}}[P^{(1,g)}_{A_1 B}] \circ \cdots \circ P_{\sqrt{\delta}}[P^{(n,g)}_{A_1^n B}](\rho_{A_1^n B}). \tag{5.164}$$

This bound essentially says that the $D_{\max}$ between these two states is small:

$$D_{\max}(\eta_{A_1^n B}\|\sigma_{A_1^n B}) \leq n\log(1+\sqrt{\delta}) \tag{5.165}$$

Combining this with the smooth $D_{\max}$ bound between ρ and η (Eq. 5.160) shows the bound in the claim. □

The only thing left to do now is to show that min-entropy of the state σ is large, i.e., $\gtrsim \sum_{k=1}^{n} \lambda_k$ (recall $\lambda_k \coloneqq \bar{H}^{\epsilon}_{\min}(A_k|A_1^{k-1}B)_\rho$). Note that

$$\sigma_{A_1^n B} = \sum_{l_1^n \in \{g,b\}^n} \delta^{\frac{1}{2}\omega_b(l_1^n)} P^{(1,l_1)}_{A_1 B} P^{(2,l_2)}_{A_1^2 B} \cdots P^{(n,l_n)}_{A_1^n B} \; \rho_{A_1^n B} \; P^{(n,l_n)}_{A_1^n B} \cdots P^{(2,l_2)}_{A_1^2 B} P^{(1,l_1)}_{A_1 B} \tag{5.166}$$

where $\omega_b(l_1^n) \coloneqq |\{i : l_i = b\}|$ the weight of b labels in the string l_1^n. We will now bound the min-entropy of each of the terms in this summation. Roughly speaking, whenever the label $l_k = g$ (a good projector is applied), λ_k amount of min-entropy will be accumulated but when the projector is bad $O(\log|A|)$ amount of min-entropy will be lost. The reason, we are still able to accumulate a large amount of min-entropy for the state $\sigma_{A_1^n B}$ is because these terms are also weighted with the factor $\delta^{\frac{1}{2}\omega_b(l_1^n)}$. If the number of bad projectors in a term is large, then this factor ensures that the contribution of this term is small.

Claim 5.17. *For every $k \in [n]$ and $l_k \in \{g,b\}$, we have*

$$P^{(k,l_k)}_{A_1^k B} \; \rho_{A_1^k B} \; P^{(k,l_k)}_{A_1^k B} \leq \begin{cases} (1+g_1(\epsilon)) e^{-\lambda_k} \, \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B} & \text{if } l_k = g \\ 4|A|(1+g_1(\epsilon)) \, \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B} & \text{if } l_k = b \end{cases} \tag{5.167}$$

which can succinctly be written as

$$P^{(k,l_k)}_{A_1^k B} \; \rho_{A_1^k B} \; P^{(k,l_k)}_{A_1^k B} \leq (1+g_1(\epsilon)) e^{-\lambda_k \delta(l_k,g)} (4|A|)^{\delta(l_k,b)} \, \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B} \tag{5.168}$$

where $\delta(x,y)$ is the Kronecker delta function ($\delta(x,y) = 1$ if $x = y$ else it is 0).

Proof. Let's first consider the case when $l_k = g$. In this case, we have

$$\begin{aligned} P^{(k,g)}_{A_1^k B} \; \rho_{A_1^k B} \; P^{(k,g)}_{A_1^k B} &\leq (1+g_1(\epsilon)) \tilde{\rho}^{(k)}_{A_1^k B} \\ &\leq (1+g_1(\epsilon)) e^{-\lambda_k} \, \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B} \end{aligned} \tag{5.169}$$

where the first line follows from the definition of the good projectors (Eq. 5.140) and the second line follows from Eq. 5.138.

When $l_k = b$, we have

$$\begin{aligned} P^{(k,b)}_{A_1^k B} \; \rho_{A_1^k B} \; P^{(k,b)}_{A_1^k B} &\leq 2\rho_{A_1^k B} + 2 P^{(k,g)}_{A_1^k B} \; \rho_{A_1^k B} \; P^{(k,g)}_{A_1^k B} \\ &\leq 2\rho_{A_1^k B} + 2(1+g_1(\epsilon)) \tilde{\rho}^{(k)}_{A_1^k B} \\ &\leq 2(1+g_1(\epsilon)) \left(|A| \, \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B} + e^{-\lambda_k} \, \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B} \right) \\ &\leq 4|A|(1+g_1(\epsilon)) \, \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B} \end{aligned} \tag{5.170}$$

where in the first line we have used Lemma 5.12, in the second line we have used Eq. 5.140, in the third line we have used $\rho_{A_1^k B} \leq |A| \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B}$ and Eq. 5.138, and in the last line we use $\lambda_k \geq -\log|A|$. □

For every k, let $c_k(l_k) := e^{-\lambda_k \delta(l_k,g)} (4|A|)^{\delta(l_k,b)}$ so that

$$P^{(k,l_k)}_{A_1^k B} \; \rho_{A_1^k B} \; P^{(k,l_k)}_{A_1^k B} \leq (1 + g_1(\epsilon)) c_k(l_k) \mathbb{1}_{A_k} \otimes \rho_{A_1^{k-1} B}. \tag{5.171}$$

Now, observe that for each term in the summation in Eq. 5.166

$$\begin{aligned}
&P^{(1,l_1)}_{A_1 B} P^{(2,l_2)}_{A_1^2 B} \cdots P^{(n,l_n)}_{A_1^n B} \; \rho_{A_1^n B} \; P^{(n,l_n)}_{A_1^n B} \cdots P^{(2,l_2)}_{A_1^2 B} P^{(1,l_1)}_{A_1 B} \\
&\quad \leq (1 + g_1(\epsilon)) c_n(l_n) P^{(1,l_1)}_{A_1 B} P^{(2,l_2)}_{A_1^2 B} \cdots P^{(n-1,l_{n-1})}_{A_1^{n-1} B} \; \mathbb{1}_{A_n} \otimes \rho_{A_1^{n-1} B} \; P^{(n-1,l_{n-1})}_{A_1^{n-1} B} \cdots P^{(2,l_2)}_{A_1^2 B} P^{(1,l_1)}_{A_1 B} \\
&\quad = (1 + g_1(\epsilon)) c_n(l_n) \mathbb{1}_{A_n} \otimes P^{(1,l_1)}_{A_1 B} P^{(2,l_2)}_{A_1^2 B} \cdots P^{(n-1,l_{n-1})}_{A_1^{n-1} B} \; \rho_{A_1^{n-1} B} \; P^{(n-1,l_{n-1})}_{A_1^{n-1} B} \cdots P^{(2,l_2)}_{A_1^2 B} P^{(1,l_1)}_{A_1 B} \\
&\quad \leq \cdots \\
&\quad \leq (1 + g_1(\epsilon))^n \left(\prod_{k=1}^n c_k(l_k) \right) \mathbb{1}_{A_1^n} \otimes \rho_B.
\end{aligned} \tag{5.172}$$

Plugging this bound into the expression for σ in Eq. 5.166, we get

$$\begin{aligned}
\sigma_{A_1^n B} &= \sum_{l_1^n \in \{g,b\}^n} \delta^{\frac{1}{2}\omega_b(l_1^n)} P^{(1,l_1)}_{A_1 B} P^{(2,l_2)}_{A_1^2 B} \cdots P^{(n,l_n)}_{A_1^n B} \; \rho_{A_1^n B} \; P^{(n,l_n)}_{A_1^n B} \cdots P^{(2,l_2)}_{A_1^2 B} P^{(1,l_1)}_{A_1 B} \\
&\leq (1 + g_1(\epsilon))^n \left(\sum_{l_1^n \in \{g,b\}^n} \delta^{\frac{1}{2}\omega_b(l_1^n)} \prod_{k=1}^n c_k(l_k) \right) \mathbb{1}_{A_1^n} \otimes \rho_B.
\end{aligned} \tag{5.173}$$

Let us now bound the expression

$$\begin{aligned}
\sum_{l_1^n \in \{g,b\}^n} \delta^{\frac{1}{2}\omega_b(l_1^n)} \prod_{k=1}^n c_k(l_k) &= \sum_{l_1^n \in \{g,b\}^n} \prod_{k=1}^n \sqrt{\delta}^{\delta(l_k,b)} \prod_{k=1}^n e^{-\lambda_k \delta(l_k,g)} (4|A|)^{\delta(l_k,b)} \\
&= \sum_{l_1^n \in \{g,b\}^n} \prod_{k=1}^n e^{-\lambda_k \delta(l_k,g)} (4|A|\sqrt{\delta})^{\delta(l_k,b)} \\
&= \prod_{k=1}^n \left(e^{-\lambda_k} + 4|A|\sqrt{\delta} \right) \\
&= \prod_{k=1}^n e^{-\lambda_k} \left(1 + 4|A| e^{\lambda_k} \sqrt{\delta} \right) \\
&\leq \prod_{k=1}^n e^{-\lambda_k} \left(1 + 4|A|^2 \frac{\sqrt{\delta}}{1 - \epsilon^2} \right) \\
&= \left(1 + 4|A|^2 \frac{\sqrt{\delta}}{1 - \epsilon^2} \right)^n e^{-\sum_{k=1}^n \lambda_k}
\end{aligned} \tag{5.174}$$

where we have used $\lambda_k \leq \log \frac{|A|}{1-\epsilon^2}$ in the second last line. Combining Eq. 5.173 and Eq. 5.174, we get

$$\begin{aligned} H_{\min}(A_1^n|B)_\sigma &\geq -D_{\max}(\sigma_{A_1^n B} \| \mathbb{1}_{A_1^n} \otimes \rho_B) \\ &\geq \sum_{k=1}^n \lambda_k - n \log\left(1 + 4|A|^2 \frac{\sqrt{\delta}}{1-\epsilon^2}\right) - n\log(1+g_1(\epsilon)) \\ &\geq \sum_{k=1}^n \bar{H}^\epsilon_{\min}(A_k|A_1^{k-1}B)_\rho - n \log\left(1 + 4|A|^2 \frac{\sqrt{\delta}}{1-\epsilon^2}\right) - n\log(1+g_1(\epsilon)) \end{aligned} \tag{5.175}$$

Finally, we can use the entropic triangle inequality in Lemma 3.6 along with Eq. 5.161 to get

$$\begin{aligned} H^\mu_{\min}(A_1^n|B)_\rho \geq \sum_{k=1}^n \bar{H}^\epsilon_{\min}(A_k|A_1^{k-1}B)_\rho &- n \log\left(1 + 4|A|^2 \frac{\sqrt{\delta}}{1-\epsilon^2}\right) - n\log(1+g_1(\epsilon)) \\ &- n\log(1+\sqrt{\delta}) - n(\mu+\mu^3) - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \end{aligned} \tag{5.176}$$

where $\mu = \left(g_2(\epsilon)\log\frac{1}{\delta}\right)^{1/3}$. We choose the parameter $\delta = \epsilon^2$, so that we have

$$\begin{aligned} H^\mu_{\min}(A_1^n|B)_\rho \geq \sum_{k=1}^n \bar{H}^\epsilon_{\min}(A_k|A_1^{k-1}B)_\rho &- n \log\left(1 + 4|A|^2 \frac{\epsilon}{1-\epsilon^2}\right) - n\log(1+g_1(\epsilon)) \\ &- n\log(1+\epsilon) - n(\mu+\mu^3) - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} \end{aligned} \tag{5.177}$$

for $\mu = \left(8(1+\epsilon^{1/3})\log\frac{1}{\epsilon}\right)^{1/3}\epsilon^{1/9} = O(\epsilon^{1/9}\left(\log\frac{1}{\epsilon}\right)^{1/3})$. □

We complete the proof of the universal chain rule by transforming all the entropies to $H^{\downarrow,\epsilon}_{\min}$ in the following theorem.

Theorem 5.18. *For a state $\rho_{A_1^n B}$ and $\epsilon \in (0,1)$, we have the chain rule*

$$\begin{aligned} H^{\downarrow,2\mu+\epsilon/4}_{\min}(A_1^n|B)_\rho \geq \sum_{k=1}^n H^{\downarrow,\epsilon/4}_{\min}(A_k|A_1^{k-1}B)_\rho &- n \log\left(1 + 4|A|^2 \frac{\epsilon}{1-\epsilon^2}\right) - n\log(1+g_1(\epsilon)) \\ &- n\log(1+\epsilon) - n(\mu+\mu^3) - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} - \log\left(\frac{2}{\mu^2} + \frac{1}{1-\mu}\right) \end{aligned} \tag{5.178}$$

where

$$\mu := \left(8(1+\epsilon^{1/3})\log\frac{1}{\epsilon}\right)^{1/3}\epsilon^{1/9} = O\left(\epsilon^{1/9}\left(\log\frac{1}{\epsilon}\right)^{1/3}\right) \tag{5.179}$$

$$g_1(\epsilon) := \frac{8}{3}(1+\epsilon^{1/3})\epsilon^{1/3}\log\frac{1}{\epsilon} + (1+\epsilon^{1/3}+\epsilon^{2/3})\epsilon^{1/3} = O\left(\epsilon^{1/3}\log\frac{1}{\epsilon}\right) \tag{5.180}$$

as long as $\mu \in (0,1)$.

Proof. **Case 1:** If $\rho_{A_1^n B}$ is full rank, then we can use the lemma above along with Eq. 2.42 and Lemma 5.11 to show that

$$H_{\min}^{\downarrow,2\mu}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H_{\min}^{\downarrow,\epsilon/2}(A_k|A_1^{k-1}B)_\rho - n\log\left(1+4|A|^2\frac{\epsilon}{1-\epsilon^2}\right) - n\log(1+g_1(\epsilon)) \\ - n\log(1+\epsilon) - n(\mu+\mu^3) - \frac{1}{\mu^2} - \log\frac{1}{1-\mu^2} - \log\left(\frac{2}{\mu^2}+\frac{1}{1-\mu}\right). \tag{5.181}$$

Case 2: Now, we can follow the same argument as the one in Case 2 of the proof of Theorem 5.7 to derive the bound in the theorem statement for all states ρ. □

5.5. Conclusion

We developed a powerful proof technique combining the entropic triangle inequality, the generalised Golden-Thompson inequality and the substate theorem for proving entropic bounds for approximation chains. We used this technique to prove novel chain rules– the universal smooth min-entropy chain rule and the unstructured approximate entropy accumulation theorem. Importantly, both of these chain rules can be used meaningfully for arbitrarily large number of systems.

As far as applications are concerned, we use the unstructured approximate EAT to prove the security of parallel DIQKD in the next chapter. We expect the universal chain rule to aid in transforming von Neumann entropy based arguments to one-shot arguments. Furthermore, it provides a straightforward solution to problems such as the approximately independent registers problem discussed in Sec. 3.3. It is also our conviction that the proof technique introduced in this chapter will be useful for tackling other problems. For instance, it should be possible to use it to derive similar chain rules for one-shot variants (e.g., $I_{\max}$) of the (multipartite) mutual information.

As discussed in Sec. 5.3.2, it seems possible to decouple the smoothing parameter and the approximation parameter in certain scenarios, especially those involving DIQKD. We leave the problem of determining whether these parameters can be decoupled using testing for future work. This is an interesting and important question, which could potentially lead to significant improvements in the security proof of parallel DIQKD and the analysis of DIQKD with leakage.

Finally, we did not attempt to optimise our bounds here, but it should be interesting to study the absolute limits of the entropic error terms and the smoothing errors in our chain rules.

Chapter 6

Security for parallel DIQKD

6.1. Introduction

We introduced device-independent quantum key distribution (DIQKD) in Sec. 2.8. These protocols involve multiple rounds of a non-local game (Definition 2.32), which can be played either sequentially or in parallel, depending on which the protocols themselves are termed sequential or parallel. Sequential protocols are comparatively easier to analyse as they can be broken down into smaller steps, each depending only on the preceding steps. Each of these steps is itself simply an instance of the non-local game with some state shared between the Alice and Bob. In contrast, during parallel DIQKD, the two parties input all the questions for the multiple games into their devices and receive all the answers at once. This simultaneous nature makes the analysis of these protocols significantly more challenging, as there really is just one quantum channel depending on all the questions which produces all the answers. There is no natural way to decompose it further. From the viewpoint of implementations, however, parallel DIQKD protocols could potentially lead to faster implementations. Moreover, from a foundational perspective, these protocols remove the sequential or time-ordering assumption in secure key distribution. Thus, investigating the security of these protocols is an important and interesting question.

The challenge of breaking down parallel protocols into smaller, more manageable steps for analysis is also encountered when studying the parallel repetition of non-local games. A parallel repetition of a non-local game G, denoted G^n, consists of n instances of the game G played simultaneously. Formally, it is defined as:

Definition 6.1 (Parallel repetition of a non-local game)**.** *The n-parallel repetition of a non-local game, $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \Pi_{XY}, V)$, is the non-local game $G^n = (\mathcal{X}^n, \mathcal{Y}^n, \mathcal{A}^n, \mathcal{B}^n, \Pi_{XY}^{\otimes n}, V^n)$, where Alice and Bob are given the questions $x_1^n \in \mathcal{X}^n$ and $y_1^n \in \mathcal{Y}^n$ respectively sampled*

according to the distribution $\Pi_{XY}^{\otimes n}$ *(independently and identically according to* Π_{XY}*). Alice and Bob reply with answers* $a_1^n \in \mathcal{A}^n$ *and* $b_1^n \in \mathcal{B}^n$ *according to their (classical or quantum) strategy. They win the game if for every* $1 \leq i \leq n$*, they satisfy the predicate* $V(x_i, y_i, a_i, b_i)$*.*

For the parallelly repeated game G^n, one can always play the optimal strategy for game G independently n times. This yields a lower bound on the winning probability (Definition 2.32): $\omega_S(G)^n \leq \omega_S(G^n)$ (where strategy S can be either classical or quantum). However, it's natural to ask whether it's possible to significantly outperform this strategy for G^n. While it may not be immediately apparent, there are examples of games where $\omega_S(G)^n < \omega_S(G^n)$. For instance, the FFL game [**Hol09**, Appendix A] exhibits $\omega_c(G^2) = \omega_c(G) < 1$ and $\omega_q(G^2) = \omega_q(G) < 1$. Roughly speaking, this improvement arises because players can correlate their answers across parallel games, thereby correlating the winning conditions and achieving a higher overall winning probability than independent play would allow. The parallel repetition question asks whether the winning probability of G^n decays exponentially in n. For classical strategies, this exponential decay was proven in [**Raz98**, **Hol09**]. In the quantum case, it has been demonstrated for large classes of games [**CSUU08**, **JPY14**, **CS14**, **CWY15**, **DSV14**, **BVY21**], but remains an open question for general games. Currently, for general quantum games, the best known bound for $\omega_q(G^n)$ decays only polynomially in n [**Yue16**].

The fundamental idea behind these works on parallel repetition is that one can simulate the probability distributions and states created by the strategy for the parallelly repeated game G^n, conditioned on an event Ω defined in terms of a small subset ($\leq \delta n$ for small $\delta > 0$) of the questions and answers, using a single-round strategy for the game G. In this single-round simulation strategy, the game G is actually embedded in a particular round j of the game G^n. Specifically, it is shown that the distribution of questions at index j, conditioned on the event Ω, has the same distribution as the questions of the game G. Furthermore, it is demonstrated that one can define appropriate measurements which produce answers A_j and B_j for Alice and Bob, respectively, with the same distribution as they would have in G^n conditioned on Ω. Using this simulation, it can then be shown that the winning probability for round j cannot be significantly larger than that of winning the single-round game G.

In this work, we apply techniques developed for the analysis of the parallel repetition of anchored games [**BVY21**] to create a proof for parallel DIQKD which uses CHSH games. Specifically, we demonstrate that for a random subset of size δn of the games, Alice's answer for every game in the subset can be approximately viewed as the output of a single-round strategy, similar to the parallel repetition setting. In the language of approximation chains,

we show that the state produced at the end of the protocol, $\rho_{A_1^n B_1^n X_1^n Y_1^n E}$, where X_1^n, Y_1^n are the questions, A_1^n, B_1^n are the answers for the non-local games, and E is the adversary's register, has an approximation chain $\left(\sigma^{(k)}_{A_1^k B_1^k X_1^k Y_1^k E}\right)_{k=1}^{n}$ satisfying:

$$\sigma^{(k)}_{A_1^k B_1^k X_1^k Y_1^k E} = \mathcal{M}_k^{R_k \to X_k Y_k A_k B_k}\left(\sigma^{(k,0)}_{A_1^{k-1} B_1^{k-1} X_1^{k-1} Y_1^{k-1} R_k E}\right) \tag{6.1}$$

where $\mathcal{M}_k$ is the channel applied by participants playing a single-round of a variant of the CHSH game. With this result, we can leverage the fact that Alice's answers for a single-round CHSH game are random with respect to the adversary when the CHSH game is won with high probability. We first present this idea with a proof sketch for security based on von Neumann entropies in Section 6.6. This allows us to concretely demonstrate how parallel repetition techniques can be used for proving security. However, the von Neumann entropy based bounds, we derive here, are not strong enough to prove security. In the subsequent section, Section 6.7, we use the unstructured approximate EAT to port the bound to a smooth min-entropy bound.

6.1.1. Comparison with previous work

[JMS20] provided the first proof for parallel DI-QKD. Their QKD protocol is based on the Magic Square game. The security proof for this protocol relies on using the parallel repetition theorem for free games (games where questions have a product distribution) with multiple (more than 2) players. [JMS20] views the setting of DIQKD with Alice, Bob, and Eve as the parallel repetition of a multiplayer game. In this context, $\exp\left(-H^{\epsilon}_{\min}(\text{Raw Key}|\text{Eve's information})\right)$, where the raw key consists of Alice's answers on a random subset, can be interpreted as the winning probability for this game. Since the winning probability decays exponentially in the number of rounds due to the parallel repetition result, the smooth min-entropy can be bounded by $\Omega(n)$. This proof relies on three key properties of the Magic Square game: (1) it samples questions uniformly, (2) there exists a quantum strategy to win it perfectly, and (3) under this perfect strategy, Alice and Bob receive a perfectly correlated uniform random bit.

[Vid17] significantly simplified the security proof given by [JMS20]. The key idea remains viewing the QKD protocol as a parallel repetition of a 3-player game between Alice, Bob and Eve. This 3-player Magic Square game is won if Alice and Bob win the Magic Square game and if Eve correctly guesses Alice's answer. A technique developed for studying non-local games called "immunization" [KKM+08] is used by [Vid17] to prove that the 3-player game has a winning probability strictly less than 1. The parallel repetition theorem for anchored games [BVY21] is then used to show that the winning probability of the

repeated game is $2^{-\Omega(n)}$. This is subsequently used to bound Eve's guessing probability of Alice's answers and hence the min-entropy for Alice's answers given Eve's system. [Vid17]'s proof also utilises the same properties of the Magic Square game as [JMS20]. Building on these works, [JK22] also proves the security of a similar parallel DIQKD protocol in the presence leakage from Alice and Bob's devices.

In comparison to these proofs, our proof is not limited by the properties of the games used for the DIQKD protocol. We demonstrate our protocol using the CHSH game, showing how to convert it into an anchored game suitable for parallel DIQKD– a technique we believe should be applicable to other games as well. Further, our work offers an alternative security proof for parallel DIQKD, employing a more information-theoretic approach. This method decomposes the large quantum device playing parallel CHSH games into smaller single-round CHSH game playing devices. In contrast, [JMS20] and [Vid17] reduce the security proof to bounding the winning probability of a parallelly repeated game. We hope that our approach can provide greater insight into the problem and aid in solving open problems like the security of parallel device-independent randomness expansion. From another perspective, techniques from parallel repetition lie at the heart of both our proof and those of [JMS20] and [Vid17], highlighting the fundamental importance of these methods in analysing parallel protocols.

On the downside, our proof strategy couples the security parameter of the DIQKD protocol to its rate. For a choice of security parameter of $\tilde{O}(\epsilon)$, our approach can only prove security for a rate of $\Omega(\epsilon^{192})$. This is not a limitation of the proofs in [JMS20] and [Vid17]. It might be possible to break this linkage by further refining the unstructured approximate EAT as mentioned in the previous chapter, but we leave this for future work. Finally, it is important to note that our protocol and the protocols in [JMS20] and [Vid17] are intended as proofs of concept. The key rates for these parallel DIQKD protocols are currently too small for practical implementation.

6.2. Preliminaries

In this chapter, we follow [BVY21] in using the notation $[\![x]\!] = |x\rangle\langle x|$ to represent a classical value x. We also denote the density operator for a pure state $|\psi\rangle$ as ψ.

Recall that a non-local game G is represented as $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \Pi_{XY}, V)$, where $\mathcal{X}$ and $\mathcal{Y}$ are the sets of Alice and Bob's questions, $\mathcal{A}$ and $\mathcal{B}$ are the sets of their answers, Π_{XY} is

the probability distribution of their questions and V is the winning predicate.

For the standard CHSH game, we have $\mathcal{X} = \mathcal{Y} = \mathcal{A} = \mathcal{B} = \{0,1\}$, Π_{XY} is the uniform distribution on all possible questions and the predicate $V(x,y,a,b) = \neg[a \oplus b \oplus (x \wedge y)]$. We refer to this game as the CHSH game or the 2CHSH game.

This nomenclature helps distinguish it from the *3CHSH game*, which is usually used for DIQKD (see 2.8.2). In this game, Alice's questions lie in $\{0,1\}$ and Bob's question lie in $\{0,1,2\}$. These questions are sampled according to the probability distribution

$$P_{XY}(x,y) = \begin{cases} 1-\nu & \text{if } x = 0, y = 2 \\ \nu/4 & \text{if } x, y \in \{0,1\} \end{cases}$$

where $\nu \in (0,1)$ is a parameter which we will fix later. For the questions $x,y \in \{0,1\}$, Alice and Bob win this game if they win the standard CHSH game, that is, if $\neg[a \oplus b \oplus (x \wedge y)]$ is true. On questions $(x,y) = (0,2)$, they win if their answers are equal.

We require the following *anchoring transform* for non-local games in order to describe our protocol.

Definition 6.2 (Anchoring transform [BVY21])**.** *Let $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \Pi_{XY}, V)$ be a non-local game and $0 < \alpha \leq 1$. In the α-anchored game $G_\perp := (\mathcal{X} \cup \{\perp\}, \mathcal{Y} \cup \{\perp\}, \mathcal{A}, \mathcal{B}, \Pi^\perp_{XY}, V_\perp)$, the Referee first uses Π_{XY} to sample questions x,y for Alice and Bob. Then, randomly and independently with probability α, he replaces each of x and y with an auxiliary "anchor" symbol $\perp$ to obtain the questions for the anchored game $G_\perp$. Alice and Bob win the game if either one of the questions was $\perp$, or if their answers a,b satisfy the original game's predicate, that is, $V(x,y,a,b) = 1$. The quantum winning probability for the anchored game satisfies*

$$\omega_q(G_\perp) = 1 - (1-\alpha)^2(1-\omega_q(G)).$$

We call the game obtained after applying the anchoring transform for $\alpha \in (0,1)$ to the 3CHSH game, the $3\text{CHSH}_\perp$ game. Once again, we will consider α to be a parameter and fix it later.

A strategy S for a non-local game $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \Pi_{XY}, V)$ consists of the tuple $(\Psi_{E_A E_B}, \{A_x\}_{x \in \mathcal{X}}, \{B_y\}_{y \in \mathcal{Y}})$ where $\Psi_{E_A E_B}$ is the quantum state shared by Alice and Bob, A_x is the measurement used by Alice on question x, and B_y is the measurement used by Bob

on question y.

If a quantum strategy wins the 2CHSH game with a probability strictly greater than 3/4, then Alice's answer is guaranteed to be random with respect to any purification of the initial state held by the adversary, Eve, and the questions for the game. This statement is quantified in the entropic bound given in Lemma 2.33. We state a counterpart for this lemma for the 3CHSH$_\perp$ game. It follows fairly easily from Lemma 2.33. We prove it in Appendix D.1.

Lemma 6.3. *Suppose that a given quantum strategy for the 3CHSH$_\perp$ game starting with $\rho^{(0)}_{E_A E_B E}$ wins the 3CHSH$_\perp$ game with probability $\omega \in \left[1 - \frac{(1-\alpha)^2\nu}{4}, 1 - \frac{2-\sqrt{2}}{4}(1-\alpha)^2\nu\right]$. Let X and Y be Alice and Bob's questions during the game, and A and B be their answers produced according to this strategy. Then, for the post measurement state ρ_{XYABE}, we have*

$$H(AB|EXY)_\rho \geq H(A|EX)_\rho \geq (1-\alpha)F(g_{\alpha,\nu}(\omega)) \tag{6.2}$$

where the functions F and $g_{\alpha,\nu}(\omega)$ are given by

$$F(x) = \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{3 - 16\,x\,(1-x)}\right) \quad \text{for} \quad g_{\alpha,\nu}(\omega) = 1 - \frac{1-\omega}{\nu(1-\alpha)^2}. \tag{6.3}$$

6.3. Protocol

Before we describe our protocol for parallel DIQKD, we must introduce a key result from [BVY21]. For every α-anchored game $G_\perp := (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, P_{XY}, V)$ (a game produced by applying the anchoring transform), [BVY21, Section 4.1] shows that the probability distribution P_{XY} can be extended to the distribution $\hat{P}_{\Omega XY}$ such that

$$\hat{P}_{XY} := P_{XY} \tag{6.4}$$

$$\hat{P}_{\Omega XY} = \hat{P}_\Omega \hat{P}_{X|\Omega} \hat{P}_{Y|\Omega}. \tag{6.5}$$

This extension plays a crucial role in our protocol, as we will demonstrate in the subsequent sections. In this work, we will call the random variable Ω the *seed randomness* for the questions. We note that Ω is defined such that it can be sampled efficiently by Alice. Given Ω, Alice and Bob can independently sample their questions for the game, $G_\perp$.

For the rest of the chapter, let the tuple $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, P_{XY}, V)$ represent the 3CHSH$_\perp$ game. We will drop the hat notation while referring to the extension distribution for this question distribution. We simply refer to it as $P_{\Omega XY}$.

Parameters:

- $\alpha, \nu \in (0, 0.1)$ are parameters for the 3CHSH$_\perp$
- $\delta \in \left(0, \frac{1}{2}\right)$ determines the size of the raw key
- $\omega_{\text{th}} \in \left(1 - \frac{(1-\alpha)^2\nu}{4}, 1 - \frac{2-\sqrt{2}}{4}(1-\alpha)^2\nu\right)$ is the threshold for the winning probability on the test round
- $\gamma \in (0,1)$ parameter for sampling testing rounds.

Parallel DIQKD protocol

(1) Alice randomly samples Ω_1^n independently and identically.

(2) Alice sends Ω_1^n to Bob.

(3) Alice and Bob use Ω_1^n to sample the questions X_1^n and Y_1^n for the 3CHSH$_\perp$ game.

(4) Alice and Bob use their questions to play n 3CHSH$_\perp$ games. Let A_1^n and B_1^n be the answers.

(5) Alice randomly selects a subset $J = \{I_1, I_2, \cdots, I_t\}$ of size $t = \frac{\delta}{\log|\mathcal{A}||\mathcal{B}|+\delta} n$. She announces this subset to Bob.

(6) For each $i \in [t]$, Alice randomly selects a $T_j \in \{0,1\}$ with probability $\Pr(T_j = 1) = \gamma$. Let $S := \{I_j : j \in [t] \text{ and } T_j = 1\} \subseteq J$. She announces S, her questions X_S, and answers A_S for this subset of the games.

(7) Bob checks whether $\sum_{i \in S} V(X_i, Y_i, A_i, B_i) \geq \gamma\omega_{\text{th}} t$. Alice and Bob abort if this is not satisifed.

(8) A_J and B_J are Alice and Bob's raw keys. They use information reconciliation and privacy amplification to create a secret key.

Protocol 6.1

We present the protocol for parallel DIQKD in Protocol 6.1. Note that this protocol does not fully reveal the questions to Eve. Unlike sequential DI-QKD, where Alice and Bob can reveal all of their questions to Eve, the security proofs for parallel DI-QKD require that Alice and Bob only reveal a small fraction of their questions to Eve[(1)]. Similarly, in our setting, since only Ω_1^n are revealed to Eve, only a fraction of the information about the questions is leaked to Eve. Importantly, this is the main obstacle to extending these techniques to prove security for parallel device-independent randomness extraction. It should be noted that the previous protocols and proofs [JMS20, Vid17] required the probability distributions of Alice and Bob's questions to be a product distribution, so that both Alice and Bob could sample their questions independently. However, by utilising the anchoring transform and the seed randomness we are able to relax this constraint.

[(1)] In the parallel DIQKD protocol of [JMS20] only a small fraction of the questions are announced publicly. In the protocol used by [Vid17], Alice and Bob can reveal any fraction smaller than 1.

6.4. Setup

As indicated in Protocol 6.1 we let Ω_1^n be the randomness seed for the questions shared by Alice. X_1^n and Y_1^n denote Alice and Bob's questions for the n 3CHSH$_\perp$ games during the protocol, and A_1^n and B_1^n denote their answers for these games.

In order to analyse the protocol, we fix a strategy for Eve. Let's suppose that Eve distributes the registers E_A and E_B of the pure state $\psi_{E_A E_B E}$ between Alice and Bob, keeping register E for herself[(2)]. Let Alice and Bob use measurements $\{A_{x_1^n}^{E_A}(a_1^n)\}_{a_1^n}$ and $\{B_{y_1^n}^{E_B}(b_1^n)\}_{b_1^n}$ to measure their registers E_A and E_B respectively given questions x_1^n and y_1^n.

The state after all the 3CHSH$_\perp$ games have been played will be denoted as ρ. We have that

$$\rho_{\Omega_1^n X_1^n Y_1^n A_1^n B_1^n E} = \sum_{\omega_1^n, x_1^n, y_1^n} P_{\Omega_1^n X_1^n Y_1^n}(\omega_1^n, x_1^n, y_1^n) [\![\omega_1^n, x_1^n, y_1^n]\!] \otimes \sum_{a_1^n, b_1^n} [\![a_1^n, b_1^n]\!] \otimes \mathrm{tr}_{E_A E_B}\left(A_{x_1^n}^{E_A}(a_1^n) \otimes B_{y_1^n}^{E_B}(b_1^n) \psi_{E_A E_B E}\right) \tag{6.6}$$

where $P_{\Omega_1^n X_1^n Y_1^n}$ is the i.i.d. distribution $P_{\Omega XY}^{\otimes n}$. We can also write the above as

$$\rho_{\Omega_1^n X_1^n Y_1^n A_1^n B_1^n E} = \sum_{\omega_1^n, x_1^n, y_1^n} P_{\Omega_1^n X_1^n Y_1^n}(\omega_1^n, x_1^n, y_1^n) [\![\omega_1^n, x_1^n, y_1^n]\!] \otimes \sum_{a_1^n, b_1^n} P_{A_1^n B_1^n | X_1^n Y_1^n}(a_1^n, b_1^n | x_1^n, y_1^n) [\![a_1^n, b_1^n]\!] \otimes \rho_E^{(x_1^n, y_1^n, a_1^n, b_1^n)} \tag{6.7}$$

for $P_{A_1^n B_1^n | X_1^n Y_1^n}(a_1^n, b_1^n | x_1^n, y_1^n) = \mathrm{tr}\left(A_{x_1^n}^{E_A}(a_1^n) \otimes B_{y_1^n}^{E_B}(b_1^n) \psi_{E_A E_B E}\right)$ and

$$\rho_E^{(x_1^n, y_1^n, a_1^n, b_1^n)} := \frac{\mathrm{tr}_{E_A E_B}\left(A_{x_1^n}^{E_A}(a_1^n) \otimes B_{y_1^n}^{E_B}(b_1^n) \psi_{E_A E_B E}\right)}{P_{A_1^n B_1^n | X_1^n Y_1^n}(a_1^n, b_1^n | x_1^n, y_1^n)}. \tag{6.8}$$

We will also use the notation $P_{\Omega_1^n X_1^n Y_1^n A_1^n B_1^n} := \rho_{\Omega_1^n X_1^n Y_1^n A_1^n B_1^n}$.

Using ρ in the form in Eq. 6.7, we can further define the state of register E conditioned on other classical variables, for example register E conditioned on ω_1^n is

$$\rho_E^{(\omega_1^n)} = \sum_{x_1^n, y_1^n, a_1^n, b_1^n} P_{X_1^n Y_1^n A_1^n B_1^n | \Omega_1^n}(x_1^n, y_1^n, a_1^n, b_1^n | \omega_1^n) \rho_E^{(x_1^n, y_1^n, a_1^n, b_1^n)} \tag{6.9}$$

$$= \mathop{\mathbb{E}}_{x_1^n y_1^n a_1^n b_1^n | \omega_1^n} \left[\rho_E^{(x_1^n, y_1^n, a_1^n, b_1^n)}\right]. \tag{6.10}$$

Finally, we let F (for "fail") denote the event that the protocol aborts.

[(2)] The state can be considered pure. If it were not, then purifying it and providing Eve the purification register would only increase Eve's information.

6.5. Key results from [BVY21]

In Protocol 6.1, we have chosen

$$t = \frac{\delta}{\log|\mathcal{A}||\mathcal{B}| + \delta} n. \tag{6.11}$$

In our security analysis, we will fix a subset $C \subseteq J$ and show that for any such set we can embed a single-round 3CHSH$_\perp$ game in a random index outside this set. The value of t above guarantees that for any such subset $C \subseteq J$, we have

$$\frac{|C|}{n - |C|} \log|\mathcal{A}||\mathcal{B}| \leq \delta. \tag{6.12}$$

The results established in [BVY21] for parallel repetition settings remain applicable in our context, with minor modifications (see Appendix D.2 for a detailed discussion). Specifically, these results can be adapted to our setting by introducing a reference register E. In this section, we present the key definitions and results from [BVY21] which are used in our security proof.

Following [BVY21, Section 4.2], for the subset $C \subseteq [n]$ and $i \in C^c = [n] \setminus C$, we define the *dependency breaking variable* R_{-i} as

$$R_{-i} := (\Omega_j)_{j\in[n]\setminus(C\cup\{i\})} \cup (X_C, Y_C, A_C, B_C). \tag{6.13}$$

We also use the shorthand Ω_{-i} for $(\Omega_j)_{j\in[n]\setminus(C\cup\{i\})} \cup (X_C, Y_C)$.

For questions x_1^n and y_1^n, we define the following measurements for answers on the subset $C \subseteq [n]$:

$$A^{E_A}_{x_1^n}(a_C) = \sum_{a_1^n|a_C} A^{E_A}_{x_1^n}(a_1^n) \tag{6.14}$$

$$B^{E_B}_{y_1^n}(b_C) = \sum_{b_1^n|b_C} B^{E_B}_{y_1^n}(b_1^n). \tag{6.15}$$

Here $a_1^n|a_C$ denotes strings a_1^n consistent with a_C. $b_1^n|b_C$ is defined similarly.

For subset $C \subseteq [n]$, $i \in C^c$, the seed $\omega_{-i} = (\omega_j)_{j\in[n]\setminus(C\cup\{i\})} \cup (x_C, y_C)$ (an instantiation of Ω_{-i}), and questions x and y we also define the measurements:

$$A^{E_A}_{\omega_{-i},x}(a_1^n) := \mathop{\mathbb{E}}_{X_1^n|\Omega_{-i}=\omega_{-i},X_i=x} A^{E_A}_{x_1^n}(a_1^n) \tag{6.16}$$

$$= \sum_{x_1^n} P_{X_1^n|\Omega_{-i}X_i}(x_1^n|\omega_{-i}, x) A^{E_A}_{x_1^n}(a_1^n) \tag{6.17}$$

and

$$B^{E_B}_{\omega_{-i},y}(b_1^n) := \mathop{\mathbb{E}}_{Y_1^n|\Omega_{-i}=\omega_{-i},Y_i=y} B^{E_B}_{y_1^n}(b_1^n) \tag{6.18}$$

$$= \sum_{y_1^n} P_{Y_1^n|\Omega_{-i}Y_i}(y_1^n|\omega_{-i},y) B^{E_B}_{y_1^n}(b_1^n). \tag{6.19}$$

Also define

$$A^{E_A}_{\omega_{-i},x}(a_C) := \sum_{a_1^n|a_C} A^{E_A}_{\omega_{-i},x}(a_1^n) \tag{6.20}$$

$$B^{E_B}_{\omega_{-i},y}(b_C) := \sum_{b_1^n|b_C} B^{E_B}_{\omega_{-i},y}(b_1^n). \tag{6.21}$$

For subset $C \subseteq [n]$, $i \in C^c$, $r_{-i} = (\omega_{-i}, a_C, b_C)$ (an instance of R_{-i}) and questions x and y, we also define the unnormalised state

$$|\Phi^{(r_{-i},x,y)}\rangle_{E_AE_BE} := \sqrt{A^{E_A}_{\omega_{-i},x}(a_C)} \otimes \sqrt{B^{E_B}_{\omega_{-i},y}(b_C)}\,|\psi\rangle_{E_AE_BE} \tag{6.22}$$

and its normalisation

$$|\tilde{\Phi}^{(r_{-i},x,y)}\rangle_{E_AE_BE} := \frac{1}{\||\Phi^{(r_{-i},x,y)}\rangle\|}\,|\Phi^{(r_{-i},x,y)}\rangle_{E_AE_BE}\,. \tag{6.23}$$

It is shown in [BVY21, Proposition 4.9] that

$$\||\Phi^{(r_{-i},x,y)}\rangle\| = \left(P_{A_CB_C|\Omega_{-i}X_iY_i}(a_c,b_c|\omega_{-i},x,y)\right)^{1/2}. \tag{6.24}$$

The following result from [BVY21] is the key towards showing that it is possible to simulate the answers produced by the parallelly repeated strategy at a random index outside the set C, using a strategy for the single-round 3CHSH$_\perp$ game which embeds the game at this index.

Proposition 6.4 ([BVY21, Proposition 5.1]). *For every $C \subseteq J$, $i \in C^c$, dependency breaking variable r_{-i}, and questions x and y, there exist unitaries $U^{E_A}_{r_{-i},x}$ acting on E_A and $V^{E_B}_{r_{-i},y}$ acting on E_B such that*

$$\mathop{\mathbb{E}}_{I}\mathop{\mathbb{E}}_{R_{-i}}\mathop{\mathbb{E}}_{XY}\left\|U^{E_A}_{r_{-i},x} \otimes V^{E_B}_{r_{-i},y} \otimes \mathbb{1}^E\,|\tilde{\Phi}^{(r_{-i},\perp,\perp)}\rangle_{E_AE_BE} - |\tilde{\Phi}^{(r_{-i},x,y)}\rangle_{E_AE_BE}\right\| = O(\delta^{1/16}/\alpha^3), \tag{6.25}$$

where $\mathbb{E}_I$ denotes expectation over index I which is sampled uniformly at random from C^c, $\mathbb{E}_{R_{-i}}$ denotes expectation over r_{-i} sampled from $P_{R_{-i}}$ and $\mathbb{E}_{XY}$ denotes expectation over the questions sampled according to the question distribution for the single-round game P_{XY}.

6.6. Proof approach with von Neumann entropies

We will first demonstrate our proof strategy by proving the requisite entropic statements with von Neumann entropies instead of one-shot entropies. This approach allows us to separate the inherent complexity of one-shot entropies from the core concepts of the proof. We can focus on using the results from the parallel repetition setting for analysing DIQKD and establishing a clear roadmap for the security proof, instead of grappling with the various technicalities associated with one-shot entropies. In the next section, we will proceed to develop a comprehensive one-shot security proof.

For simplicity, we also set aside the testing procedure and the conditioning of the state on the protocol not aborting. Instead, we simply assume that the state $\psi_{E_A E_B E}$ and the measurements set up by Eve are such that

$$\Pr_{\rho}\left[\frac{1}{n}\sum_{i\in[n]} W_i \geq \omega_{\text{th}}\right] \geq 1-\epsilon \tag{6.26}$$

where ϵ is negligibly small. This assumption is effectively equivalent to stating that the protocol only aborts with a negligible probability. Once again, this restriction allows us to concentrate on the more challenging aspects of the security proof. In the next section, we will see that once we properly account for the testing procedure, this assumption can be eliminated.

In Sec. 2.7.2, we discussed the entropic bounds required for proving the security of QKD. The same bounds are also sufficient for proving the security of parallel DIQKD. Specifically, in order to prove that the protocol can securely produce a key of length $\Omega(n)$ (that is, it has a positive key rate), we need to show that the difference

$$H^{\epsilon}_{\min}(A_J|E\Omega_1^n S J A_S X_S)_{\rho_{|\neg F}} - H^{\epsilon'}_{\max}(A_J|B_J J)_{\rho_{\text{honest}}} \geq \Omega(n) \tag{6.27}$$

for small ϵ and ϵ'. ρ_{honest} above is the state produced at the end of a protocol with no adversary, Eve.

We use the conditional entropy $H(A_J|E\Omega_1^n J)_\rho$ as proxy for the entropy $H^{\epsilon}_{\min}(A_J|E\Omega_1^n S J A_S X_S)_{\rho_{|\neg F}}$, which quantifies the amount of randomness that can be safely extracted from the Alice's raw key using privacy amplification. Similarly, the entropy $H(A_J|B_J J)_{\rho_{\text{honest}}}$ serves as a proxy for the information reconciliation cost, which is given by $H^{\epsilon'}_{\max}(A_J|B_J J)_{\rho_{\text{honest}}}$. In our von Neumann security proof, we aim to demonstrate that

$$H(A_J|E\Omega_1^n J)_\rho - H(B_J|A_J J)_{\rho_{\text{honest}}} \geq \Omega(n). \tag{6.28}$$

6.6.1. Bounding entropy production for privacy amplification

We begin by showing that Eve's uncertainty of Alice's answers A_J measured using von Neumann entropy is high, that is,

$$H(A_J|E\Omega_1^n J)_\rho \geq \Omega(t). \tag{6.29}$$

This entropy can be expanded using the chain rule as

$$H(A_J|E\Omega_1^n J)_\rho = \sum_{k=1}^{t} H(A_{I_k}|E\Omega_1^n A_{I_1}\cdots A_{I_{k-1}} J)_\rho. \tag{6.30}$$

Further, we can write the term inside the summation above as the expectation

$$H(A_{I_k}|E\Omega_1^n A_{I_1}\cdots A_{I_{k-1}} J)_\rho = \mathop{\mathbb{E}}_{i_1^{k-1}} \left[H(A_{I_k}|E\Omega_1^n A_{i_1}\cdots A_{i_{k-1}} I_k)_\rho\right]. \tag{6.31}$$

For the rest of this section, we focus our attention on this term. Let us fix the choice of random variables $I_1^{k-1} = i_1^{k-1}$. We then define the subset $C := \{i_1, i_2, \ldots, i_{k-1}\} \subseteq J$. This notation allows us to maintain clarity and enables us to collectively bound the term inside the expectation for various values of k and i_1^{k-1}.

The term $H(A_{I_k}|E\Omega_1^n A_{i_1}\cdots A_{i_{k-1}} I_k)_\rho$ will be bounded in two steps. In the first step, we lower bound it using $H(A_{I_k}|EI_k R_{-I_k} X_{I_k})_\rho$, where R_{-I_k} is the dependency breaking random variable defined with respect to C in Eq. 6.13. In the second step, we show that it is possible to approximately simulate the state $\rho_{A_{I_k} I_k R_{-I_k} X_{I_k} E}$ using a quantum strategy for a single instance of the game, which has high winning probability. This allows us to use the single-round entropy bound for the 3CHSH$_\perp$ game (Lemma 6.3) to lower bound the entropy of the answer A_{I_k} for the simulated state, and subsequently for ρ as well.

Step 1: Reduction to parallel repetition variables

In the first step, we will show that

$$H(A_{I_k}|E\Omega_1^n A_{i_1}\cdots A_{i_{k-1}} I_j)_\rho \geq H(A_{I_k}|EI_k R_{-I_k} X_{I_k})_\rho \tag{6.32}$$

where $R_{-I_k} = (\Omega_{-I_k}, X_C, Y_C, A_C, B_C)$. This is fairly simply. It only requires one to use Markov chain properties and data processing. Informally speaking, Alice and Bob's quantum devices only ever get to *see* the questions X_1^n, hence, the uncertainty of the answers only decreases if we change some Ω_is with X_is. We formalise this argument in the following lemma.

Lemma 6.5. *Let $\rho^{(0)}_{\Omega_1^2 X_1^2 E_A E} = \rho_{\Omega_1^2 X_1^2} \otimes \rho^{(0)}_{E_A E}$ be a classical-quantum density operator, where Ω_1^2 and X_1^2 are classical registers and E_A and E are quantum registers. Further, suppose that $\Omega_1 X_1$ and $\Omega_2 X_2$ are sampled independently, that is, $\rho_{\Omega_1^2 X_1^2} = \rho_{\Omega_1 X_1}\rho_{\Omega_2 X_2}$. If the register*

E_A is measured according to measurement operators $\{A^{E_A}_{x_1^2}(a_1^2)\}_{a_1^2}$, which depend only on the classical register X_1^2, then $\Omega_1 \leftrightarrow X_1 \leftrightarrow \Omega_2 A_1^2 E$ forms a Markov chain.

Proof. We can write $\rho^{(0)}_{\Omega_1^2 X_1^2 E_A E}$ as

$$\rho^{(0)}_{\Omega_1^2 X_1^2 E_A E} = \sum_{\omega_1^2 x_1^2} \rho(x_1^2)\rho(\omega_1|x_1)\rho(\omega_2|x_2)[\![\omega_1^2 x_1^2]\!] \otimes \rho^{(0)}_{E_A E}.$$

The post-measurement state is

$$\rho_{\Omega_1^2 X_1^2 A_1^2 E} = \sum_{\omega_1^2 x_1^2} \rho(x_1^2)\rho(\omega_1|x_1)\rho(\omega_2|x_2)[\![\omega_1^2 x_1^2]\!] \otimes \sum_{a_1^2}[\![a_1^2]\!] \otimes \mathrm{tr}_{E_A}(A^{E_A}_{x_1^2}(a_1^2)\rho^{(0)}_{E_A E}),$$

and

$$\begin{aligned}
\rho_{\Omega_1^2 X_1 A_1^2 E} &= \sum_{\omega_1^2 x_1^2} \rho(x_1^2)\rho(\omega_1|x_1)\rho(\omega_2|x_2)[\![\omega_1^2 x_1]\!] \otimes \sum_{a_1^2}[\![a_1^2]\!] \otimes \mathrm{tr}_{E_A}(A^{E_A}_{x_1^2}(a_1^2)\rho^{(0)}_{E_A E}) \\
&= \sum_{\omega_1^2 x_1} \rho(x_1)\rho(\omega_1|x_1)\rho(\omega_2)[\![\omega_1^2 x_1]\!] \otimes \sum_{a_1^2}[\![a_1^2]\!] \otimes \mathrm{tr}_{E_A}(\sum_{x_2}\rho(x_2|\omega_2)A^{E_A}_{x_1^2}(a_1^2)\rho^{(0)}_{E_A E}) \\
&= \sum_{\omega_1^2 x_1} \rho(x_1)\rho(\omega_1|x_1)\rho(\omega_2)[\![\omega_1^2 x_1]\!] \otimes \sum_{a_1^2}[\![a_1^2]\!] \otimes \mathrm{tr}_{E_A}(A^{E_A}_{x_1\omega_2}(a_1^2)\rho^{(0)}_{E_A E})
\end{aligned}$$

where we define $A^{E_A}_{x_1\omega_2}(a_1^2) \coloneqq \sum_{x_2}\rho(x_2|\omega_2)A^{E_A}_{x_1^2}(a_1^2)$. This state can also be written as

$$\begin{aligned}
\rho_{X_1\Omega_1\Omega_2 A_1^2 E} = \sum_{x_1}\rho(x_1)[\![x_1]\!] \otimes &\left(\sum_{\omega_1}\rho(\omega_1|x_1)[\![\omega_1]\!]\right) \\
&\otimes \left(\sum_{\omega_2}\rho(\omega_2)[\![\omega_2]\!] \otimes \sum_{a_1^2}[\![a_1^2]\!] \otimes \mathrm{tr}_{E_A}(A^{E_A}_{x_1\omega_2}(a_1^2)\rho^{(0)}_{E_A E})\right).
\end{aligned}$$

This proves that $\Omega_1 \leftrightarrow X_1 \leftrightarrow \Omega_2 A_1^2 E$ form a Markov chain in the state $\rho_{X_1\Omega_1\Omega_2 A_1^2 E}$. □

Note that for the variables in the lemma above, we also have $\Omega_1 \leftrightarrow \Omega_2 X_1 A_1 E \leftrightarrow A_2$ using properties of Markov chains. We will use this fact in the following.

Fix $I_j = i_j$. Define $C' \coloneqq C \cup \{i_j\}$. The state $\rho^{(0)}_{\Omega_{C'}\Omega_{C'^c}X_{C'}X_{C'^c}E_A E} \coloneqq \rho_{\Omega_{C'}\Omega_{C'^c}X_{C'}X_{C'^c}} \otimes \psi_{E_A E}$ satisfies the conditions for Lemma 6.5. Moreover, the register E_A is measured using the measurement $A_{x_1^n}(\cdot)$, which only depends on X_1^n to create A_1^n. Therefore, using the lemma above, we have that the state ρ satisfies

$$\Omega_{C'} \leftrightarrow \Omega_{C'^c}X_{C'}A_C E \leftrightarrow A_{C^c} \tag{6.33}$$

$$\Rightarrow \Omega_{C'} \leftrightarrow \Omega_{C'^c}X_{C'}A_C E \leftrightarrow A_{i_j}. \tag{6.34}$$

Since, this is true for every i_j, we have

$$\Omega_C\Omega_{I_j} \leftrightarrow I_j\Omega_{C'^c}X_C X_{I_j}A_C E \leftrightarrow A_{I_j}. \tag{6.35}$$

This allows us to bound

$$\begin{aligned} H(A_{I_j}|EI_j\Omega_1^n A_C)_\rho &\geq H(A_{I_j}|EI_j\Omega_1^n X_C X_{I_j} A_C)_\rho \\ &= H(A_{I_j}|EI_j\Omega_{C'^c} X_C X_{I_j} A_C)_\rho \\ &\geq H(A_{I_j}|EI_j R_{-I_j} X_{I_j})_\rho. \end{aligned} \tag{6.36}$$

In the second line above, we used the fact that if $A \leftrightarrow B \leftrightarrow C$, then $H(A|BC) = H(A|B)$.

Step 2: Simulating ρ using a simple single-round strategy

We will now show that it is possible to approximate the state $\rho_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E}$ using a quantum strategy for a single instance of the 3CHSH$_\perp$ game, which uses (I_j, R_{-I_j}) as shared classical random variables between Alice and Bob. This step is much more challenging and requires us to use deeper results from [BVY21]. We use these results primarily to show that the random variable R_{-I_j} is not too correlated with the answer A_{I_j} and hence $H(A_{I_j}|EI_j R_{-I_j} X_{I_j})_\rho$ is large.

Consider Proposition 6.4 applied to the subset $C \subseteq J$ defined above. Let $U_{r_{-i},x}^{E_A}$ and $V_{r_{-i},y}^{E_B}$ be the unitaries provided by this proposition. Then, we have that

$$\mathop{\mathbb{E}}_{I} \mathop{\mathbb{E}}_{R_{-i}} \mathop{\mathbb{E}}_{XY} \left\| U_{r_{-i},x}^{E_A} \otimes V_{r_{-i},y}^{E_B} \otimes \mathbb{1}^E \, |\tilde{\Phi}^{(r_{-i},\perp,\perp)}\rangle_{E_A E_B E} - |\tilde{\Phi}^{(r_{-i},x,y)}\rangle_{E_A E_B E} \right\| = O(\delta^{1/16}/\alpha^3), \tag{6.37}$$

where i is sampled uniformly at random from C^c, r_{-i} sampled from $P_{R_{-i}}$ and x,y are sampled from P_{XY}. This relation hints at a plausible simulation strategy: Eve samples and distributes I, R_{-I} and the state $|\Phi^{(r_{-i},\perp,\perp)}\rangle_{E_A E_B E}$ between Alice and Bob. Then, given questions x and y during the single-round game, Alice and Bob apply the unitaries $U_{r_{-i},x}^{E_A}$ and $V_{r_{-i},y}^{E_B}$ to their registers. This allows them to bring their shared state close to the state $|\Phi^{(r_{-i},x,y)}\rangle_{E_A E_B E}$. Finally, Alice and Bob use appropriately defined measurements to sample their answers, simulating a state close to $\rho_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E}$ through a single-round strategy for 3CHSH$_\perp$.

The measurements used by Alice and Bob in this last step are defined as

$$\hat{A}_{r_{-i},x}(a_i) := A_{\omega_{-i},x}(a_c)^{-1/2} \left(\sum_{a_1^n | a_i, a_C} A_{\omega_{-i},x}(a_1^n) \right) A_{\omega_{-i},x}(a_C)^{-1/2} \tag{6.38}$$

$$\hat{B}_{r_{-i},y}(b_i) := B_{\omega_{-i},y}(b_c)^{-1/2} \left(\sum_{b_1^n | b_i, b_C} B_{\omega_{-i},y}(b_1^n) \right) B_{\omega_{-i},y}(b_C)^{-1/2} \tag{6.39}$$

Single-round protocol for simulating $\rho_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E}$:

(1) Eve chooses the random variable I uniformly at random from C^c, the random variable R_{-I} according to $P_{R_{-I}}$ depending on the value of I, and distributes both of them to Alice and Bob.

(2) Eve distributes the registers E_A and E_B of the state $|\tilde{\Phi}^{(r_{-i},\perp,\perp)}\rangle_{E_A E_B E}$ between Alice and Bob.

(3) Alice and Bob play the $3\text{CHSH}_\perp$ game as follows:

  (a) Let (i, r_{-i}) be the classical variables provided to them by Eve in the first step, and let x and y be their questions.

  (b) Alice applies the unitary $U^{E_A}_{r_{-i},x}$ to her register E_A. Similarly, Bob applies $V^{E_B}_{r_{-i},y}$ to his register E_B.

  (c) Alice and Bob measure their registers with $\{\hat{A}_{r_{-i},x}(a)\}_a$ and $\{\hat{B}_{r_{-i},y}(b)\}_b$ to generate their answers for the game.

Box 6.1

where $r_{-i} = (\omega_{-i}, a_C, b_C)$. We state the simulation protocol for the state $\rho_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E}$ in Box 6.1. Let the state obtained through this simulation procedure be σ.

The proof that the state produced by the simulation in Box 6.1 is close to $\rho_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E}$ parallels the arguments used to prove [BVY21, Lemma 6.2]. We cannot directly use the results in [BVY21, Section 6.1] because they only focus on the classical distributions of the answers, whereas we also need to take into account Eve's partial state.

The following lemma shows that if Alice and Bob share the state $\tilde{\Phi}^{(r_{-i},x,y)}_{E_A E_B E}$ between them and measure it with the measurements $\hat{A}_{r_{-i},x}$ and $\hat{B}_{r_{-i},y}$, then the resulting answers and Eve's state are distributed as in the parallely repeated strategy conditioned on the classical variables r_{-i}, x, y, a, b. Its proof follows the proof of [BVY21, Claim 6.3].

Lemma 6.6. *Let $\rho_E^{(r_{-i},x,y,a,b)}$ represent the state ρ conditioned on the classical variables $R_{-i} = r_{-i}, X_i = x, Y_i = y, A_i = a$, and $B_i = b$, that is,*

$$\rho_E^{(r_{-i},x,y,a,b)} = \mathop{\mathbb{E}}_{x_1^n, y_1^n, a_1^n, b_1^n | r_{-i},x,y,a,b} \left[\rho_E^{(x_1^n, y_1^n, a_1^n, b_1^n)} \right]. \tag{6.40}$$

We have the equality

$$\mathrm{tr}_{E_A E_B} \left(\hat{A}_{r_{-i},x}(a) \otimes \hat{B}_{r_{-i},y}(b) \tilde{\Phi}^{(r_{-i},x,y)}_{E_A E_B E} \right) = P_{A_i B_i | R_{-i} X_i Y_i}(a,b|r_{-i},x,y) \rho_E^{(r_{-i},x,y,a,b)}. \tag{6.41}$$

Proof. Using the definitions of $\tilde{\Phi}^{(r_{-i},x,y)}_{E_AE_BE}$, $\hat{A}_{r_{-i},x}(a)$ and $\hat{B}_{r_{-i},y}(b)$, we have

$$\begin{aligned}
&\mathrm{tr}_{E_AE_B}\left(\hat{A}_{r_{-i},x}(a)\otimes\hat{B}_{r_{-i},y}(b)\tilde{\Phi}^{(r_{-i},x,y)}_{E_AE_BE}\right) && (6.42)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}\mathrm{tr}_{E_AE_B}\left(\hat{A}_{r_{-i},x}(a)\otimes\hat{B}_{r_{-i},y}(b)A_{\omega_{-i},x}(a_C)^{1/2}\otimes B_{\omega_{-i},y}(b_C)^{1/2}\Psi\right.\\
&\qquad\qquad\left. A_{\omega_{-i},x}(a_C)^{1/2}\otimes B_{\omega_{-i},y}(b_C)^{1/2}\right) && (6.43)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}\mathrm{tr}_{E_AE_B}\left(\sum_{a_1^n|a_i,a_C}A_{\omega_{-i},x}(a_1^n)\otimes\sum_{b_1^n|b_i,b_C}B_{\omega_{-i},y}(b_1^n)\Psi\right) && (6.44)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}\sum_{\substack{a_1^n|a_i,a_C\\ b_1^n|b_i,b_C}}\mathrm{tr}_{E_AE_B}\left(\mathop{\mathbb{E}}_{x_1^n|\omega_{-i},x}A_{x_1^n}(a_1^n)\otimes\mathop{\mathbb{E}}_{y_1^n|\omega_{-i},y}B_{y_1^n}(b_1^n)\Psi\right) && (6.45)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}\mathop{\mathbb{E}}_{x_1^ny_1^n|\omega_{-i},x,y}\sum_{\substack{a_1^n|a_i,a_C\\ b_1^n|b_i,b_C}}\mathrm{tr}_{E_AE_B}\left(A_{x_1^n}(a_1^n)\otimes B_{y_1^n}(b_1^n)\Psi\right) && (6.46)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}\mathop{\mathbb{E}}_{x_1^ny_1^n|\omega_{-i},x,y}\sum_{\substack{a_1^n|a_i,a_C\\ b_1^n|b_i,b_C}}P_{A_1^nB_1^n|X_1^nY_1^n}(a_1^n,b_1^n|x_1^n,y_1^n)\rho_E^{(x_1^n,y_1^n,a_1^n,b_1^n)} && (6.47)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}\sum_{x_1^ny_1^n}\sum_{\substack{a_1^n|a_ia_C\\ b_1^n|b_ib_C}}P_{X_1^nY_1^n|\Omega_{-i}X_iY_i}(x_1^n,y_1^n|\omega_{-i},x,y)P_{A_1^nB_1^n|X_1^nY_1^n}(a_1^n,b_1^n|x_1^n,y_1^n)\rho_E^{(x_1^n,y_1^n,a_1^n,b_1^n)} && (6.48)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}\sum_{x_1^ny_1^n}\sum_{\substack{a_1^n|a_ia_C\\ b_1^n|b_ib_C}}P_{X_1^nY_1^nA_1^nB_1^n|\Omega_{-i}X_iY_i}(x_1^n,y_1^n,a_1^n,b_1^n|\omega_{-i},x,y)\rho_E^{(x_1^n,y_1^n,a_1^n,b_1^n)} && (6.49)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}P_{A_CB_CA_iB_i|\Omega_{-i}X_iY_i}(a_C,b_C,a_i,b_i|\omega_{-i},x,y)\\
&\qquad\sum_{x_1^ny_1^n}\sum_{\substack{a_1^n|a_ia_C\\ b_1^n|b_ib_C}}P_{X_1^nY_1^nA_1^nB_1^n|\Omega_{-i}X_iY_iA_CB_CA_iB_i}(x_1^n,y_1^n,a_1^n,b_1^n|\omega_{-i},x,y,a_C,b_C,a_i,b_i)\rho_E^{(x_1^n,y_1^n,a_1^n,b_1^n)} && (6.50)\\
&= \left\|\Phi^{(r_{-i},x,y)}\right\|^{-2}P_{A_CB_CA_iB_i|\Omega_{-i}X_iY_i}(a_C,b_C,a_i,b_i|\omega_{-i},x,y)\mathop{\mathbb{E}}_{x_1^n,y_1^n,a_1^n,b_1^n|r_{-i},x,y,a_i,b_i}\left[\rho_E^{(x_1^n,y_1^n,a_1^n,b_1^n)}\right] && (6.51)\\
&= P_{A_CB_C|\Omega_{-i}X_iY_i}(a_C,b_C|\omega_{-i},x,y)^{-1}P_{A_CB_CA_iB_i|\Omega_{-i}X_iY_i}(a_C,b_C,a_i,b_i|\omega_{-i},x,y)\rho_E^{(r_{-i},x,y,a,b)} && (6.52)\\
&= P_{A_iB_i|R_{-i}XY}(a,b|r_{-i}x,y)\rho_E^{(r_{-i},x,y,a,b)} && (6.53)
\end{aligned}$$

where we have used the fact that $P_{X_1^nY_1^n|\Omega_{-i},X_i,Y_i} = P_{X_1^n|\Omega_{-i},X_i}P_{Y_1^n|\Omega_{-i},Y_i}$ in Eq. 6.46 and the fact that given the questions the answers do not depend on Ω_{-i} in Eq. 6.49. □

Define the auxiliary state $\theta^{(0)}$ as

$$\theta^{(0)}_{R_{-i}X_iY_iE_AE_BE} := \sum_{r_{-i},x,y}P_{R_{-i}}(r_{-i})P_{XY}(x,y)[\![r_{-i},x,y]\!]\otimes\tilde{\Phi}^{(r_{-i},x,y)}_{E_AE_BE}. \tag{6.54}$$

One can view this as the state produced when r_{-i} is sampled according to $P_{R_{-i}}$, x,y are sampled according to the question distribution P_{XY} and the state $\tilde{\Phi}^{(r_{-i},x,y)}_{E_A E_B E}$ is distributed between Alice, Bob and Eve. We also define θ to be the state produced when Alice and Bob use the measurements $\hat{A}_{r_{-i},x}$ and $\hat{B}_{r_{-i},y}$ to measure $\theta^{(0)}$

$$\begin{aligned}
&\theta_{R_{-i}X_iY_iA_iB_iE} \\
&:= \sum_{r_{-i},x,y} P_{R_{-i}}(r_{-i})P_{XY}(x,y)[\![r_{-i},x,y]\!] \otimes \sum_{a,b}[\![a,b]\!] \otimes \mathrm{tr}_{E_A E_B}\left(\hat{A}_{r_{-i},x}(a) \otimes \hat{B}_{r_{-i},y}(b)\tilde{\Phi}^{(r_{-i},x,y)}_{E_A E_B E}\right) \qquad (6.55)\\
&= \sum_{r_{-i},x,y} P_{R_{-i}}(r_{-i})P_{XY}(x,y)[\![r_{-i},x,y]\!] \otimes \sum_{a,b} P_{A_iB_i|R_{-i}X_iY_i}(a,b|r_{-i},x,y)[\![a,b]\!] \otimes \rho^{(r_{-i},x,y,a,b)}_E. \qquad (6.56)
\end{aligned}$$

We will show that θ is close to both the simulated state and the real state. Using the triangle inequality, this will imply that the simulated state is also close to the real state.

Similar to $\theta^{(0)}$ above, we also define the state $\sigma^{(0)}$ after Step 3b in Box 6.1 (conditioned on $I = i$) as

$$\sigma^{(0)}_{R_{-i}X_iY_iE_AE_BE} := \sum_{r_{-i},x,y} P_{R_{-i}}(r_{-i})P_{XY}(x,y)[\![r_{-i},x,y]\!] \otimes \left(U^{E_A}_{r_{-i},x} \otimes V^{E_B}_{r_{-i},y}\right)\tilde{\Phi}^{(r_{-i},\perp,\perp)}_{E_A E_B E}\left(U^{E_A\dagger}_{r_{-i},x} \otimes V^{E_B\dagger}_{r_{-i},y}\right). \qquad (6.57)$$

The simulated state $\sigma_{R_{-i}X_iY_iA_iB_iE}$ for $I = i$ is simply the state obtained when $\sigma^{(0)}$ is measured using the measurements $\hat{A}_{r_{-i}x}$ and $\hat{B}_{r_{-i}y}$.

Using Proposition 6.4, it is straightforward to show that for a random $i \in C^c$, θ and σ are close to each other. The following lemma is essentially a generalisation of [**BVY21**, Claim 6.4].

Lemma 6.7. *For the state θ defined in Eq. 6.55 and the simulated state σ (conditioned on $I = i$), we have that*

$$\mathop{\mathbb{E}}_{I} \|\theta_{R_{-i}X_iY_iA_iB_iE} - \sigma_{R_{-i}X_iY_iA_iB_iE}\|_1 \leq O(\delta^{1/16}/\alpha^3)$$

where I is uniformly distributed at random in C^c.

Proof. We have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{I} \|\theta_{R_{-i}X_iY_iA_iB_iE} - \sigma_{R_{-i}X_iY_iA_iB_iE}\|_1 &\leq \mathop{\mathbb{E}}_{I} \|\theta^{(0)}_{R_{-i}X_iY_iE_AE_BE} - \sigma^{(0)}_{R_{-i}X_iY_iE_AE_BE}\|_1 \\
&= \mathop{\mathbb{E}}_{I}\mathop{\mathbb{E}}_{R_{-i}}\mathop{\mathbb{E}}_{XY}\left[\|\tilde{\Phi}^{(r_{-i},x,y)}_{E_AE_BE} - \left(U^{E_A}_{r_{-i},x} \otimes V^{E_B}_{r_{-i},y}\right)\tilde{\Phi}^{(r_{-i},\perp,\perp)}_{E_AE_BE}\left(U^{E_A\dagger}_{r_{-i},x} \otimes V^{E_B\dagger}_{r_{-i},y}\right)\|_1\right] \\
&\leq O(\delta^{1/16}/\alpha^3)
\end{aligned}$$

where the first line follows from the data processing inequality for norms, and the last line follows from Proposition 6.4 and the fact that $\|\psi - \phi\|_1 \leq \sqrt{2}\,\||\psi\rangle - |\phi\rangle\|$ for all pure states ψ and ϕ. □

Real state of the protocol ρ conditioned on $I_j = i$ can be written as

$$\rho_{R_{-i}X_iY_iA_iB_iE} = \sum_{r_{-i}} P_{R_{-i}}(r_{-i})[\![r_{-i}]\!] \otimes \sum_{x,y} P_{X_iY_i|R_{-i}}(x,y|r_{-i})[\![x,y]\!] \otimes \sum_{a,b} P_{A_iB_i|R_{-i}X_iY_i}(a,b|r_{-i},x,y)[\![a,b]\!] \otimes \rho_E^{(r_{-i},x,y,a,b)}. \tag{6.58}$$

Lemma 6.8. *The real state of the protocol ρ (conditioned on $I_j = i$) and the auxiliary state θ satisfy*

$$\mathop{\mathbb{E}}_{I} \|\theta_{R_{-i}X_iY_iA_iB_iE} - \rho_{R_{-i}X_iY_iA_iB_iE}\|_1 \leq O(\delta^{1/2}/\alpha^2)$$

where I is uniformly distributed at random in C^c.

Proof.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{I}\left[\|\theta_{R_{-i}X_iY_iA_iB_iE} - \rho_{R_{-i}X_iY_iA_iB_iE}\|_1\right] \\
&= \mathop{\mathbb{E}}_{I}\left[\left\|\sum_{r_{-i},x,y} P_{R_{-i}}(r_{-i})P_{XY}(x,y)[\![r_{-i},x,y]\!] \otimes \rho_{A_iB_iE}^{(r_{-i},x,y)} - \sum_{r_{-i},x,y} P_{R_{-i}XY}(r_{-i},x,y)[\![r_{-i},x,y]\!] \otimes \rho_{A_iB_iE}^{(r_{-i},x,y)}\right\|_1\right] \\
&= \mathop{\mathbb{E}}_{I}\left[\|P_{R_{-i}}P_{XY} - P_{R_{-i}X_iY_i}\|_1\right] \\
&\leq O(\delta^{1/2}/\alpha^2),
\end{aligned}$$

where the last line follows from the equation after Eq. 88 in [BVY21] (setting W_C to be the trivial event). □

Lemma 6.9. *The state σ produced by the single-round protocol for the 3CHSH$_\perp$ game in Box 6.1 approximates the state ρ. Specifically,*

$$\left\|\rho_{I_jR_{-I_j}X_{I_j}Y_{I_j}A_{I_j}B_{I_j}E} - \sigma_{IR_{-I}X_IY_IA_IB_IE}\right\|_1 = O(\delta^{1/16}/\alpha^3).$$

Proof. Using the fact that I_j in ρ is distributed uniformly at random in C^c, and Lemma 6.7 and 6.8, we have

$$\begin{aligned}
&\left\|\rho_{I_jR_{-I_j}X_{I_j}Y_{I_j}A_{I_j}B_{I_j}E} - \sigma_{IR_{-I}X_IY_IA_IB_IE}\right\|_1 = \mathop{\mathbb{E}}_{I}\|\rho_{R_{-i}X_iY_iA_iB_iE} - \sigma_{R_{-i}X_iY_iA_iB_iE}\|_1 \\
&\leq \mathop{\mathbb{E}}_{I}\|\theta_{R_{-i}X_iY_iA_iB_iE} - \rho_{R_{-i}X_iY_iA_iB_iE}\|_1 + \mathop{\mathbb{E}}_{I}\|\theta_{R_{-i}X_iY_iA_iB_iE} - \sigma_{R_{-i}X_iY_iA_iB_iE}\|_1 \\
&\leq O(\delta^{1/16}/\alpha^3)
\end{aligned}$$

where $\mathbb{E}_I$ represents expectation over I, which is sampled uniformly at random from C^c. □

Step 3: Bounding $H(A_{I_j}|EI_jR_{-I_j}X_{I_j})_\rho$ using the single-round entropy bound

Using Lemma 6.9 above, we show that the single-round strategy in Box 6.1 has a large winning probability if ρ has a large average winning probability. This will be helpful for bounding the entropy of the answers of the game given Eve's register while using the single-round bound in Lemma 6.3.

Let W_i denote the indicator random variable for the event that game G_i is won by the players. Since, W_i is a deterministic function of X_i, Y_i, A_i, B_i we have

$$|\Pr_\rho(W_{I_j}) - \Pr_\sigma(W_I)| \leq \left\|\rho_{I_jR_{-I_j}X_{I_j}Y_{I_j}A_{I_j}B_{I_j}W_{I_j}E} - \sigma_{IR_{-I}X_IY_IA_IB_IW_IE}\right\|_1 \leq O\left(\delta^{1/16}/\alpha^3\right). \quad (6.59)$$

For the state, ρ, I_j is an index chosen uniformly at random in C^c. Using Eq. 6.26, we have

$$\begin{aligned}\Pr_\rho(W_{I_j}) &\geq (1-\epsilon)\frac{\omega_{\text{th}}n - |C|}{n-|C|} \\ &\geq (1-\epsilon)\left(\omega_{\text{th}} - \delta\right) \end{aligned} \quad (6.60)$$

Combining this with Eq. 6.59, we get that

$$\Pr_\sigma(W_I) \geq \omega_{\text{th}} - O\left(\epsilon + \delta^{1/16}/\alpha^3\right) \quad (6.61)$$

We now bound $H(A_{I_j}|EI_jR_{-I_j}X_{I_j})_\rho$ using the entropy bound for the single-round 3CHSH$_\perp$ game. To use Lemma 6.3, we first use the Alicki-Fannes-Winter (AFW) continuity bound for the conditional entropy [**AF04**, **Win16**] to lower bound the entropy on ρ with the corresponding entropy on the simulated state σ:

$$\begin{aligned}H(A_{I_j}|EI_jR_{-I_j}X_{I_j})_\rho &\geq H(A_I|EIR_{-I}X_I)_\sigma - g(O(\delta^{1/16}/\alpha^3)) \\ &\geq F\left(g_{\alpha,\nu}(\Pr_\sigma[W])\right) - g(O(\delta^{1/16}/\alpha^3)) \\ &\geq F\left(g_{\alpha,\nu}(\omega_{\text{th}} - O(\epsilon + \delta^{1/16}/\alpha^3))\right) - O\left(\frac{\delta^{1/16}}{\alpha^3}\log\frac{1}{\delta}\right) \\ &\geq F\left(g_{\alpha,\nu}(\omega_{\text{th}})\right) - O((F\circ g_{\alpha,\nu})'(\omega_{\text{th}})(\epsilon + \delta^{1/16}/\alpha^3) - O\left(\frac{\delta^{1/16}}{\alpha^3}\log\frac{1}{\delta}\right) \\ &\geq F\left(g_{\alpha,\nu}(\omega_{\text{th}})\right) - O\left(\frac{\epsilon}{\nu} + \frac{\delta^{1/16}}{\alpha^3\nu}\log\frac{1}{\delta}\right)\end{aligned}$$

where $g(x) = 2x\log(|A|) + (x+1)\log(x+1) - x\log(x) = O\left(x\log\frac{|A|}{x}\right)$. We use the single-round entropy bound (Lemma 6.3) for the 3CHSH$_\perp$ game in the second line. This is valid because in the simulation of σ, Eve classically distributes I and R_{-I} to the players, so we can assume that she also holds copies of these registers. $(F\circ g_{\alpha,\nu})'$ represents the derivative of

the function $F \circ g_{\alpha,\nu}$ with respect to ω. We have used the fact that $F \circ g_{\alpha,\nu}$ is convex and increasing, and the bound $(F \circ g_{\alpha,\nu})'(\omega_{\text{th}}) \leq O(\frac{1}{\nu})$ above.

Finally, if we plug in the bound above in Eq. 6.30, we get

$$\begin{aligned} H(A_J|E\Omega_1^n J)_\rho &\geq \sum_{k=1}^{t} \mathop{\mathbb{E}}_{i_1^{k-1}} \left[H(A_{I_k}|EI_k R_{-I_k} X_{I_k})_\rho\right] \\ &\geq t \cdot \left(F\left(g_{\alpha,\nu}(\omega_{\text{th}})\right) - O\left(\frac{\epsilon}{\nu} + \frac{\delta^{1/16}}{\alpha^3 \nu} \log \frac{1}{\delta}\right)\right) \end{aligned} \tag{6.62}$$

6.6.2. Bounding information reconciliation cost

To take the information reconciliation cost into account, we consider the quantity $H(A_J|B_J J)_{\rho_{\text{honest}}}$ for the honest protocol (with no Eve) under noisy conditions. We model the noise between Alice and Bob as a depolarising channel with noise parameter $2Q$. If Alice sends one half of a perfect Bell state to Bob for each round in the honest protocol, then their shared state is $\eta_{E_A E_B}^{\otimes n}$ where

$$\eta_{E_A E_B} = (1 - 2Q) \left|\Phi^+\right\rangle \left\langle\Phi^+\right| + 2Q \tau_{E_A E_B} \tag{6.63}$$

where $|\Phi^+\rangle$ is a Bell state and $\tau_{E_A E_B}$ is the completely mixed state. The noise parameter, Q can be measured using the qubit error rate through the equation [**AF20**, Section 4.2.4]

$$\Pr[A \neq B | X, Y = (0,2)] = Q. \tag{6.64}$$

We will assume that $Q \leq 0.1$. With probability at least $(1-\alpha)^2(1-\nu)(1-Q)$, Alice and Bob's answers are equal in every round. Since the shared state is i.i.d, we can simply bound the entropy as

$$\begin{aligned} H(A_J|B_J J)_{\rho_{\text{honest}}} &= t \cdot H(A|B)_\eta \\ &\leq t \cdot h(1 - (1-\alpha)^2(1-\nu)(1-Q)) \\ &\leq t \cdot h(2\alpha + \nu + Q) \end{aligned}$$

where the last line is true for $\alpha, \nu, Q \in (0, 0.1)$.

6.6.3. Bounding key length

We can now estimate the length of the key produced during the parallel DIQKD protocol as:

$$\begin{aligned} &H(A_J|EJ\Omega_1^n)_\rho - H(A_J|B_J J)_{\rho_{\text{honest}}} \\ &\quad \geq t \left(F\left(g_{\alpha,\nu}(\omega_{\text{th}})\right) - O\left(\frac{\epsilon}{\nu} + \frac{\delta^{1/16}}{\alpha^3 \nu} \log \frac{1}{\delta}\right) - h(2\alpha + \nu + Q)\right) \end{aligned}$$

where $t = \frac{\delta}{\log|\mathcal{A}||\mathcal{B}|+\delta} n = \Omega(n)$. Note that we can choose ω_{th} close to the maximum probability of winning so that $F \circ g_{\alpha,\nu}(\omega_{\text{th}})$ is at least some constant, say $\frac{1}{2}\log(2)$. α, ν and Q can be chosen close to zero so that the information reconciliation term is small. Finally, with these parameters fixed, we can choose δ small enough so that the key length is $\Omega(n)$ and the key rate is positive. This completes the proxy von Neumann entropy based security proof.

6.7. Security proof

The von Neumann entropy based security argument illuminates the fundamental mechanism behind entropy accumulation in parallel DIQKD. It demonstrates that Alice's answers A_J are random with respect to the adversary Eve because, for every $k \in [t]$, one can approximate the questions and answers in the partial state $\rho_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E}$ as the output of a single-round strategy for the $3\text{CHSH}_\perp$ game, which has a high winning probability. Now, we need to port the lower bound on the von Neumann entropy of Alice's answers to a smooth min-entropy lower bound. To chart the course forward, we can compare this situation with sequential DIQKD, where Alice's answers in each round directly result from a single-round strategy. This direct relationship ensures their randomness according to the single-round entropy bound. Consequently, when the average winning probability is high, entropy accumulates across all rounds. The entropy accumulation theorem (EAT) (Sec. 2.6) serves as the primary information-theoretic tool for demonstrating this accumulation [**AFDF+18**, **AFRV19**]. We sketched the security proof for sequential DIQKD in Sec. 2.8.3.

To prove security for the parallel DIQKD protocol, we require a tool analogous to EAT. Drawing insights from the von Neumann entropy based argument, we can identify key features necessary for such a tool. The most crucial distinction between the setting for EAT (Fig. 2.1) and the state ρ produced in parallel DIQKD is the absence of a sequential or structured process generating the answers in ρ. Instead, Alice and Bob produce answers *in parallel* using a single measurement channel. The primary source of structure, and the reason for expecting entropy accumulation, lies in our ability to approximate Alice and Bob's questions and answers on round I_k along with Eve's information using the output of a single-round strategy. Consequently, the entropic tool we employ should not rely on a sequential structure for the state. Rather, it should demonstrate entropy accumulation when the partial state of the given state can be approximated as the output of a suitable channel. These requirements are all addressed by the unstructured approximate entropy accumulation theorem (Theorem 5.8 and 5.10) developed in the previous chapter.

Note that we cannot use Theorem 3.12 here. Unlike the unstructured approximate EAT, this theorem applies only to sequential processes. Moreover, it considers much stricter channel approximations to the channels $\mathcal{M}_k$ forming this process. The unstructured approximate EAT's relaxed assumptions come with an inherent limitation: the smoothing parameter must depend on the approximation parameter. Nevertheless, in our security proof for parallel DIQKD, this limitation is not significant since we can choose the approximation parameter using the protocol parameter δ.

In the following section, we apply Theorem 5.10 to prove security for parallel DIQKD in the one-shot regime.

6.7.1. Using unstructured approximate EAT for the smooth min-entropy bound

In order to prove security for Protocol 6.1, we need to bound the smooth min-entropy of Alice's raw key with respect to Eve's information, that is, we need to lower bound

$$H_{\min}^{\epsilon}(A_J|E\Omega_1^n JT_1^t X_S A_S)_{\rho_{|\neg F}}. \tag{6.65}$$

Recall that $J = \{I_1, I_2, \cdots, I_t\}$. For simplicity, for a sequence of variables V_1^n (like X_1^n, A_1^n etc.) and $j \in [t]$, we define the variable

$$\hat{V}_j := V_{I_j}. \tag{6.66}$$

For example the entropy above in Eq. 6.65 can be written as $H_{\min}^{\epsilon}(\hat{A}_1^t|E\Omega_1^n I_1^t T_1^t X_S A_S)_{\rho_{|\neg F}}$. We will begin by using the unstructured approximate EAT to prove that

$$H_{\min}^{\epsilon}(\hat{A}_1^t\hat{B}_1^t|\hat{X}_1^t\hat{Y}_1^t T_1^t I_1^t E\Omega_{J^c})_{\rho_{|\neg F}} \geq \Omega(t). \tag{6.67}$$

Through a minor modification of the simulation in Box 6.1, we will show that the states $\rho_{\hat{A}_1^j\hat{B}_1^j\hat{X}_1^j\hat{Y}_1^j T_1^j I_1^t E\Omega_{J^c}}$ can be approximated by states, which are produced by playing a single-round 3CHSH$_\perp$ game. This modified simulation is presented in Box 6.2. Let $\sigma^{(j)}$ denote the state produced by this simulation.

Claim 6.10. *For every $j \in [t]$, the state $\sigma^{(j)}$ produced through Box 6.2 satisfies*

$$\left\| \rho_{\hat{A}_1^j\hat{B}_1^j\hat{X}_1^j\hat{Y}_1^j T_1^j I_1^t E\Omega_{J^c}} - \sigma^{(j)}_{\hat{A}_1^j\hat{B}_1^j\hat{X}_1^j T_1^j\hat{Y}_1^j I_1^t E\Omega_{J^c}} \right\|_1 \leq O\left(\frac{\delta^{1/16}}{\alpha^3}\right). \tag{6.68}$$

Proof. In the first step of Box 6.2, condition on the choice of the indices $I_1^{j-1} = i_1^{j-1}$. This fixes the subset $C \subseteq J$. For this choice of C, consider the state $\sigma^{(j)}_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E|C=i_1^{j-1}}$ produced

Single-round protocol for producing $\sigma^{(j)}_{\hat{A}_1^j\hat{B}_1^j\hat{X}_1^j\hat{Y}_1^j I_1^t T_1^j E\Omega_{J^c}}$:

(1) Eve randomly samples $J \coloneqq \{I_1, I_2, \cdots, I_t\} \subseteq [n]$ of size t. Define $C \coloneqq I_1^{j-1}$. She also samples T_1^{j-1} i.i.d. from $\{0,1\}$ with $\Pr(T_i = 1) = \gamma$.

(2) Eve randomly samples the random variable $R_{-I_j} = (A_C, B_C, X_C, Y_C, \Omega_{(C\cup\{i\})^c}) = (\hat{A}_1^{j-1}, \hat{B}_1^{j-1}, \hat{X}_1^{j-1}, \hat{Y}_1^{j-1}, \Omega_{J^c}, \hat{\Omega}_{j+1}^t)$ according to $P_{R_{-I_j}}$ depending on the value of I_j.

(3) Eve distributes C, I_j, R_{-I_j} to both Alice and Bob. Call their copies $C^{(A)}, I_j^{(A)}, R_{-I_j}^{(A)}$ and $C^{(B)}, I_j^{(B)}, R_{-I_j}^{(B)}$.

(4) Let $I_j = i$ and $R_{-I_j} = r_{-i}$. Eve distributes the registers E_A and E_B of the state $|\tilde{\Phi}^{(r_{-i},\perp,\perp)}\rangle_{E_A E_B E}$ between Alice and Bob.

(5) Alice and Bob play the $3\text{CHSH}_\perp$ (EAT map):

  (a) The questions x and y are sampled according to P_{XY} and sent to Alice and Bob.

  (b) Let $U^{E_A}_{r_{-i},x}$ and $V^{E_B}_{r_{-i},y}$ be the unitaries defined by Proposition 6.4. Alice applies the unitary $U_{r_{-i},x}$ to her register E_A. Similarly, Bob applies $V_{r_{-i},y}$ to his register E_B.

  (c) Alice and Bob measure their registers with $\{\hat{A}_{r_{-i},x}(a)\}_a$ and $\{\hat{B}_{r_{-i},y}(b)\}_b$ to generate their answers for the game.

  (d) Alice randomly samples $T_j \in \{0,1\}$ with $\Pr(T_j = 1) = \gamma$.

(6) Eve traces over $\hat{\Omega}_{j+1}^t$ in her R_{-I_j}.

Box 6.2

by the process in Box 6.2 before Step 6 (ignoring I_{j+1}^t and T_1^j for now). Observe that this state is identical to the state produced using the simulation in Box 6.1 for the fixed subset C. Using Lemma 6.9, we have that

$$\left\| \rho_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E | C = i_1^{j-1}} - \sigma^{(j)}_{I_j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E | C = i_1^{j-1}} \right\|_1 \leq O\left(\frac{\delta^{1/16}}{\alpha^3}\right). \tag{6.69}$$

Now, in the real protocol the subset $C = I_1^{j-1}$ is distributed as a random $j-1$ size subset of $[n]$. This is exactly how I_1^{j-1} are also distributed in the simulation protocol in Box 6.2. Therefore, we have

$$\left\| \rho_{I_1^j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E} - \sigma^{(j)}_{I_1^j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E} \right\|_1 \leq O\left(\frac{\delta^{1/16}}{\alpha^3}\right). \tag{6.70}$$

We also need to incorporate I_{j+1}^t in the inequality above. Let $\Phi : I_1^j \to I_1^t$ be the map which simply reads I_1^j and randomly samples I_j^t outside this set in $[n]$. This channel produces the

correct distribution on I_1^t when applied to both $\rho_{I_1^j}$ and $\sigma^{(j)}_{I_1^j}$. Further, observe that both ρ and $\sigma^{(j)}$ satisfy the Markov chains $I_{j+1}^t \leftrightarrow I_1^j \leftrightarrow R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E$. Therefore, using the data processing inequality for the trace norm we have

$$\begin{aligned}
&\left\| \rho_{I_1^t R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E} - \sigma^{(j)}_{I_1^t R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E} \right\|_1 \\
&\qquad = \left\| \Phi\left(\rho_{I_1^j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E} \right) - \Phi\left(\sigma^{(j)}_{I_1^j R_{-I_j} X_{I_j} Y_{I_j} A_{I_j} B_{I_j} E} \right) \right\|_1 \\
&\qquad \leq O\left(\frac{\delta^{1/16}}{\alpha^3} \right). \qquad (6.71)
\end{aligned}$$

We can now also consider the last step in Box 6.2 by tracing over $\hat{\Omega}^t_{j+1}$. This gives us

$$\left\| \rho_{I_1^t \hat{X}_1^j \hat{Y}_1^j \hat{A}_1^j \hat{B}_1^j \Omega_{J^c} E} - \sigma^{(j)}_{I_1^t \hat{X}_1^j \hat{Y}_1^j \hat{A}_1^j \hat{B}_1^j \Omega_{J^c} E} \right\|_1 \leq O\left(\frac{\delta^{1/16}}{\alpha^3} \right). \qquad (6.72)$$

Lastly, we also need to account for T_1^j. These are sampled independently in both ρ and $\sigma^{(j)}$ with the same distribution. Hence, we can account for these simply and the claim follows from the bound above. □

We will now apply Theorem 5.10 to the state ρ. The approximation chain given by the simulation procedure in Box 6.2 is useful for this purpose. For $j \in [t]$, let's define

$$\tilde{E}_A^{(j)} := E_A I_j^{(A)} C^{(A)} R^{(A)}_{-I_j} \text{ and } \tilde{E}_B^{(j)} := E_B I_j^{(B)} C^{(B)} R^{(B)}_{-I_j} \qquad (6.73)$$

in Box 6.2 to be the entirety of Alice and Bob's registers before they start playing the 3CHSH$_\perp$ game in Step 5 of Box 6.2. Let $\mathcal{M}_j : \tilde{E}_A^{(j)} \tilde{E}_B^{(j)} \to \hat{X}_j \hat{Y}_j T_j \hat{A}_j \hat{B}_j$ be the channel applied by Alice and Bob in Step 5. This channel samples the questions $\hat{X}_j$ and $\hat{Y}_j$ for the 3CHSH$_\perp$ game, applies Alice and Bob's measurements on $\tilde{E}_A^{(j)}$ and $\tilde{E}_B^{(j)}$ to produce their answers and also randomly samples T_j. Note that the Steps 5 and 6 commute, so we can change their order without affecting Claim 6.10. For the following assume that Eve performs Step 6 before Alice and Bob play the game (Step 5). Finally, let the state $\sigma^{(j,0)}_{\hat{A}_1^{j-1} \hat{B}_1^{j-1} \hat{X}_1^{j-1} \hat{Y}_1^{j-1} T_1^{j-1} I_1^t \tilde{E}_A^{(j)} \tilde{E}_B^{(j)} E \Omega_{J^c}}$ be the state between Alice, Bob and Eve before the 3CHSH$_\perp$ game is played. With these definitions, we have that for every $j \in [t]$

$$\sigma^{(j)}_{\hat{A}_1^j \hat{B}_1^j \hat{X}_1^j \hat{Y}_1^j T_1^j I_1^t E \Omega_{J^c}} = \mathcal{M}_j \left(\sigma^{(j,0)}_{\hat{A}_1^{j-1} \hat{B}_1^{j-1} \hat{X}_1^{j-1} \hat{Y}_1^{j-1} T_1^{j-1} I_1^t \tilde{E}_A^{(j)} \tilde{E}_B^{(j)} E \Omega_{J^c}} \right). \qquad (6.74)$$

Using Claim 6.10 we further have that for every $j \in [t]$

$$\frac{1}{2} \left\| \rho_{\hat{A}_1^j \hat{B}_1^j \hat{X}_1^j \hat{Y}_1^j T_1^j I_1^t E \Omega_{J^c}} - \sigma^{(j)}_{\hat{A}_1^j \hat{B}_1^j \hat{X}_1^j \hat{Y}_1^j T_1^j I_1^t E \Omega_{J^c}} \right\|_1 \leq \epsilon \qquad (6.75)$$

for $\epsilon = O\left(\frac{\delta^{1/16}}{\alpha^3}\right)$. We choose the testing maps $\mathcal{T}_j : \hat{X}_j\hat{Y}_jT_j\hat{A}_j\hat{B}_j \to \hat{X}_j\hat{Y}_jT_j\hat{A}_j\hat{B}_jW_j$ to be the classical channel which outputs the register W_j according to

$$W_j = \begin{cases} V(\hat{X}_j, \hat{Y}_j, \hat{A}_j, \hat{B}_j) & \text{if } T_j = 1 \\ \perp & \text{if } T_j = 0 \end{cases} \tag{6.76}$$

We make the following choices in order to use Theorem 5.10,

$$A_j \leftarrow \hat{A}_j\hat{B}_j \tag{6.77}$$

$$B_j \leftarrow \hat{X}_j\hat{Y}_jT_j \tag{6.78}$$

$$X_j \leftarrow W_j \tag{6.79}$$

$$E \leftarrow EI_1^t\Omega_{J^c}. \tag{6.80}$$

Note that the side information $\hat{X}_j\hat{Y}_jT_j$ is sampled independent of the input state by the channel $\mathcal{M}_j$. Lastly, we will condition on the event $\neg F$, which is equivalent to the event $\text{freq}(W_1^t)(1) \geq \gamma\omega_{\text{th}}$.

We will use Lemma 6.3 to define an affine min-tradeoff function for the channels $\{\mathcal{M}_j\}_{j=1}^t$. Fix $j \in [t]$. For an arbitrary state $\nu^{(0)}_{\tilde{E}_A^{(j)}\tilde{E}_B^{(j)}R}$, the state $\nu_{\hat{X}_j\hat{Y}_jT_j\hat{A}_j\hat{B}_jW_j\tilde{R}} = \mathcal{T}_j \circ \mathcal{M}_j\left(\nu^{(0)}\right)$ satisfies

$$\nu_{T_j} = \gamma\left|1\right\rangle\left\langle 1\right| + (1-\gamma)\left|0\right\rangle\left\langle 0\right|. \tag{6.81}$$

Further, we can write the output state ν on register W_j as

$$\nu_{W_j} = \gamma(1-\omega)\left|0\right\rangle\left\langle 0\right| + \gamma\omega\left|1\right\rangle\left\langle 1\right| + (1-\gamma)\left|\perp\right\rangle\left\langle \perp\right| \tag{6.82}$$

for some $\omega \in [0,1]$. Observe that ω here is actually the winning probability of the $3\text{CHSH}_\perp$ game for the strategy given by the initial state $\nu^{(0)}$ and the measurements that comprise the channel $\mathcal{M}_j$. Let's define the function $F_{\alpha,\nu}$ piecewise as

$$F_{\alpha,\nu}(x) := \begin{cases} (1-\alpha)F\left(g_{\alpha,\nu}\left(\frac{x}{\gamma}\right)\right) & \text{if } \frac{x}{\gamma} \in [\omega_{\min}, \omega_{\max}] \\ 0 & \text{else} \end{cases} \tag{6.83}$$

where $\omega_{\min} := 1 - \frac{(1-\alpha)^2\nu}{4}$, $\omega_{\max} := 1 - \frac{2-\sqrt{2}}{4}(1-\alpha)^2\nu$, and the functions F and $g_{\alpha,\nu}$ are defined in Eq. 6.3. Using Lemma 6.3 for the state ν, we have

$$H(AB|EXY)_\nu \geq F_{\alpha,\nu}\left(\nu_{W_j}(1)\right) \tag{6.84}$$

We can transform this lower bound into an affine function by using the fact that $F_{\alpha,\nu}$ is convex and that a convex function lies above its slope. We consider the slope at point $\gamma\omega_{\text{th}}$. For $0 \leq x \leq \gamma\omega_{\max}$, we have

$$F_{\alpha,\nu}(x) \geq F_{\alpha,\nu}(\gamma\omega_{\text{th}}) + F'_{\alpha,\nu}(\gamma\omega_{\text{th}})\left(x - \gamma\omega_{\text{th}}\right) \tag{6.85}$$

$$=: \bar{F}_{\alpha,\nu}(x)\,. \tag{6.86}$$

We have defined the right-hand side above as the linear function $\bar{F}_{\alpha,\nu}$. Note that Eq. 6.82 implies that irrespective of the input state $\nu^{(0)}$, $\nu_{W_j}(\perp) = 1-\gamma$. So, for any probability distribution q_{W_k} such that $q(\perp) \neq 1-\gamma$, we have that $\Sigma(q|\mathcal{M}_j) = \varnothing$ (Eq. 5.101). This is also true for any distribution q such that $q(1)/\gamma > \omega_{\max}$, which is the maximum winning probability for the 3CHSH$_\perp$ game. As a result, for the distribution q_{W_j}, the function $q \mapsto \bar{F}_{\alpha,\nu}(q(1))$ is a min-tradeoff function for the channels $\{\mathcal{M}_j\}_{j=1}^t$.

One can easily evaluate the derivative of $F_{\alpha,\nu}$:

$$F'_{\alpha,\nu}(\omega_{\text{th}}) = \frac{1}{\nu\gamma(1-\alpha)} F'(g_{\alpha,\nu}(\omega_{\text{th}})) \leq O\left(\frac{1}{\nu\gamma}\right) \tag{6.87}$$

This gives us that $\left\|\nabla \bar{F}_{\alpha,\nu}\right\|_\infty \leq O\left(\frac{1}{\nu\gamma}\right)$, which is required while applying Theorem 5.10. Finally, applying the approximate EAT to the state ρ as described above, we get that if $\Pr(\neg F) \geq 2\mu$[(3)], then

$$H_{\min}^{\mu'+\epsilon'}(\hat{A}_1^t \hat{B}_1^t | \hat{X}_1^t \hat{Y}_1^t T_1^t I_1^t \Omega_{J^c} E)_{\rho_{|\neg F}} \geq t\left((1-\alpha)F\left(g_{\alpha,\nu}(\omega_{\text{th}})\right) - O\left(\frac{\sqrt{\mu}}{\nu\gamma}\right)\right) - O(1) \tag{6.88}$$

where

$$\delta \in (0,1) \tag{6.89}$$

$$t = \frac{\delta}{\log|\mathcal{A}||\mathcal{B}| + \delta} n \tag{6.90}$$

$$\epsilon = O\left(\frac{\delta^{1/16}}{\alpha^3}\right) \tag{6.91}$$

$$\mu := \left(4(4\sqrt{\epsilon}+\epsilon)\log\frac{|\mathcal{A}||\mathcal{B}||\mathcal{X}||\mathcal{Y}|||\mathcal{T}|}{\epsilon}\right)^{1/3} = O\left(\epsilon^{1/6}\left(\log\frac{1}{\epsilon}\right)^{1/3}\right) \tag{6.92}$$

$$\mu' := 2\sqrt{\frac{\mu}{\Pr_\rho(\neg F)}} = O\left(\epsilon^{1/12}\left(\log\frac{1}{\epsilon}\right)^{1/6}\right) \tag{6.93}$$

and $\epsilon' = \Omega(1) \in (0,1)$ such that $\mu' + \epsilon' < 1$.

In Appendix D.3, using standard techniques, we derive a bound for the entropy of Alice's raw key with respect to Eve's information, $H_{\min}^{O(\mu')}(\hat{A}_1^t | T_1^t I_1^t \Omega_1^n E X_S A_S)_{\rho_{|\neg F}}$ from the bound above. We use the fact that if the 3CHSH$_\perp$ games are won with a high probability, then Alice's and Bob's answers have a small relative distance. Consequently, Alice's answers alone possess high entropy relative to Eve's information. We account for the information disclosed

[(3)] If $\Pr(\neg F) < 2\mu$, then one can show that the secrecy condition for QKD is satisfied for a security parameter greater than 2μ (Eq. 2.60)

during the testing procedure, X_S and A_S, by applying a straightforward dimension bound. Through these arguments, we prove that:

$$H_{\min}^{\mu'+8\epsilon'}(A_J|JT_1^t\Omega_1^n EX_S A_S)_{\rho_{|\neg F}} - \text{leak}_{\text{IR}}$$
$$\geq t\left((1-\alpha)F\left(g_{\alpha,\nu}(\omega_{\text{th}})\right) - O\left(\frac{\sqrt{\mu}}{\nu\gamma}\right) - 2h(2(\nu+\alpha+\delta_1)) - 2\log|\mathcal{A}|\gamma\right) - O(1) \quad (6.94)$$

where leak_{IR} is the information reconciliation cost, $\delta_1 \in (0,1)$ is a small parameter and the rest of the parameters are defined and chosen as in Eq. 6.93.

We can make the lower bound above at least $\gtrsim t\frac{\log(2)}{2}$ by choosing ω_{th} such that $g_{\alpha,\nu}(\omega_{\text{th}}) \approx 0.84$ (the winning probability of the CHSH games). This choice results in $F(g_{\alpha,\nu}(\omega_{\text{th}})) \geq \frac{3\log(2)}{4}$. Note that this winning probability threshold is sufficiently below the maximal winning probability to allow for a robust implementation that accounts for experimental imperfections. We can further choose $\alpha = \gamma = \nu = \delta_1 = 10^{-3}$. With these choices, the combined terms $\alpha F(g_{\alpha,\nu}(\omega_{\text{th}})) + 2h(2(\nu+\alpha+\delta)) + 2\gamma\log|\mathcal{A}|$ sum together less than $0.1\log(2)$. By choosing δ small enough, the term $O\left(\frac{\sqrt{\mu}}{\nu\gamma}\right)$ can be made smaller than $0.1\log(2)$ and the security parameter μ' can also be made small enough. Once we fix a value of δ, we obtain a fixed small key rate ($\geq \frac{\delta\log(2)}{2(\log|\mathcal{A}||\mathcal{B}|+\delta)}$ in this example) for a fixed security parameter given by the corresponding value of μ'. It is important to note that one limitation of our approach is the interdependence of the key rate and the security parameter. We have essentially proven that the protocol is $\tilde{O}(\epsilon_s)$ secure for a rate $\Omega(\epsilon_s^{192})$. Consequently, if one wishes to make the security parameter smaller, they must also reduce the key rate.

6.8. Conclusion

In this work, we have developed an alternative approach to proving the security of parallel DIQKD. This approach is based on using results from the work on parallel repetition for anchored games to break the multi-round parallel strategy used by Alice and Bob's quantum devices into multiple approximately single-round strategies. We can then prove that entropy accumulates in this protocol using the unstructured approximate EAT. Our approach yields a more information theoretic and general proof compared to those presented by [JMS20] and [Vid17].

However, a major drawback of our technique is that it couples the key rate of the protocol with the security parameter. The most important and immediate problem arising from this work is whether this dependence can be broken. We expect that it might be possible to use stronger properties of testing in the unstructured approximate EAT to break

this relation. We believe that efforts to enhance the performance of our entropic method to match and potentially surpass that of [JMS20] and [Vid17] could reveal deeper insights about approximation chains. Furthermore, it would also be interesting to explore whether the results and techniques employed in this chapter have implications in the broader context of parallel repetition.

At a broader level, whether we can improve the rates of parallel DIQKD to match those of its sequential counterpart still remains an open problem. Given the complexity inherent to general non-local game and their parallel repetitions, it seems unlikely that one can match the performance of sequential DIQKD without relying on specific properties of the underlying games. A promising direction could be to select or engineer these games such that the parallel key rates approach those of sequential DIQKD. For example, certain games like XOR game satisfy strong parallel repetition [CSUU08], that is, the strategy of playing each game in the parallel repetition independently using the optimal strategy yields the optimal winning probability. It's conceivable that for a similar class of games, Eve's optimal attack in parallel DIQKD might be constrained to be i.i.d. or close to it. Finally, we note that proving the security of parallel device-independent randomness expansion still remains an interesting open problem.

Chapter 7

Exploring further...

In each of the chapters, we have discussed the immediate questions and research directions emerging from the work presented. In this chapter, we broaden our focus to explore approximation chains in the wider field of quantum information and how our work may impact their analysis.

(1) **Imperfections and leakage in cryptographic protocols**: Incorporating imperfections and leakages into security proofs is crucial for analysing real-world implementations of quantum cryptographic protocols. We previously highlighted this while motivating the development of the first approximate entropy accumulation theorem in Chapter 3. The unstructured approximate EAT presented in Chapter 5 may prove more suitable for analysing imperfections, provided one can utilise additional testing to decouple the smoothing parameter from the approximation parameter in this theorem. Exploring the application of these theorems to prove the security of cryptographic protocols under imperfections and leakages represents a promising line of research. Furthermore, it would be valuable to understand the tradeoffs and assumptions required for such proofs in comparison to other recently developed tools for similar tasks [**Tan23**, **AMT24**].

(2) **Parallel repetition based proofs**: The arguments and techniques developed for analysing parallel repetition and decomposing parallel protocols [**Raz98**, **Hol09**, **BVY22**] are incredibly strong and general. In this thesis, we use them for proving the security of parallel DIQKD. Additionally, variants of these arguments have been successfully used to establish direct product results in communication complexity [**JPY12**, **BRWY13**, **JK21**, **JK23**]. We have also been able to leverage the parallel repetition argument to provide an alternative proof for

the strong converse of classical channel capacity [MD24d]. The general quantum problems for both these tasks remains an open problem.

Given the generality of parallel repetition techniques, we expect these to serve as important tools for tackling such problems. As we saw in Chapter 6, one of the main components of these arguments is the development of a structured approximation chain, which is easy to manipulate. We anticipate that the tools developed in this thesis will facilitate the application of these techniques in the future.

(3) **Fault-tolerant channel coding**: This direction once again explores protocol performance under error conditions. While calculating the communication capacities for quantum channels, the quantum gates implementing encoders and decoders are assumed to be noiseless. However, this is an unrealistic assumption, especially for near term quantum devices. Research on fault-tolerant channel coding aims to use fault-tolerant computation techniques developed for quantum computing to communicate information over quantum channels with noisy local quantum gates, achieving rates close to the capacities with noiseless gates [CMH24, BCMH24].

The techniques for analysing approximation chains might prove useful for these problems as well. Under certain conditions, it may be possible to view the states produced by the encoders in these problems as an approximation chain (or a similar construct) of the perfectly encoded state. For instance, consider a scenario where the encoder produces a state on registers X_1^n, and on each register X_k, the state produced is almost perfect. In this case, it might be possible to demonstrate that the information-theoretic distance between the real and ideal state is $\sim n\epsilon$. Then, the entropic triangle inequality could be used to prove that the decrease in the amount of information communicated with the real states is also $\sim n\epsilon$, which would further imply that the rate decreases by $\sim \epsilon$.

(4) **Analysing infinite-dimensional protocols**: Cryptographic and communication protocols are often implemented using optical signals, which are represented by infinite-dimensional Hilbert spaces. A common strategy for reducing the analysis of such protocols to the finite-dimensional case involves using a large projector to project the signal onto a finite-dimensional space (see, for example, [KGL+23]). However, if there are n signals communicated during the protocol and the truncation causes a deviation of ϵ from the real protocol for each signal, then the protocol with

the truncated signals will be $n\epsilon$ far from the original implementation. Similar to the problems addressed in this thesis, it is generally preferable to treat such a linearly growing term as an information or entropy loss rather than see it added to the protocol error. It is seems that such an analysis can be modelled using an approximation chain, and that the techniques presented in this thesis would be beneficial in these contexts as well.

References

[ABJT19] Anurag Anshu, Mario Berta, Rahul Jain, and Marco Tomamichel. A minimax approach to one-shot entropy inequalities. *Journal of Mathematical Physics*, 60(12):122201, dec 2019. doi:10.1063/1.5126723.

[ABJT20] Anurag Anshu, Mario Berta, Rahul Jain, and Marco Tomamichel. Partially smoothed information measures. *IEEE Transactions on Information Theory*, 66(8):5022–5036, 2020. doi:10.1109/TIT.2020.2981573.

[AF04] R Alicki and M Fannes. Continuity of quantum conditional information. *Journal of Physics A: Mathematical and General*, 37(5):L55–L57, Jan 2004. doi:10.1088/0305-4470/37/5/l01.

[AF20] Rotem Arnon-Friedman. *Device-Independent Quantum Information Processing*. Springer International Publishing, 2020. doi:10.1007/978-3-030-60231-4.

[AFDF+18] Rotem Arnon-Friedman, Frédéric Dupuis, Omar Fawzi, Renato Renner, and Thomas Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nature Communications*, 9(1):459, 2018. doi:10.1038/s41467-017-02307-4.

[AFRV19] Rotem Arnon-Friedman, Renato Renner, and Thomas Vidick. Simple and tight device-independent security proofs. *SIAM Journal on Computing*, 48(1):181–225, jan 2019. doi:10.1137/18m1174726.

[AMT24] Amir Arqand, Tony Metger, and Ernest Y. Z. Tan. Mutual information chain rules for security proofs robust against device imperfections, 2024, arXiv:2407.20396. URL: https://arxiv.org/abs/2407.20396.

[Aud14] Koenraad M. R. Audenaert. Telescopic relative entropy. In Dave Bacon, Miguel Martin-Delgado, and Martin Roetteler, editors, *Theory of Quantum Computation, Communication, and Cryptography*, pages 39–52, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[BACGPH22] Andreas Bluhm, Ángela Capel, Paul Gondolf, and Antonio Pérez-Hernández. Continuity of quantum entropic quantities via almost convexity, 2022, arXiv:2208.00922.

[BB84] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *Proceedings of the International Conference on Computers, Systems & Signal Processing, Bangalore, India*, pages 175–179, 1984.

[BCMH24] Paula Belzig, Matthias Christandl, and Alexander Müller-Hermes. Fault-tolerant coding for entanglement-assisted communication. *IEEE Trans. Inf. Theor.*, 70(4):2655–2673, January 2024. doi:10.1109/TIT.2024.3354319.

[Ben92] Charles H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68:3121–3124, May 1992. doi:10.1103/PhysRevLett.68.3121.

[BF10] Niek J. Bouman and Serge Fehr. Sampling in a quantum population, and applications. In Tal Rabin, editor, *Advances in Cryptology – CRYPTO 2010*, pages 724–741, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[BFF21] Peter Brown, Hamza Fawzi, and Omar Fawzi. Computing conditional entropies for quantum correlations. *Nature Communications*, 12(1):575, 2021. doi:10.1038/s41467-020-20018-1.

[BFT17] Mario Berta, Omar Fawzi, and Marco Tomamichel. On variational expressions for quantum relative entropies. *Letters in Mathematical Physics*, 107(12):2239–2265, 2017. doi:10.1007/s11005-017-0990-7.

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997.

[Bha07] Rajendra Bhatia. *Positive Definite Matrices*. Princeton University Press, Princeton, 2007. doi:10.1515/9781400827787.

[BHK05] Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95:010503, Jun 2005. doi:10.1103/PhysRevLett.95.010503.

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 746–755, 2013. doi:10.1109/FOCS.2013.85.

[BS94] Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. In Tor Helleseth, editor, *Advances in Cryptology — EUROCRYPT '93*, pages 410–423, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg.

[BSD21] Bjarne Bergh, Robert Salzmann, and Nilanjana Datta. The $\alpha \to 1$ limit of the sharp quantum rényi divergence. *Journal of Mathematical Physics*, 62(9):092205, 2021. doi:10.1063/5.0049791.

[BSST99] Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83:3081–3084, Oct 1999. doi:10.1103/PhysRevLett.83.3081.

[BVY21] Mohammad Bavarian, Thomas Vidick, and Henry Yuen. Anchored parallel repetition for nonlocal games, 2021, arXiv:1509.07466. URL: `https://arxiv.org/abs/1509.07466`.

[BVY22] Mohammad Bavarian, Thomas Vidick, and Henry Yuen. Anchored parallel repetition for nonlocal games. *SIAM Journal on Computing*, 51(2):214–253, 2022, arXiv:https://doi.org/10.1137/21M1405927. doi:10.1137/21M1405927.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969. doi:10.1103/PhysRevLett.23.880.

[CLPK+23] Guillermo Currás-Lorenzo, Margarida Pereira, Go Kato, Marcos Curty, and Kiyoshi Tamaki. A security framework for quantum key distribution implementations, 2023, arXiv:2305.05930.

[CMH17] Matthias Christandl and Alexander Müller-Hermes. Relative entropy bounds on quantum, private and repeater capacities. *Communications in Mathematical Physics*, 353(2):821–852, 2017.

[CMH24] Matthias Christandl and Alexander Müller-Hermes. Fault-tolerant coding for quantum communication. *IEEE Transactions on Information Theory*, 70(1):282–317, 2024. doi:10.1109/TIT.2022.3169438.

[CMS02] N. J. Cerf, S. Massar, and S. Schneider. Multipartite classical and quantum secrecy monotones. *Physical Review A*, 66(4), Oct 2002. doi:10.1103/physreva.66.042309.

[CMW16] Tom Cooney, Milán Mosonyi, and Mark M. Wilde. Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication. *Communications in Mathematical Physics*, 344(3):797–829, May 2016. doi:10.1007/s00220-016-2645-4.

[CS14] André Chailloux and Giannicola Scarpa. Parallel repetition of entangled games with exponential decay via the superposed information cost. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming*, pages 296–307, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[CSUU08] Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum xor proof systems. *computational complexity*, 17(2):282–299, 2008. doi:10.1007/s00037-008-0250-4.

[CWY15] Kai-Min Chung, Xiaodi Wu, and Henry Yuen. Parallel repetition for entangled player games via fast quantum search. In *Proceedings of the 30th Conference on Computational Complexity*, CCC '15, page 512–536, Dagstuhl, DEU, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[DBWR14] Frédéric Dupuis, Mario Berta, Jürg Wullschleger, and Renato Renner. One-shot decoupling. *Communications in Mathematical Physics*, 328(1):251–284, 2014. doi:10.1007/s00220-014-1990-4.

[DF19] Frédéric Dupuis and Omar Fawzi. Entropy accumulation with improved second-order term. *IEEE Transactions on Information Theory*, 65(11):7596–7612, 2019. doi:10.1109/TIT.2019.2929564.

[DFR20] Frédéric Dupuis, Omar Fawzi, and Renato Renner. Entropy accumulation. *Communications in Mathematical Physics*, 379(3):867–913, 2020. doi:10.1007/s00220-020-03839-5.

[DSV14] Irit Dinur, David Steurer, and Thomas Vidick. A parallel repetition theorem for entangled projection games. In *Proceedings of the 2014 IEEE 29th Conference on Computational Complexity*, CCC '14, page 197–208, USA, 2014. IEEE Computer Society. doi:10.1109/CCC.2014.28.

[Eke91] Artur K. Ekert. Quantum cryptography based on bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991. doi:10.1103/PhysRevLett.67.661.

[FF21] Hamza Fawzi and Omar Fawzi. Defining quantum divergences via convex optimization. *Quantum*, 5:387, jan 2021. doi:10.22331/q-2021-01-26-387.

[FM18] Honghao Fu and Carl A. Miller. Local randomness: Examples and application. *Physical Review A*, 97(3), mar 2018. doi:10.1103/physreva.97.032324.

[fu23] fedja (`https://mathoverflow.net/users/1131/fedja`). Does approximate equality of quantum states imply operator inequality in a large subspace? MathOverflow `https://mathoverflow.net/q/451757`, 2023. (version: 2023-08-01).

[GLvH+22] Ian George, Jie Lin, Thomas van Himbeeck, Kun Fang, and Norbert Lütkenhaus. Finite-key analysis of quantum key distribution with characterized devices using entropy accumulation, 2022, arXiv:2203.06554. doi:10.48550/ARXIV.2203.06554.

[HML+23] Anqi Huang, Akihiro Mizutani, Hoi-Kwong Lo, Vadim Makarov, and Kiyoshi Tamaki. Characterization of state-preparation uncertainty in quantum key distribution. *Phys. Rev. Appl.*, 19:014048, Jan 2023. doi:10.1103/PhysRevApplied.19.014048.

[Hol98] A.S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998. doi:10.1109/18.651037.

[Hol09] Thomas Holenstein. Parallel repetition: Simplifications and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009. doi:10.4086/toc.2009.v005a008.

[Hor94] Ryszard Horodecki. Informationally coherent quantum systems. *Physics Letters A*, 187(2):145–150, 1994. doi:https://doi.org/10.1016/0375-9601(94)90052-3.

[HP91] Fumio Hiai and Dénes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, 143(1):99–114, 1991. doi:10.1007/BF02100287.

[IRS17] Raban Iten, Joseph M. Renes, and David Sutter. Pretty good measures in quantum information theory. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 3195–3199. IEEE Press, 2017. doi:10.1109/ISIT.2017.8007119.

[JK21] Rahul Jain and Srijita Kundu. A direct product theorem for one-way quantum communication. In *Proceedings of the 36th Computational Complexity Conference*, CCC '21, Dagstuhl, DEU, 2021. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2021.27.

[JK22] Rahul Jain and Srijita Kundu. A direct product theorem for quantum communication complexity with applications to device-independent qkd. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1285–1295, 2022. doi:10.1109/FOCS52979.2021.00125.

[JK23] Rahul Jain and Srijita Kundu. A direct product theorem for quantum communication complexity with applications to device-independent cryptography, 2023, arXiv:2106.04299.

[JMS20] R. Jain, C. A. Miller, and Y. Shi. Parallel device-independent quantum key distribution. *IEEE Transactions on Information Theory*, 66(9):5567–5584, 2020. doi:10.1109/TIT.2020.2986740.

[JN11] Rahul Jain and Ashwin Nayak. Short proofs of the quantum substate theorem, 2011, arXiv:1103.6067. doi:10.48550/ARXIV.1103.6067.

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 167–176, 2012. doi:10.1109/FOCS.2012.42.

[JPY14] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A parallel repetition theorem for entangled two-player one-round games under product distributions. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 209–216, 2014. doi:10.1109/CCC.2014.29.

[JRS02] R. Jain, J. Radhakrishnan, and P. Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 429–438, 2002. doi:10.1109/SFCS.2002.1181967.

[KGL+23] Florian Kanitschar, Ian George, Jie Lin, Twesh Upadhyaya, and Norbert Lütkenhaus. Finite-size security for discrete-modulated continuous-variable quantum key distribution protocols. *PRX Quantum*, 4:040306, Oct 2023. doi:10.1103/PRXQuantum.4.040306.

[KKM+08] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, Ben Toner, and Thomas Vidick. Entangled games are hard to approximate. *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 447–456, 2008.

[Koa09] M Koashi. Simple security proof of quantum key distribution based on complementarity. *New Journal of Physics*, 11(4):045018, apr 2009. doi:10.1088/1367-2630/11/4/045018.

[KRS09] Robert König, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, Sep 2009. doi:10.1109/tit.2009.2025545.

[KWW12] Robert Konig, Stephanie Wehner, and Jürg Wullschleger. Unconditional security from noisy quantum storage. *IEEE Transactions on Information Theory*, 58(3):1962–1984, 2012. doi:10.1109/TIT.2011.2177772.

[KZF19] Guo-Dong Kang, Qing-Ping Zhou, and Mao-Fa Fang. Measurement-device-independent quantum key distribution with uncharacterized coherent sources. *Quantum Information Processing*, 19(1):1, 2019.

[Led16] Felix Leditzky. *Relative entropies and their use in quantum information theory*. PhD thesis, 2016, arXiv:1611.08802.

[LKDW18] Felix Leditzky, Eneet Kaur, Nilanjana Datta, and Mark M. Wilde. Approaches for approximate additivity of the holevo information of quantum channels. *Physical Review A*, 97(1), jan 2018. doi:10.1103/physreva.97.012332.

[LMT20] Norbert Lütkenhaus, Ashutosh Marwah, and Dave Touchette. Erasable bit commitment from temporary quantum trust. *IEEE Journal on Selected Areas in Information Theory*, 1(2):536–554, 2020. doi:10.1109/JSAIT.2020.3017054.

[LWW+10] Lars Lydersen, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar, and Vadim Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature Photonics*, 4(10):686–689, 2010. doi:10.1038/nphoton.2010.214.

[MAG06] Ll. Masanes, A. Acin, and N. Gisin. General properties of nonsignaling theories. *Phys. Rev. A*, 73:012112, Jan 2006. doi:10.1103/PhysRevA.73.012112.

[Mat18] Keiji Matsumoto. A new quantum version of f-divergence. In Masanao Ozawa, Jeremy Butterfield, Hans Halvorson, Miklós Rédei, Yuichiro Kitajima, and Francesco Buscemi, editors, *Reality and Measurement in Algebraic Quantum Theory*, pages 229–273, Singapore, 2018. Springer Singapore.

[MD24a] Ashutosh Marwah and Frédéric Dupuis. Proving security of BB84 under source correlations, 2024, arXiv:2402.12346. URL: https://arxiv.org/abs/2402.12346.

[MD24b] Ashutosh Marwah and Frédéric Dupuis. Security for parallel DIQKD. *Manuscript in preparation*, 2024.

[MD24c] Ashutosh Marwah and Frédéric Dupuis. Smooth min-entropy lower bounds for approximation chains. *Communications in Mathematical Physics*, 405(9):211, 2024. doi:10.1007/s00220-024-05074-8.

[MD24d] Ashutosh Marwah and Frédéric Dupuis. Strong converse for classical channel capacity using parallel repetition. *Manuscript in preparation*, 2024.

[MD24e] Ashutosh Marwah and Frédéric Dupuis. Universal chain rules from entropic triangle inequalities, 2024, arXiv:2412.06723. URL: https://arxiv.org/abs/2412.06723.

[MFSR24] Tony Metger, Omar Fawzi, David Sutter, and Renato Renner. Generalised entropy accumulation. *Communications in Mathematical Physics*, 405(11):261, 2024. doi:10.1007/s00220-024-05121-4.

[MLDS+13] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, 2013. doi:10.1063/1.4838856.

[MR22] Tony Metger and Renato Renner. Security of quantum key distribution from generalised entropy accumulation, 2022, arXiv:2203.04993. doi:10.48550/ARXIV.2203.04993.

[MS17] Carl A. Miller and Yaoyun Shi. Randomness in nonlocal games between mistrustful players. *Quantum information & computation*, 17 7:595–610, 2017.

[MST+19] Akihiro Mizutani, Toshihiko Sasaki, Yuki Takeuchi, Kiyoshi Tamaki, and Masato Koashi. Quantum key distribution with simply characterized light sources. *npj Quantum Information*, 5(1):87, 2019.

[MY98] D. Mayers and A. Yao. Quantum cryptography with imperfect apparatus. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No.98CB36280)*, pages 503–509, 1998. doi:10.1109/SFCS.1998.743501.

[MY04] Dominic Mayers and Andrew Yao. Self testing quantum apparatus. *Quantum Info. Comput.*, 4(4):273–286, July 2004.

[MYS12] M McKague, T H Yang, and V Scarani. Robust self-testing of the singlet. *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, oct 2012. doi:10.1088/1751-8113/45/45/455304.

[PAB+09] Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, Apr 2009. doi:10.1088/1367-2630/11/4/045021.

[PCLN+22] Margarida Pereira, Guillermo Currás-Lorenzo, Álvaro Navarrete, Akihiro Mizutani, Go Kato, Marcos Curty, and Kiyoshi Tamaki. Modified BB84 quantum key distribution protocol robust to source imperfections, 2022. doi:10.48550/ARXIV.2210.11754.

[Pet86] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics*, 23(1):57–65, 1986. doi:https://doi.org/10.1016/0034-4877(86)90067-4.

[Pet88] Dénes Petz. A variational expression for the relative entropy. *Communications in Mathematical Physics*, 114(2):345–349, 1988. doi:10.1007/BF01225040.

[PR14] Christopher Portmann and Renato Renner. Cryptographic security of quantum key distribution, 2014, arXiv:1409.3525.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998, arXiv:https://doi.org/10.1137/S0097539795280895. doi:10.1137/S0097539795280895.

[Ren06] Renato Renner. *Security of Quantum Key Distribution*. PhD thesis, 2006, arXiv:quant-ph/0512258.

[RK05] Renato Renner and Robert König. Universally composable privacy amplification against quantum adversaries. In Joe Kilian, editor, *Theory of Cryptography*, pages 407–425, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[SBT17] David Sutter, Mario Berta, and Marco Tomamichel. Multivariate trace inequalities. *Communications in Mathematical Physics*, 352(1):37–58, 2017.

[Sch95] Benjamin Schumacher. Quantum coding. *Phys. Rev. A*, 51:2738–2747, Apr 1995. doi:10.1103/PhysRevA.51.2738.

[SFR21] David Sutter, Omar Fawzi, and Renato Renner. Bounds on lyapunov exponents via entropy accumulation. *IEEE Transactions on Information Theory*, 67(1):10–24, jan 2021. doi:10.1109/tit.2020.3026959.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

[SP00] Peter W. Shor and John Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85:441–444, Jul 2000. doi:10.1103/PhysRevLett.85.441.

[SRK+15] Shihan Sajeed, Igor Radchenko, Sarah Kaiser, Jean-Philippe Bourgoin, Anna Pappa, Laurent Monat, Matthieu Legré, and Vadim Makarov. Attacks exploiting deviation of mean photon number in quantum key distribution and coin tossing. *Phys. Rev. A*, 91:032326, Mar 2015. doi:10.1103/PhysRevA.91.032326.

[Sut18] David Sutter. Approximate quantum markov chains. In *Approximate Quantum Markov Chains*, pages 75–100. Springer International Publishing, 2018. doi:10.1007/978-3-319-78732-9_5.

[SW88] Stephen J. Summers and Reinhard Werner. Maximal violation of bell's inequalities for algebras of observables in tangent spacetime regions. *Annales de l'I.H.P. Physique théorique*, 49(2):215–243, 1988. URL: http://eudml.org/doc/76413.

[SW97] Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, Jul 1997. doi:10.1103/PhysRevA.56.131.

[Tan23] Ernest Y. Z. Tan. Robustness of implemented device-independent protocols against constrained leakage, 2023, arXiv:2302.13928.

[TCR09] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009. doi:10.1109/TIT.2009.2032797.

[TH13] Marco Tomamichel and Masahito Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59(11):7693–7710, 2013. doi:10.1109/TIT.2013.2276628.

[TL17] Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, July 2017. doi:10.22331/q-2017-07-14-14.

[TLGR12] Marco Tomamichel, Charles Ci Wen Lim, Nicolas Gisin, and Renato Renner. Tight finite-key analysis for quantum cryptography. *Nature Communications*, 3(1):634, 2012. doi:10.1038/ncomms1631.

[Tom12] Marco Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, 2012, arXiv:1203.2142. doi:10.48550/ARXIV.1203.2142.

[Tom16] Marco Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016. doi:10.1007/978-3-319-21891-5.

[TRSS10] Marco Tomamichel, Renato Renner, Christian Schaffner, and Adam Smith. Leftover hashing against quantum side information. In *IEEE International Symposium on Information Theory*, pages 2703 –2707, June 2010. doi:10.1109/ISIT.2010.5513652.

[VDTR13] Alexander Vitanov, Frédéric Dupuis, Marco Tomamichel, and Renato Renner. Chain rules for smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 59(5):2603–2612, 2013. doi:10.1109/TIT.2013.2238656.

[Vid17] Thomas Vidick. Parallel DIQKD from parallel repetition, 2017. doi:10.48550/ARXIV.1703.08508.

[Wat60] Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.*, 4(1):66–82, jan 1960. doi:10.1147/rd.41.0066.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. doi:10.1017/9781316848142.

[Wat20] John Watrous. Lecture 5: Min-relative entropy, conditional max-entropy, and hypothesis-testing relative entropy. https://cs.uwaterloo.ca/~watrous/QIT-notes/QIT-notes.05.pdf, 2020. Accessed: 2023-06-15.

[Wie83] Stephen Wiesner. Conjugate coding. *SIGACT News*, 15(1):78–88, jan 1983. doi:10.1145/1008908.1008920.

[Wil13] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013. doi:10.1017/CBO9781139525343.

[Win16] Andreas Winter. Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints. *Communications in Mathematical Physics*, 347(1):291–313, 2016. doi:10.1007/s00220-016-2609-8.

[WR12] Ligong Wang and Renato Renner. One-shot classical-quantum capacity and hypothesis testing. *Phys. Rev. Lett.*, 108:200501, May 2012. doi:10.1103/PhysRevLett.108.200501.

[WWY14] Mark M. Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and hadamard channels via a sandwiched Rényi relative entropy. *Communications in Mathematical Physics*, 331(2):593–622, 2014.

[Yue16] Henry Yuen. A parallel repetition theorem for all entangled games, 2016. doi:10.48550/ARXIV.1604.04340.

Appendix A

Appendices for Chapter 3

A.1. Entropic triangle inequalities cannot be improved much

In this section, we will construct a classical counterexample to show that it is not possible to improve Lemma 3.5 to get a result like

$$H_{\min}^{\epsilon'}(A|B)_\rho \geq H_{\min}^{\epsilon}(A|B)_\eta - O(D_{\max}^{\epsilon''}(\rho\|\eta)) \tag{A.1}$$

where $\epsilon, \epsilon' > 0$ and the constant in front of $D_{\max}^{\epsilon''}(\rho\|\eta)$ is independent of the dimensions $|A|$ and $|B|$.

Consider the probability distribution p_{AB} where B is chosen to be equal to 1 with probability $1-\epsilon$ and 0 with probability ϵ, and A is chosen to be a random n-bit string if $B = 1$ otherwise A is chosen to be the all 0 string. Let E be the event that $B = 0$. Then, we have

$$p_{AB|E} \leq \frac{1}{p(E)} p_{AB} = \frac{1}{\epsilon} p_{AB}$$

or equivalently $D_{\max}(p_{AB|E}\|p_{AB}) \leq \log \frac{1}{\epsilon}$. In this case, we have $H_{\min}^{\epsilon}(A|B)_p = n\log(2)$ (where we are smoothing in the trace distance) and $H_{\min}^{\epsilon'}(A|B)_{p_{|E}} = \log \frac{1}{1-\epsilon'} = O(1)$ (independent of n). If Eq. A.1, were true then we would have

$$\begin{aligned} n\log(2) - O\left(\log \frac{1}{\epsilon}\right) &\leq H_{\min}^{\epsilon}(A|B)_P - O(D_{\max}^{\epsilon''}(p_{AB|E}\|p_{AB})) \\ &\leq H_{\min}^{\epsilon'}(A|B)_{p_{|E}} = O(1) \end{aligned}$$

which would lead to a contradiction because n is a free parameter, and we can let $n \to \infty$.

The same example can be used to show that it is not possible to improve Corollary 3.7 to an equation of the form

$$H(A|B)_\rho \geq H(A|B)_\eta - O(D(\rho\|\eta)).$$

For $\rho = P_{|E}$ and $\eta = P$, such a bound would imply that

$$0 \geq (1-\epsilon)n\log(2) - \log\frac{1}{\epsilon}$$

which is not true for large n.

A.2. Bounds for $D_\alpha^\#$ of the form in Lemma 3.14 necessarily diverge in the limit $\alpha = 1$

Classically, we have the following bound for Rényi entropies.

Lemma A.1. *Suppose $\epsilon \in (0,1]$, $d \geq \epsilon^{1/2}$, and p and q are two distributions over an alphabet $\mathcal{X}$ such that $\frac{1}{2}\|p-q\|_1 \leq \epsilon$ and $D_{\max}(p\|q) \leq d < \infty$, for $\alpha > 1$ we have*

$$D_\alpha(p\|q) \leq \frac{1}{\alpha-1}\log\left((1+\sqrt{\epsilon})^{\alpha-1}(1-2\sqrt{\epsilon}) + e^{d(\alpha-1)+1}\sqrt{\epsilon}\right). \tag{A.2}$$

In the limit, $\alpha \to 1$, we get the bound

$$D(p\|q) \leq (1-2\sqrt{\epsilon})\log(1+\sqrt{\epsilon}) + 2\sqrt{\epsilon}d. \tag{A.3}$$

Proof. Classically, we have that the set $S := \{x \in \mathcal{X} : p(x) \leq (1+\sqrt{\epsilon})q(x)\}$ is such that $p(S) \geq 1-2\sqrt{\epsilon}$ using Lemma 3.8. Thus, for $\alpha > 1$ we have

$$\begin{aligned}
\sum_{x\in\mathcal{X}} p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} &= \sum_{x\in S} p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} + \sum_{x\notin S} p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} \\
&\leq \sum_{x\in S}(1+\sqrt{\epsilon})^{\alpha-1}p(x) + \sum_{x\notin S} e^{d(\alpha-1)}p(x) \\
&= (1+\sqrt{\epsilon})^{\alpha-1}p(S) + e^{d(\alpha-1)}p(S^c) \\
&\leq (1+\sqrt{\epsilon})^{\alpha-1}(1-2\sqrt{\epsilon}) + e^{d(\alpha-1)+1}\sqrt{\epsilon}
\end{aligned}$$

where in the second line we used the definition of set S and the fact that $D_{\max}(p\|q) \leq d$, in the last line we use the fact that since $d \geq \sqrt{\epsilon} \geq \log(1+\sqrt{\epsilon})$, the convex sum is maximised for the largest possible value of $p(S^c)$, which is $2\sqrt{\epsilon}$. The bound now follows. □

We observed in Sec. 3.4.2 that the bound in Lemma 3.14 for $D_\alpha^\#$ tends to ∞ as $\alpha \to 1$ for a fixed $\epsilon > 0$. One may wonder if a bound like Eq. A.3 exists for $\lim_{\alpha\to1} D_\alpha^\#(\rho\|\sigma) = \hat{D}(\rho\|\sigma)$ [BSD21]. We show in the following that such a bound is not possible.

Suppose, that for all $\epsilon \in [0,a)$ (a small neighborhood of 0), $1 \leq d < \infty$, states ρ and σ, which satisfy $\frac{1}{2}\left\|\rho - \sigma\right\|_1 \leq \epsilon$ and $\rho \leq e^d \sigma$, the following bound holds

$$\hat{D}(\rho\|\sigma) \leq f(\epsilon, d) \tag{A.4}$$

where $f(\epsilon, d)$ is such that $\lim_{\epsilon\to 0} f(\epsilon, d) = f(0,d) = 0$ for every $1 \leq d < \infty$. Note that the upper bound in Eq. A.3 is of this form. It is known that for pure states ρ, $\hat{D}(\rho\|\sigma) = D_{\max}(\rho\|\sigma)$. We will use this to construct a contradiction.

Lemma A.2. [(1)] *For a pure state $\rho = |\rho\rangle\langle\rho|$ and a state σ, we have*

$$\hat{D}(\rho\|\sigma) = D_{\max}(\rho\|\sigma) = \langle\rho|\sigma^{-1}|\rho\rangle .$$

Proof. First, we can evaluate $\hat{D}$ as

$$\begin{aligned}
\hat{D}(\rho\|\sigma) &= \operatorname{tr}\left(\rho \log\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right) \\
&= \operatorname{tr}\left(|\rho\rangle\langle\rho| \log\left(|\rho\rangle\langle\rho|\sigma^{-1}|\rho\rangle\langle\rho|\right)\right) \\
&= \operatorname{tr}\left(|\rho\rangle\langle\rho| \log(\langle\rho|\sigma^{-1}|\rho\rangle)|\rho\rangle\langle\rho|\right) \\
&= \log\langle\rho|\sigma^{-1}|\rho\rangle .
\end{aligned}$$

Next, we have that

$$\begin{aligned}
D_{\max}(\rho\|\sigma) &= \log\left\|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right\|_\infty \\
&= \log\left\|\sigma^{-\frac{1}{2}}|\rho\rangle\langle\rho|\sigma^{-\frac{1}{2}}\right\|_\infty \\
&= \log\operatorname{tr}\left(\sigma^{-\frac{1}{2}}|\rho\rangle\langle\rho|\sigma^{-\frac{1}{2}}\right) \\
&= \log\langle\rho|\sigma^{-1}|\rho\rangle .
\end{aligned}$$

□

To obtain a contradiction, let $\epsilon \in [0, a^2)$. Define the states

$$\begin{aligned}
\rho &:= |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\
\sigma'_\epsilon &:= (\sqrt{1-\epsilon}\,|0\rangle + \sqrt{\epsilon}\,|1\rangle)(\sqrt{1-\epsilon}\,|0\rangle + \sqrt{\epsilon}\,|1\rangle)^\dagger \\
&= \begin{pmatrix} 1-\epsilon & \sqrt{\epsilon(1-\epsilon)} \\ \sqrt{\epsilon(1-\epsilon)} & \epsilon \end{pmatrix} \\
\sigma_\epsilon &:= (1-\delta)\sigma'_\epsilon + \delta\rho
\end{aligned}$$

[(1)] This lemma was pointed out to us by Omar Fawzi.

$$= \begin{pmatrix} (1-\epsilon)(1-\delta)+\delta & (1-\delta)\sqrt{\epsilon(1-\epsilon)} \\ (1-\delta)\sqrt{\epsilon(1-\epsilon)} & (1-\delta)\epsilon \end{pmatrix}$$

where $\{|0\rangle, |1\rangle\}$ is the standard basis and $\delta \in (0,1)$ is a parameter, which will be chosen later. Observe that $F(\rho, \sigma_\epsilon) = \langle e_0, \sigma_\epsilon e_0 \rangle = 1 - \epsilon(1-\delta)$, which implies that $\frac{1}{2}\|\rho - \sigma_\epsilon\|_1 \leq \sqrt{\epsilon} \in [0,a)$. For these definitions, we have

$$\sigma_\epsilon^{-1} = \frac{1}{(1-\delta)\delta\epsilon} \begin{pmatrix} (1-\delta)\epsilon & -(1-\delta)\sqrt{\epsilon(1-\epsilon)} \\ -(1-\delta)\sqrt{\epsilon(1-\epsilon)} & (1-\epsilon)(1-\delta)+\delta \end{pmatrix}$$

which implies that $\hat{D}(\rho\|\sigma_\epsilon) = \log\frac{1}{\delta}$ using Lemma A.2. We can fix $\delta = \frac{1}{10}$. Note that $\hat{D}(\rho\|\sigma_\epsilon) > 0$ is independent of ϵ. Now observe that if the bound in Eq. A.4 were true, then as $\epsilon \to 0$, $\hat{D}(\rho\|\sigma_\epsilon) = \log(10) \to 0$, which leads us to a contradiction. Thus, we cannot have bounds of the form in Eq. A.4 (also see [**BACGPH22**]). Consequently, any kind of bound on $\hat{D}_\alpha$ or $D_\alpha^\#$ which results in a bound of the form in Eq. A.4 as $\alpha \to 1$, for example, the bound in Eq. A.2, is also not possible at least close to $\alpha = 1$.

It should be noted that the reason we can have bounds of the form in Lemma 3.14, despite the fact that no good bound on $\hat{D} = \lim_{\alpha\to 1} D_\alpha^\#$ can be produced is that $D_\alpha^\#$, unlike the conventional generalizations of the Rényi divergence, is **not monotone** in α [**FF21**, Remark 3.3](otherwise the above counterexample would also give a no-go argument for $D_\alpha^\#$).

A.3. Transforming lemmas for EAT from $\tilde{H}_\alpha^\downarrow$ to $\tilde{H}_\alpha^\uparrow$

We have to redo the lemmas used in [**DFR20**] using $\tilde{H}_\alpha^\uparrow$ because we were only able to prove the dimension bound we need $(\tilde{H}_\alpha^\uparrow(A|BC) \geq \tilde{H}_\alpha^\uparrow(A|B) - 2\log|C|)$ in terms of $\tilde{H}_\alpha^\uparrow$

Lemma A.3 ([**DFR20**, Lemma 3.1]). *For $\rho_{A_1A_2B}$ and σ_B be states and $\alpha \in (0,\infty)$, we have the chain rule*

$$\tilde{D}_\alpha(\rho_{A_1B}\|\mathbb{1}_{A_1}\otimes\sigma_B) - \tilde{D}_\alpha(\rho_{A_1A_2B}\|\mathbb{1}_{A_1A_2}\otimes\sigma_B) = \tilde{H}_\alpha^\downarrow(A_2|A_1B)_\nu \tag{A.5}$$

where the state $\nu_{A_1A_2B}$ is defined as

$$\nu_{A_1B} := \frac{\left(\rho_{A_1B}^{\frac{1}{2}}\sigma_B^{-\alpha'}\rho_{A_1B}^{\frac{1}{2}}\right)^\alpha}{\mathrm{tr}\left(\rho_{A_1B}^{\frac{1}{2}}\sigma_B^{-\alpha'}\rho_{A_1B}^{\frac{1}{2}}\right)^\alpha}$$

$$\nu_{A_1A_2B} := \nu_{A_1B}^{\frac{1}{2}}\rho_{A_2|A_1B}\nu_{A_1B}^{\frac{1}{2}}$$

and $\alpha' := \frac{\alpha-1}{\alpha}$.

Corollary A.4 (Chain rule for $\tilde{H}^{\downarrow}_{\alpha}$ [DFR20, Theorem 3.2])**.** *For $\alpha \in (0,\infty)$, a state $\rho_{A_1A_2B}$, we have the chain rule*

$$\tilde{H}^{\downarrow}_{\alpha}(A_1A_2|B)_{\rho} = \tilde{H}^{\downarrow}_{\alpha}(A_1|B)_{\rho} + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_{\nu} \tag{A.6}$$

where the state $\nu_{A_1A_2B}$ is defined as

$$\nu_{A_1B} := \frac{\left(\rho^{\frac{1}{2}}_{A_1B}\rho_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^{\alpha}}{\operatorname{tr}\left(\rho^{\frac{1}{2}}_{A_1B}\rho_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^{\alpha}}$$
$$\nu_{A_1A_2B} := \nu^{\frac{1}{2}}_{A_1B}\rho_{A_2|A_1B}\nu^{\frac{1}{2}}_{A_1B}$$

and $\alpha' := \frac{\alpha-1}{\alpha}$.

We can modify [DFR20, Theorem 3.2], which is in terms of $\tilde{H}^{\downarrow}_{\alpha}$, to the following, which is a chain rule in terms of $\tilde{H}^{\uparrow}_{\alpha}$. The chain rule in this Corollary was also observed in [DFR20].

Corollary A.5 (Chain rule for $\tilde{H}^{\uparrow}_{\alpha}$)**.** *For $\alpha \in (0,\infty)$, a state $\rho_{A_1A_2B}$ and for any state σ_B such that $\tilde{H}^{\uparrow}_{\alpha}(A_1|B)_{\rho} = -\tilde{D}_{\alpha}(\rho_{A_1B}\|\mathbb{1}_{A_1}\otimes\sigma_B)$, we have*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B)_{\rho} + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_{\nu} \tag{A.7}$$

where the state $\nu_{A_1A_2B}$ is defined as

$$\nu_{A_1B} := \frac{\left(\rho^{\frac{1}{2}}_{A_1B}\sigma_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^{\alpha}}{\operatorname{tr}\left(\rho^{\frac{1}{2}}_{A_1B}\sigma_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^{\alpha}}$$
$$\nu_{A_1A_2B} := \nu^{\frac{1}{2}}_{A_1B}\rho_{A_2|A_1B}\nu^{\frac{1}{2}}_{A_1B}$$

and $\alpha' := \frac{\alpha-1}{\alpha}$. For $\alpha \in (0,\infty)$, state $\rho_{A_1A_2B}$ and any state σ_B such that $\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_{\rho} = -\tilde{D}_{\alpha}(\rho_{A_1A_2B}\|\mathbb{1}_{A_1A_2}\otimes\sigma_B)$, we have

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_{\rho} \leq \tilde{H}^{\uparrow}_{\alpha}(A_1|B)_{\rho} + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_{\nu} \tag{A.8}$$

where the state $\nu_{A_1A_2B}$ is defined the same as above.

Proof. Let σ_B be a state such that $\tilde{H}^{\uparrow}_{\alpha}(A_1|B)_{\rho} = -\tilde{D}_{\alpha}(\rho_{A_1B}\|\mathbb{1}\otimes\sigma_B)$. Then, using Lemma A.3, we have

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_{\rho} &\geq -\tilde{D}_{\alpha}(\rho_{A_1A_2B}\|\mathbb{1}_{A_1A_2}\otimes\sigma_B)\\ &= -\tilde{D}_{\alpha}(\rho_{A_1B}\|\mathbb{1}_{A_1}\otimes\sigma_B) + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_{\nu}\\ &= \tilde{H}^{\uparrow}_{\alpha}(A_1|B)_{\rho} + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_{\nu}\end{aligned}$$

for $\nu_{A_1A_2B}$ defined as in the lemma. Similarly, if $\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_\rho = -\tilde{D}_\alpha(\rho_{A_1A_2B}\|\mathbb{1}_{A_1A_2}\otimes\sigma_B)$, then

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_\rho &= -\tilde{D}_\alpha(\rho_{A_1A_2B}\|\mathbb{1}_{A_1A_2}\otimes\sigma_B)\\ &= -\tilde{D}_\alpha(\rho_{A_1B}\|\mathbb{1}_{A_1}\otimes\sigma_B) + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_\nu\\ &\leq \tilde{H}^{\uparrow}_{\alpha}(A_1|B)_\rho + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_\nu\end{aligned}$$

for $\nu_{A_1A_2B}$ defined as in the lemma. □

We transform [DFR20, Theorem 3.3] to a statement about $\tilde{H}^{\uparrow}_{\alpha}$ in the following.

Lemma A.6. *Let $\alpha \in \left[\frac{1}{2},\infty\right)$ and $\rho_{A_1A_2B_1B_2}$ be a state which satisfies the Markov chain $A_1 \leftrightarrow B_1 \leftrightarrow B_2$. Then, we have*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B_1B_2)_\rho \geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1)_\rho + \inf_{\nu} \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B_1B_2)_\nu \tag{A.9}$$

where the infimum is taken over all states $\nu_{A_1A_2B_1B_2}$ such that $\nu_{A_2B_2|A_1B_1} = \rho_{A_2B_2|A_1B_1}$.

Proof. Since, ρ satisfies the Markov chain $A_1 \leftrightarrow B_1 \leftrightarrow B_2$, there exists a decomposition of the system B_1 as [Sut18, Theorem 5.4]

$$B_1 = \bigoplus_{j\in J} a_j \otimes c_j$$

such that

$$\rho_{A_1B_1B_2} = \bigoplus_{j\in J} p(j)\rho_{A_1a_j}\otimes\rho_{c_jB_2}. \tag{A.10}$$

Let $J' \subseteq J$ be the set $\{j \in J : p(j) > 0\}$. Note, that we can replace J by J' in the above equation.

We can define the CPTP recovery map $\mathcal{R}_{B_1\to B_1B_2}$ for $\rho_{A_1B_1B_2}$ as

$$\mathcal{R}_{B_1\to B_1B_2}(X) := \bigoplus_{j\in J} \mathrm{tr}_{c_j}\left(\Pi_{a_j}\otimes\Pi_{c_j}X\Pi_{a_j}\otimes\Pi_{c_j}\right)\otimes\rho_{c_jB_2} \tag{A.11}$$

where $\Pi_{a_j}\otimes\Pi_{c_j}$ is the projector on the subspace $a_j\otimes c_j$. This recovery channel satisfies

$$\mathcal{R}_{B_1\to B_1B_2}(\rho_{A_1B_1}) = \rho_{A_1B_1B_2}. \tag{A.12}$$

We can now show that the optimisation for the conditional entropy $\tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_\rho$ can be restricted to states of the form $\mathcal{R}_{B_1\to B_1B_2}(\sigma_{B_1})$. This follows as

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_\rho &= \sup_{\sigma_{B_1B_2}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2}\|\mathbb{1}_{A_1}\otimes\sigma_{B_1B_2})\\ &\leq \sup_{\sigma_{B_1B_2}} -\tilde{D}_\alpha(\mathcal{R}_{B_1\to B_1B_2}\circ\mathrm{tr}_{B_2}(\rho_{A_1B_1B_2})\|\mathcal{R}_{B_1\to B_1B_2}\circ\mathrm{tr}_{B_2}(\mathbb{1}_{A_1}\otimes\sigma_{B_1B_2}))\\ &= \sup_{\sigma_{B_1}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2}\|\mathbb{1}_{A_1}\otimes\mathcal{R}_{B_1\to B_1B_2}(\sigma_{B_1}))\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{\sigma_{B_1B_2}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2}\|\mathbb{1}_{A_1}\otimes\sigma_{B_1B_2}) \\
&= \tilde{H}^{\uparrow}_\alpha(A_1|B_1B_2)_\rho
\end{aligned}
$$

where the second line follows from the data processing inequality for $\tilde{D}_\alpha$ for $\alpha \geq \frac{1}{2}$, the supremum in the fourth line is over all states on the registers B_1B_2,and the last line simply follows from the definition of $\tilde{H}^{\uparrow}_\alpha(A_1|B_1B_2)_\rho$. As a result, it follows that

$$
\tilde{H}^{\uparrow}_\alpha(A_1|B_1B_2)_\rho = \sup_{\sigma_{B_1}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2}\|\mathbb{1}_{A_1}\otimes\mathcal{R}_{B_1\to B_1B_2}(\sigma_{B_1})) \tag{A.13}
$$

Let $\sigma_{B_1B_2} = \mathcal{R}_{B_1\to B_1B_2}(\eta_{B_1})$ be such that $\tilde{H}^{\uparrow}_\alpha(A_1|B_1B_2)_\rho = -\tilde{D}_\alpha(\rho_{A_1B_1B_2}\|\mathbb{1}_{A_1}\otimes\sigma_{B_1B_2})$. Using Corollary A.5, for this choice of $\sigma_{B_1B_2}$, we have that

$$
\tilde{H}^{\uparrow}_\alpha(A_1A_2|B_1B_2)_\rho \geq \tilde{H}^{\uparrow}_\alpha(A_1|B_1B_2)_\rho + \tilde{H}^{\downarrow}_\alpha(A_2|A_1B_1B_2)_\nu \tag{A.14}
$$

where the state $\nu_{A_1A_2B_1B_2}$ is defined as

$$
\begin{aligned}
\nu_{A_1B_1B_2} &:= \frac{\left(\rho^{\frac{1}{2}}_{A_1B_1B_2}\sigma^{-\alpha'}_{B_1B_2}\rho^{\frac{1}{2}}_{A_1B_1B_2}\right)^\alpha}{\operatorname{tr}\left(\rho^{\frac{1}{2}}_{A_1B_1B_2}\sigma^{-\alpha'}_{B_1B_2}\rho^{\frac{1}{2}}_{A_1B_1B_2}\right)^\alpha} \\
\nu_{A_1A_2B_1B_2} &:= \nu^{\frac{1}{2}}_{A_1B_1B_2}\rho_{A_2|A_1B_1B_2}\nu^{\frac{1}{2}}_{A_1B_1B_2}.
\end{aligned}
$$

We will now show that $\nu_{A_2B_2|A_1B_1} = \rho_{A_2B_2|A_1B_1}$. For this it is sufficient to show that

$$
\nu^{-\frac{1}{2}}_{A_1B_1}\nu^{\frac{1}{2}}_{A_1B_1B_2} = \rho^{-\frac{1}{2}}_{A_1B_1}\rho^{\frac{1}{2}}_{A_1B_1B_2}.
$$

We have that

$$
\begin{aligned}
\sigma_{B_1B_2} &= \mathcal{R}_{B_1\to B_1B_2}(\eta_{B_1}) \\
&= \bigoplus_{j\in J}\operatorname{tr}_{c_j}\left(\Pi_{a_j}\otimes\Pi_{c_j}\eta_{B_1}\Pi_{a_j}\otimes\Pi_{c_j}\right)\otimes\rho_{c_jB_2} \\
&= \bigoplus_{j\in J} q(j)\omega_{a_j}\otimes\rho_{c_jB_2}
\end{aligned}
$$

where we have defined the probability distribution $q(j) := \operatorname{tr}(\Pi_{a_j}\otimes\Pi_{c_j}\eta_{B_1})$ and states $\omega_{a_j} = \frac{1}{q(j)}\Pi_{a_j}\operatorname{tr}_{c_j}\left(\Pi_{c_j}\eta_{B_1}\Pi_{c_j}\right)\Pi_{a_j}$ for every $j\in J$.

Since $\tilde{D}_\alpha(\rho_{A_1B_1B_2}\|\mathbb{1}_{A_1}\otimes\sigma_{B_1B_2}) = -\tilde{H}^{\uparrow}_\alpha(A_1|B_1B_2)_\rho \leq \log|A_1| < \infty$, we have that

$$
\begin{aligned}
&\rho_{A_1B_1B_2} \ll \mathbb{1}_{A_1}\otimes\sigma_{B_1B_2} \\
\Rightarrow &\bigoplus_{j\in J'} p(j)\rho_{A_1a_j}\otimes\rho_{c_jB_2} \ll \mathbb{1}_{A_1}\otimes\bigoplus_{j\in J} q(j)\omega_{a_j}\otimes\rho_{c_jB_2} \\
\Rightarrow &\text{for every } j\in J' : \rho_{A_1a_j} \ll \mathbb{1}_{A_1}\otimes\omega_{a_j} \text{ and } q(j) > 0.
\end{aligned} \tag{A.15}
$$

This decomposition can be used to evaluate $\nu_{A_1B_1B_2}$ as follows

$$\begin{aligned}
\nu_{A_1B_1B_2} &= \frac{1}{N}\left(\rho^{\frac{1}{2}}_{A_1B_1B_2}\sigma^{-\alpha'}_{B_1B_2}\rho^{\frac{1}{2}}_{A_1B_1B_2}\right)^{\alpha} \\
&= \frac{1}{N}\left(\bigoplus_{j\in J'} p(j)^{\frac{1}{2}}\rho^{\frac{1}{2}}_{A_1a_j}\otimes\rho^{\frac{1}{2}}_{c_jB_2}\bigoplus_{j\in J} q(j)^{-\alpha'}\omega^{-\alpha'}_{a_j}\otimes\rho^{-\alpha'}_{c_jB_2}\bigoplus_{j\in J'} p(j)^{\frac{1}{2}}\rho^{\frac{1}{2}}_{A_1a_j}\otimes\rho^{\frac{1}{2}}_{c_jB_2}\right)^{\alpha} \\
&= \frac{1}{N}\left(\bigoplus_{j\in J'} p(j)q(j)^{-\alpha'}\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\otimes\rho^{1-\alpha'}_{c_jB_2}\right)^{\alpha} \\
&= \frac{1}{N}\bigoplus_{j\in J'} p(j)^{\alpha}q(j)^{1-\alpha}\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)^{\alpha}\otimes\rho_{c_jB_2}
\end{aligned}$$

for $N \coloneqq \operatorname{tr}\left(\rho^{\frac{1}{2}}_{A_1B_1B_2}\sigma^{-\alpha'}_{B_1B_2}\rho^{\frac{1}{2}}_{A_1B_1B_2}\right)^{\alpha}$. Further, we have

$$\begin{aligned}
&\nu^{-\frac{1}{2}}_{A_1B_1}\nu^{\frac{1}{2}}_{A_1B_1B_2} \\
&\quad= \frac{1}{N^{-\frac{1}{2}}}\bigoplus_{j\in J'} p(j)^{-\frac{\alpha}{2}}q(j)^{-\frac{1-\alpha}{2}}\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)^{-\frac{\alpha}{2}}\otimes\rho^{-\frac{1}{2}}_{c_j} \\
&\qquad\qquad\cdot\frac{1}{N^{\frac{1}{2}}}\bigoplus_{j\in J'} p(j)^{\frac{\alpha}{2}}q(j)^{\frac{1-\alpha}{2}}\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)^{\frac{\alpha}{2}}\otimes\rho^{\frac{1}{2}}_{c_jB_2} \\
&\quad= \bigoplus_{j\in J'}\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)^{0}\otimes\rho^{-\frac{1}{2}}_{c_j}\rho^{\frac{1}{2}}_{c_jB_2} \\
&\quad= \bigoplus_{j\in J'}\rho^{0}_{A_1a_j}\otimes\rho^{-\frac{1}{2}}_{c_j}\rho^{\frac{1}{2}}_{c_jB_2}
\end{aligned}$$

where in the last line we have used that the projector $\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)^{0}$ is equal to the projector $\rho^{0}_{A_1a_j}$ for every $j \in J'$ (here P^0 is the projector onto the image of positive semidefinite operator P). This can be seen since for every $j \in J'$ we first have

$$\operatorname{im}\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)\subseteq\operatorname{im}\left(\rho_{A_1a_j}\right). \tag{A.16}$$

Second, we have that Eq. A.15 above implies that $\omega^0_{a_j}\rho^0_{Aa_j} = \rho^0_{Aa_j}$ for every $j \in J'$. Now, for $j \in J'$ we have the following inequality

$$\begin{aligned}
\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right) &\geq m\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{0}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right) \\
&= m\rho_{A_1a_j}
\end{aligned}$$

where $m > 0$ is the minimum non-zero eigenvalue of $\omega^{-\alpha'}_{a_j}$. Finally, raising the above to the power of 0 (this action is operator monotone)

$$\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)^{0}\geq\rho^{0}_{A_1a_j}. \tag{A.17}$$

Eq. A.16 and A.17 together imply that for $j \in J'$

$$\left(\rho^{\frac{1}{2}}_{A_1a_j}\omega^{-\alpha'}_{a_j}\rho^{\frac{1}{2}}_{A_1a_j}\right)^0 = \rho^0_{A_1a_j}.$$

Finally, we have that

$$\begin{aligned}\rho^{-\frac{1}{2}}_{A_1B_1}\rho^{\frac{1}{2}}_{A_1B_1B_2} &= \bigoplus_{j\in J'} p(j)^{-\frac{1}{2}}\rho^{-\frac{1}{2}}_{A_1a_j}\otimes\rho^{-\frac{1}{2}}_{c_j}\bigoplus_{j\in J'} p(j)^{\frac{1}{2}}\rho^{\frac{1}{2}}_{A_1a_j}\otimes\rho^{\frac{1}{2}}_{c_jB_2}\\ &= \bigoplus_{j\in J'}\rho^0_{A_1a_j}\otimes\rho^{-\frac{1}{2}}_{c_j}\rho^{\frac{1}{2}}_{c_jB_2}.\end{aligned}$$

This proves that

$$\nu^{-\frac{1}{2}}_{A_1B_1}\nu^{\frac{1}{2}}_{A_1B_1B_2} = \rho^{-\frac{1}{2}}_{A_1B_1}\rho^{\frac{1}{2}}_{A_1B_1B_2} \tag{A.18}$$

and hence

$$\begin{aligned}\nu_{A_2B_2|A_1B_1} &= \nu^{-\frac{1}{2}}_{A_1B_1}\nu^{\frac{1}{2}}_{A_1B_1B_2}\nu_{A_2|A_1B_1B_2}\nu^{\frac{1}{2}}_{A_1B_1B_2}\nu^{-\frac{1}{2}}_{A_1B_1}\\ &= \rho^{-\frac{1}{2}}_{A_1B_1}\rho^{\frac{1}{2}}_{A_1B_1B_2}\rho_{A_2|A_1B_1B_2}\rho^{\frac{1}{2}}_{A_1B_1B_2}\rho^{-\frac{1}{2}}_{A_1B_1}\\ &= \rho_{A_2B_2|A_1B_1}\end{aligned}$$

where we have used the fact that $\nu_{A_2|A_1B_1B_2} = \rho_{A_2|A_1B_1B_2}$ and Eq. A.18. We can now modify Eq. A.14 to get

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B_1B_2)_\rho \geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_\rho + \inf_{\nu}\tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B_1B_2)_\nu$$

where the infimum is over states ν such that $\nu_{A_2B_2|A_1B_1} = \rho_{A_2B_2|A_1B_1}$. We can use the data processing inequality to get

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_\rho &= \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_{\mathcal{R}_{B_1\to B_1B_2}(\rho_{AB_1})}\\ &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1)_\rho.\end{aligned}$$

Together with the above inequality this proves the lemma. □

We will use the following modification of [**DFR20**, Corollary 3.5].

Corollary A.7. *Let $\mathcal{M}_{R\to A_2B_2}$ be a channel and $\rho_{A_1A_2B_1B_2} = \mathcal{M}(\rho'_{A_1B_1R})$ such that the Markov chain $A_1 \leftrightarrow B_1 \leftrightarrow B_2$ holds. Then, we have*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B_1B_2)_\rho \geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1)_\rho + \inf_{\omega}\tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B_1B_2)_{\mathcal{M}(\omega)} \tag{A.19}$$

where the infimum is taken over all states $\omega_{A_1B_1R}$. Moreover, if $\rho'_{A_1B_1R}$ is pure then we can restrict the optimisation to pure states.

Proof. The proof is the same as [DFR20, Corollary 3.5]. We include it here for the sake of completeness.

It is sufficient to show that for every state ν such that $\nu_{A_2B_2|A_1B_1} = \rho_{A_2B_2|A_1B_1}$, there exists an $\omega_{A_1B_1R}$ such that $\nu_{A_1A_2B_1B_2} = \mathcal{M}(\omega)$. For such a ν, we can define

$$\omega_{RA_1B_1} = \nu_{A_1B_1}^{\frac{1}{2}} \rho_{A_1B_1}^{-\frac{1}{2}} \rho'_{A_1B_1R} \rho_{A_1B_1}^{-\frac{1}{2}} \nu_{A_1B_1}^{\frac{1}{2}}$$

which can be seen to be a valid state and also satisfy $\nu_{A_1A_2B_1B_2} = \mathcal{M}(\omega)$. □

A.4. Dimension bounds for conditional Rényi entropies

Lemma A.8 (Dimension bound)**.** *For $\alpha \in [\frac{1}{2}, \infty]$, a state $\rho_{A_1A_2B}$, the following bounds hold for the sandwiched conditional entropies*

$$\tilde{H}_\alpha^\downarrow(A_1|B)_\rho - \log|A_2| \leq \tilde{H}_\alpha^\downarrow(A_1A_2|B)_\rho \leq \tilde{H}_\alpha^\downarrow(A_1|B)_\rho + \log|A_2|$$
$$\tilde{H}_\alpha^\uparrow(A_1|B)_\rho - \log|A_2| \leq \tilde{H}_\alpha^\uparrow(A_1A_2|B)_\rho \leq \tilde{H}_\alpha^\uparrow(A_1|B)_\rho + \log|A_2|.$$

For $\alpha \in [0,2]$ and a state $\rho_{A_1A_2B}$, the following bounds hold for the Petz conditional entropies

$$\bar{H}_\alpha^\downarrow(A_1A_2|B)_\rho \leq \bar{H}_\alpha^\downarrow(A_1|B)_\rho + \log|A_2|$$
$$\bar{H}_\alpha^\uparrow(A_1A_2|B)_\rho \leq \bar{H}_\alpha^\uparrow(A_1|B)_\rho + \log|A_2|.$$

Proof. For the sandwiched conditional entropies, we simply use the corresponding chain rules (Corollary A.4 or Corollary A.5) along with the fact that for all states ν, $\tilde{H}_\alpha^\downarrow(A_2|A_1B)_\nu \in [-\log|A_2|, \log|A_2|]$ [Tom16, Lemma 5.2].

For the Petz conditional entropies, we will make use of the Jensen's inequality for operators [Bha97, Theorem V.2.3]. Suppose, $\{|e_i\rangle\}_{i=1}^{|X|}$ is an orthogonal basis for the space X. Then, we have for a positive operator P_{XY} and $\alpha \in [0,1]$

$$\begin{aligned}
\operatorname{tr}_X P_{XY}^\alpha &= \sum_{i=1}^{|X|} \mathbb{1}_Y \otimes \langle e_i|_X P_{XY}^\alpha \mathbb{1}_Y \otimes |e_i\rangle_X \\
&\leq |X| \left(\sum_{i=1}^{|X|} \frac{1}{|X|} \mathbb{1}_Y \otimes \langle e_i|_X P_{XY} \mathbb{1}_Y \otimes |e_i\rangle_X \right)^\alpha \\
&= |X|^{1-\alpha} P_Y^\alpha
\end{aligned} \tag{A.20}$$

where in the second step we have used the operator Jensen's inequality with the operators $\left\{ \frac{1}{\sqrt{|X|}} \mathbb{1}_Y \otimes |e_i\rangle_X \right\}_{i=1}^{|X|}$ along with the fact that the map $X \mapsto X^\alpha$ is operator concave. For

$\alpha \in [1,2]$ and positive operator P_{XY}, we can use the same argument as above and the fact that $X \mapsto X^\alpha$ is operator convex in this regime and derive

$$\operatorname{tr}_X P_{XY}^\alpha \geq |X|^{1-\alpha} P_Y^\alpha. \tag{A.21}$$

To prove the dimension bound, observe that for a positive state σ_B and $\alpha \in [0,2]$, we have

$$\begin{aligned}
-\bar{D}_\alpha(\rho_{A_1A_2B} \| \mathbb{1}_{A_1A_2} \otimes \sigma_B) &= \frac{1}{1-\alpha} \log \operatorname{tr}\left(\rho_{A_1A_2B}^\alpha \sigma_B^{1-\alpha}\right) \\
&= \frac{1}{1-\alpha} \log \operatorname{tr}\left(\operatorname{tr}_{A_2}\left(\rho_{A_1A_2B}^\alpha\right) \sigma_B^{1-\alpha}\right) \\
&\leq \frac{1}{1-\alpha} \log \operatorname{tr}\left(|A_2|^{1-\alpha} \rho_{A_1B}^\alpha \sigma_B^{1-\alpha}\right) \\
&= -\bar{D}_\alpha(\rho_{A_1B} \| \mathbb{1}_{A_1} \otimes \sigma_B) + \log |A_2|.
\end{aligned}$$

We can now take a supremum over σ_B to prove the dimension bound for $\bar{H}_\alpha^\uparrow$ or choose $\sigma_B = \rho_B$ to prove the dimension bound for $\bar{H}_\alpha^\downarrow$. □

The following lemma was originally proven in [MLDS+13, Proposition 8]. We reproduce the proof argument here.

Lemma A.9. *For $\alpha \in [\frac{1}{2}, \infty]$, a state ρ_{ABC}, we have*

$$\tilde{H}_\alpha^\uparrow(A|BC)_\rho \geq \tilde{H}_\alpha^\uparrow(AC|B)_\rho - \log |C| \tag{A.22}$$

and for $\alpha \in [0, 2]$

$$\bar{H}_\alpha^\uparrow(A|BC)_\rho \geq \bar{H}_\alpha^\uparrow(AC|B)_\rho - \log |C| \tag{A.23}$$

Proof. By the definition of the sandwiched conditional entropy, we have

$$\begin{aligned}
\tilde{H}_\alpha^\uparrow(A|BC) &= \sup_{\eta_{BC} \in D(BC)} -\tilde{D}_\alpha(\rho_{ABC} \| \mathbb{1}_A \otimes \eta_{BC}) \\
&\geq \sup_{\eta_B \in D(B)} -\tilde{D}_\alpha\left(\rho_{ABC} \| \mathbb{1}_A \otimes \frac{\mathbb{1}_C}{|C|} \otimes \eta_B\right) \\
&= \sup_{\eta_B \in D(B)} -\tilde{D}_\alpha\left(\rho_{ABC} \| \mathbb{1}_{AC} \otimes \eta_B\right) - \log |C| \\
&= \tilde{H}_\alpha^\uparrow(AC|B) - \log |C|
\end{aligned}$$

where we simply restrict the supremum in the second line to states of the form $\eta_{BC} = \eta_B \otimes \frac{\mathbb{1}_C}{|C|}$ to derive the inequality. The same proof also works with $\bar{H}_\alpha^\uparrow$ entropy. □

The following lemma was originally proven in [Led16, Proposition 3.3.5].

Lemma A.10 (Dimension bound for conditioning register)**.** *For $\alpha \in [\frac{1}{2}, \infty]$ and a state ρ_{ABC} we have*

$$\tilde{H}^{\uparrow}_{\alpha}(A|BC)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\rho} - 2\log|C|. \tag{A.24}$$

Further, if the register C is classical, then we have

$$\tilde{H}^{\uparrow}_{\alpha}(A|BC)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\rho} - \log|C|. \tag{A.25}$$

Proof. This bound can be proven by combining Lemma A.8 and Lemma A.9. In the case that C is classical, we have the inequality $\tilde{H}^{\uparrow}_{\alpha}(AC|B)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\rho}$ [Tom16, Lemma 5.3]. □

A.5. Necessity for constraints on side information size for approximate AEP and EAT and its implication for approximate GEAT

It turns out that it is necessary to place some sort of bound on the size of the side information for an approximate entropy accumulation theorem of the form in Theorem 3.12. The following classical example demonstrates this. This example also demonstrates the necessity for a bound on the size of the side information in an approximate asymptotic equipartition of the form in Theorem 3.11.

Let there be n rounds. For $k \in [n]$, the map $\mathcal{M}_k : A_1^{k-1} \to A_k B_k C_k$. This map sets the variables as follows:

(1) Measure A_1^{k-1} in the standard basis.
(2) Let $A_k \in_R \{0,1\}$ be a randomly chosen bit.
(3) Let $C_k = 0$ with probability $\frac{\epsilon}{2}$ and $C_k = 1$ otherwise.
(4) In the case that $C_k = 1$, let $B_k \in_R \{0,1\}^n$ be a randomly chosen n-bit string. Otherwise, let $B_k = A_1^k R_k$, where R_k is an $(n-k)$ bit randomly chosen string from $\{0,1\}$.

Let $\mathcal{M}'_k$ be the map which always chooses B_k to be a random n-bit string. It is easy to see that in this case, we have $H_{\min}(A_1^n|B_1^n C_1^n)_{\mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(1)} = n\log(2)$ whereas $H^{\delta}_{\min}(A_1^n|B_1^n C_1^n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(1)} = O(1)$ even though for every $k \in [n]$, the maps $\mathcal{M}_k$ are ϵ–close in diamond norm distance to the maps $\mathcal{M}'_k$. This proves that a bound on the size of the side registers is indeed necessary for approximate entropy accumulation. We show these facts formally in the following.

Lemma A.11. *Suppose $\Phi : R \to A$ and $\Phi' : R \to A$ are two channels which take a register R and measure it in the standard basis and map the resulting classical register C to the classical register A. Then, for every $\rho_{RR'}$, we have*

$$\|\Phi(\rho_{RR'}) - \Phi'(\rho_{RR'})\|_1 \leq \left\|P^{\Phi}_{AC} - P^{\Phi'}_{AC}\right\|_1 \tag{A.26}$$

where P^{Φ}_{AC} and $P^{\Phi'}_{AC}$ are the classical distributions produced when the maps Φ and Φ' are applied to the state $\rho_{RR'}$ respectively.

Proof. Let $\{|c\rangle\langle c|\}_c$ represent the measurement in the standard basis. Since, both the channels first measure register R in the standard basis, they produce the state

$$\begin{aligned}\rho_{CR'} &= \sum_c |c\rangle\langle c|_C \otimes \mathrm{tr}_R\left(|c\rangle\langle c|_R \rho_{RR'}\right) \\ &= \sum_c p(c)\,|c\rangle\langle c|_C \otimes \rho_{R'|c}\end{aligned}$$

where we have defined $p(c) \coloneqq \mathrm{tr}\left(|c\rangle\langle c|_R \rho_R\right)$ and $\rho_{R'|c} \coloneqq \frac{1}{p(c)} \mathrm{tr}_R\left(|c\rangle\langle c|_R \rho_{RR'}\right)$. Now, the action of channel Φ on register C can be represented using the conditional probability distribution $p^{\Phi}_{A|C}$ and the action of channel Φ' on register C can be similarly represented using $p^{\Phi'}_{A|C}$. We can define the states

$$\begin{aligned}\rho^{\Phi}_{ACR'} &\coloneqq \sum_{ac} p^{\Phi}_{A|C}(a|c)p(c)\,|a,c\rangle\langle a,c| \otimes \rho_{R'|c} \\ \rho^{\Phi'}_{ACR'} &\coloneqq \sum_{ac} p^{\Phi'}_{A|C}(a|c)p(c)\,|a,c\rangle\langle a,c| \otimes \rho_{R'|c}.\end{aligned}$$

Note that $\mathrm{tr}_C\left(\rho^{\Phi}_{ACR'}\right) = \Phi(\rho_{RR'})$ and $\mathrm{tr}_C\left(\rho^{\Phi'}_{ACR'}\right) = \Phi'(\rho_{RR'})$. Further, we can view the R' register of $\rho^{\Phi}_{ACR'}$ and $\rho^{\Phi'}_{ACR'}$ as being created by a channel which measures the register C and outputs the state $\rho_{R'|c}$ in the register R'. Therefore, we have

$$\begin{aligned}\|\Phi(\rho_{RR'}) - \Phi'(\rho_{RR'})\|_1 &\leq \left\|\rho^{\Phi}_{ACR'} - \rho^{\Phi'}_{ACR'}\right\|_1 \\ &\leq \left\|\rho^{\Phi}_{AC} - \rho^{\Phi'}_{AC}\right\|_1 \\ &= \left\|P^{\Phi}_{AC} - P^{\Phi'}_{AC}\right\|_1.\end{aligned}$$

□

We can use the above lemma to evaluate the distance between the channels $\mathcal{M}_k$ and $\mathcal{M}'_k$. Using the above lemma, it is sufficient to suppose that the input of the channels are classical. We can suppose that the registers A_1^{k-1} are classical and distributed as $P_{A_1^{k-1}}$. Let $P_{A_1^k B_k C_k}$ be the output of $\mathcal{M}_k$ on this distribution and $Q_{A_1^k B_k C_k}$ be the output of applying $\mathcal{M}'_k$. Then, we have

$$\left\|P_{A_1^k B_k C_k} - Q_{A_1^k B_k C_k}\right\|_1 = \sum_{a_1^k, c_k} P(a_1^{k-1})P(a_k)P(c_k)\left\|P_{B_k|a_1^k, c_k} - Q_{B_k}\right\|_1$$

$$
\begin{aligned}
&= \sum_{a_1^k} P(a_1^{k-1})P(a_k)\left(\left(1-\frac{\epsilon}{2}\right)\left\|P_{B_k|a_1^k,c_k=1}-Q_{B_k}\right\|_1+\frac{\epsilon}{2}\left\|P_{B_k|a_1^k,c_k=0}-Q_{B_k}\right\|_1\right)\\
&\leq \sum_{a_1^k} P(a_1^{k-1})P(a_k)\epsilon\\
&= \epsilon
\end{aligned}
$$

where in the first line we have used the fact that A_k and C_k are chosen independently with the same distribution in both the maps and the fact that B_k is chosen independently in $\mathcal{M}'_k$, for the third line we have used the fact that B_k is independent and has the same distribution as Q_{B_k} when $c_k = 1$. Since, this is true for all input distributions, we have $\|\mathcal{M}_k - \mathcal{M}'_k\|_\diamond \leq \epsilon$.

Now, let $R_{A_1^n B_1^n C_1^n}$ be the probability distribution created when the maps $\mathcal{M}_k$ are applied sequentially n times and $S_{A_1^n B_1^n C_1^n}$ be the probability distribution created when the maps $\mathcal{M}'_k$ are applied sequentially n times. Since, B_k and C_k are independent of A_k in the distribution S, we have

$$H_{\min}(A_1^n|B_1^n C_1^n)_S = n\log(2).$$

We will show that $H^{\epsilon'}_{\min}(A_1^n|B_1^n C_1^n)_R = O(1)$ as long as $\epsilon' \leq \frac{1}{4}$. Let $l := \frac{2}{\epsilon}\log\frac{1}{\epsilon'}$. Let E be the event that there exists a $k > n-l$ such that $C_k = 0$. For our choice of l, we have $p(E) \geq 1-\epsilon'$.

Lemma A.12. *Let P_{AB} be a subnormalised probability distribution such that $A = f(B)$ for some function f (that is, $P(a,b) > 0$ only if $a = f(b)$). Then, $H^{\epsilon}_{\min}(A|B)_P \leq \log\frac{1}{\mathrm{tr}(P)-\sqrt{2\epsilon}}$.*

Proof. Let P'_{AB} be a distribution ϵ-close to P in purified distance. Then, it is $\sqrt{2\epsilon}$ close to P in trace distance. We have that

$$
\begin{aligned}
e^{-H_{\min}(A|B)_{P'}} &= P'_{\text{guess}}(A|B)\\
&\geq \sum_b P'_{AB}(f(b),b)\\
&\geq \sum_b P_{AB}(f(b),b) - \sqrt{2\epsilon}\\
&= \mathrm{tr}(P) - \sqrt{2\epsilon}
\end{aligned}
$$

which implies that $H_{\min}(A|B)_{P'} \leq \log\frac{1}{\mathrm{tr}(P)-\sqrt{2\epsilon}}$. Since, this is true for every distribution ϵ-close to P, it also holds for $H^{\epsilon}_{\min}(A|B)_P$. □

We then have that

$$
\begin{aligned}
H^{\epsilon'}_{\min}(A_1^n|B_1^n C_1^n)_R &\leq H^{\epsilon'}_{\min}(A_1^n|B_1^n C_1^n \wedge E)_R\\
&\leq H^{\epsilon'}_{\min}(A_1^{n-l}|B_1^n C_1^n \wedge E)_R + l\log(2)
\end{aligned}
$$

$$\leq \log \frac{1}{p(E) - \sqrt{2\epsilon'}} + l \log(2)$$
$$\leq \log \frac{1}{1 - \epsilon' - \sqrt{2\epsilon'}} + l \log(2)$$
$$\leq l \log(2) + \log 8/3 = O(1)$$

where in the first line we have used [TL17, Lemma 10] in the first line, dimension bound (can be proven using Lemma A.8) in the second line, Lemma A.12 in the third line and the fact that $p(E) \geq 1 - \epsilon'$.

Also, note that the example given here satisfies

$$\left\| P_{A_1^k B_1^k C_1^k} - P_{A_1^{k-1} B_1^{k-1} C_1^{k-1}} P_{A_k B_k C_k} \right\|_1 \leq \epsilon$$

for every k. This also proves that a bound on the size of the side information registers ($B_k C_k$ here), as we have in Theorem 3.11, is necessary for an approximate version of AEP.

Further, this example also rules out the possibility of a natural approximate extension to the generalised entropy accumulation theorem (GEAT) [MFSR24] where the maps $\mathcal{M}_k \approx_\epsilon \mathcal{M}'_k$ and the maps $\mathcal{M}'_k$ satisfy the non-signalling conditions because one can write the entropy accumulation scenario in the form of a generalised entropy accumulation scenario where Eve's information contains the side information $B_1^k E$ in each step. Thus, it would not be possible to prove a meaningful bound on the smooth min-entropy without some sort of bound on the information transferred between the adversary's register E_i and the register R_i.

A.6. Classical approximate EAT

We present a simple proof for the approximate entropy accumulation theorem for classical distributions. This result requires a much weaker assumption than Theorem 3.12.

Theorem A.13. *Let $p_{A_1^n B_1^n E}$ be a classical distribution such that for every $k \in [n]$, and a_1^{k-1}, b_1^{k-1} and e*

$$\left\| p_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} - q^{(k)}_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} \right\|_\infty \leq \epsilon \tag{A.27}$$

where $\|v\|_\infty := \max_i |v(i)|$ and the $q^{(k)}_{B_k | a_1^{k-1}, b_1^{k-1}, e} = q^{(k)}_{B_k | b_1^{k-1}, e}$ or equivalently $q^{(k)}$ satisfies the Markov chain $A_k \leftrightarrow B_1^{k-1} E \leftrightarrow B_k$. Also, let $|A_k| = |A|$, $|B_k| = |B|$ for every $k \in [n]$.

Then, for $\epsilon' \in (0,1)$ and $\alpha \in \left(1, 1 + \frac{1}{\log(1+2|A|)}\right)$, we have that

$$H^{\epsilon'}_{\min}(A_1^n | B_1^n E)_p \geq \sum_{k=1}^n \inf_q H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}}$$

$$- n(\alpha-1)\log^2(2|A|+1) - \frac{\alpha}{\alpha-1} n \log(1+\epsilon|A||B|) - \frac{g_0(\epsilon')}{\alpha-1}. \tag{A.28}$$

where $g_0(x) := -\log(1-\sqrt{1-x^2})$. *The infimums are taken over all possible input probability distributions.*

For $\alpha = 1+\sqrt{\epsilon}$ (assuming $\sqrt{\epsilon} \le 1 + \frac{1}{\log(1+2|A|)}$), and using $\alpha \le 2$ and $\log(1+x) \le x$ as long as $x \ge 0$, the above bound gives us

$$\begin{aligned} H_{\min}^{\epsilon'}(A_1^n|B_1^n E)_p \ge \sum_{k=1}^n \inf_q H(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} \\ - n\sqrt{\epsilon}\left(\log^2(2|A|+1) - 2|A||B|\right) - \frac{g_0(\epsilon')}{\alpha-1} \end{aligned} \tag{A.29}$$

Proof. For every $k \in [n]$, we modify $q^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E}$ to create the distributions $r^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E}$, which are defined as follows

(1) Choose a random variable C_k from $\{0,1\}$ with probabilities $\left(\frac{|A||B|\epsilon}{1+|A||B|\epsilon}, \frac{1}{1+|A||B|\epsilon}\right)$.

(2) If $C_k = 1$, then choose random variables A_k, B_k using $q^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E}$ else choose A_k, B_k randomly with probability $\frac{1}{|\mathcal{A}||\mathcal{B}|}$.

That is, we have

$$r^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} := \frac{1}{1+|A||B|\epsilon} q^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} + \frac{|A||B|\epsilon}{1+|A||B|\epsilon} u_{A_k B_k}$$

where $u_{A_k B_k}$ is the uniform distribution on the registers A_k and B_k.

For every k, a_1^{k-1}, b_1^{k-1}, and e, we have

$$\begin{aligned} &\left\| p_{A_k B_k|a_1^{k-1},b_1^{k-1},e} - q^{(k)}_{A_k B_k|a_1^{k-1},b_1^{k-1},e} \right\|_\infty \le \epsilon \\ \Rightarrow\ & p_{A_k B_k|a_1^{k-1},b_1^{k-1},e} \le q^{(k)}_{A_k B_k|a_1^{k-1},b_1^{k-1},e} + \epsilon \mathbb{1}_{A_k B_k} \\ \Rightarrow\ & p_{A_k B_k|a_1^{k-1},b_1^{k-1},e} \le q^{(k)}_{A_k B_k|a_1^{k-1},b_1^{k-1},e} + \epsilon|A||B| u_{A_k B_k} \\ \Rightarrow\ & p_{A_k B_k|a_1^{k-1},b_1^{k-1},e} \le (1+|A||B|\epsilon) r^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} \end{aligned}$$

Define the distribution

$$r_{A_1^n B_1^n E} = \prod_{k=1}^n r^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} p_E. \tag{A.30}$$

Note that for every k, a_1^{k-1}, b_1^k, and e, we have

$$\begin{aligned} r_{B_k|A_1^{k-1}B_1^{k-1}E}(b_k|a_1^{k-1}b_1^{k-1}e) &= \frac{1}{1+|A||B|\epsilon} q^{(k)}_{B_k|A_1^{k-1}B_1^{k-1}E}(b_k|a_1^{k-1}b_1^{k-1}e) + \frac{\epsilon}{1+|A||B|\epsilon} \\ &= \frac{1}{1+|A||B|\epsilon} q^{(k)}_{B_k|B_1^{k-1}E}(b_k|b_1^{k-1}e) + \frac{\epsilon}{1+|A||B|\epsilon}, \end{aligned}$$

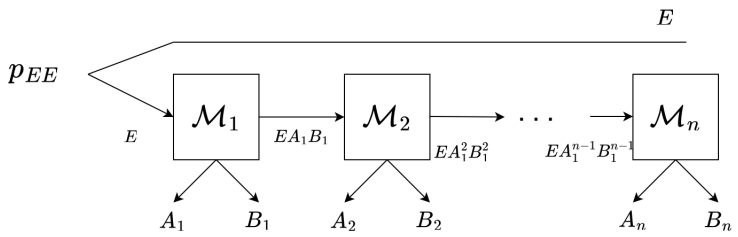

Fig. A.1. Setting for classical EAT

which implies

$$\begin{aligned} r_{B_k|B_1^{k-1}E}(b_k|b_1^{k-1}e) &= \sum_{\bar{a}_1^{k-1}} r_{A_1^{k-1}|B_1^{k-1}E}(\bar{a}_1^{k-1}|b_1^{k-1}e) r_{B_k|A_1^{k-1}B_1^{k-1}E}(b_k|\bar{a}_1^{k-1}b_1^{k-1}e) \\ &= \sum_{\bar{a}_1^{k-1}} r_{A_1^{k-1}|B_1^{k-1}E}(\bar{a}_1^{k-1}|b_1^{k-1}e) \left(\frac{1}{1+|A||B|\epsilon} q^{(k)}_{B_k|B_1^{k-1}E}(b_k|b_1^{k-1}e) + \frac{\epsilon}{1+|A||B|\epsilon} \right) \\ &= r_{B_k|A_1^{k-1}B_1^{k-1}E}(b_k|a_1^{k-1}b_1^{k-1}e). \end{aligned}$$

Thus, for every $k \in [n]$, r satisfies the Markov chain $A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$. Further, we have

$$\begin{aligned} p_{A_1^nB_1^nE}(a_1^n, b_1^n, e) &= \prod_{k=1}^n p_{A_kB_k|A_1^{k-1},B_1^{k-1},E}(a_k, b_k|a_1^{k-1}, b_1^{k-1}, e) p_E(e) \\ &\leq (1+\epsilon|A||B|)^n \prod_{k=1}^n r^{(k)}_{A_kB_k|A_1^{k-1},B_1^{k-1},E}(a_k, b_k|a_1^{k-1}, b_1^{k-1}, e) p_E(e) \\ &= (1+\epsilon|A||B|)^n r_{A_1^nB_1^nE}(a_1^n, b_1^n, e) \end{aligned}$$

which shows that $D_{\max}(p_{A_1^nB_1^nE} \| r_{A_1^nB_1^nE}) \leq n \log(1+\epsilon|A||B|)$.

The distribution $r_{A_1^nB_1^nE}$ can be viewed as the result of a series of maps as in Fig. A.1. We can now use the EAT chain rule [**DFR20**, Corollary 3.5] along with [**DFR20**, Lemma B.9] n-times to bound the entropy of this auxiliary distribution. We get

$$\begin{aligned} \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^nE)_r &\geq \sum_{k=1}^n \inf_{q_{A_1^{k-1}B_1^{k-1}E}} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_kA_1^{k-1}B_1^{k-1}E)_{r^{(k)}_{A_kB_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} \\ &\geq \sum_{k=1}^n \inf_{q_{A_1^{k-1}B_1^{k-1}E}} H(A_k|B_kA_1^{k-1}B_1^{k-1}E)_{r^{(k)}_{A_kB_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} - n(\alpha-1)\log^2(2|A|+1) \\ &\geq \sum_{k=1}^n \left(\inf_{q_{A_1^{k-1}B_1^{k-1}E}} \frac{1}{1+|A||B|\epsilon} H(A_k|B_kA_1^{k-1}B_1^{k-1}E)_{q^{(k)}_{A_kB_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} \right. \\ &\qquad \left. + \frac{\epsilon}{1+|A||B|\epsilon} \log|A| \right) - n(\alpha-1)\log^2(2|A|+1) \\ &\geq \sum_{k=1}^n \inf_{q_{A_1^{k-1}B_1^{k-1}E}} H(A_k|B_kA_1^{k-1}B_1^{k-1}E)_{q^{(k)}_{A_kB_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} - n(\alpha-1)\log^2(2|A|+1) \end{aligned}$$

for $\alpha \in \left(1, 1+\frac{1}{\log(1+2|A|)}\right)$. In the third line, we have used the concavity of the von Neumann entropy along with the definition of $r^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E}$. Using Lemma 3.5, we have

$$\begin{aligned} H^{\epsilon'}_{\min}(A_1^n|B_1^n E)_p &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_r - \frac{\alpha}{\alpha-1} D_{\max}(p_{A_1^n B_1^n E} \| r_{A_1^n B_1^n E}) - \frac{g_1(\epsilon', 0)}{\alpha-1} \\ &\geq \sum_{k=1}^{n} \inf_q H(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}} \\ &\quad - n(\alpha-1)\log^2(2|A|+1) - \frac{\alpha}{\alpha-1} n \log(1+\epsilon|A||B|) - \frac{g_0(\epsilon')}{\alpha-1}. \end{aligned}$$

□

A.7. Lemma to bound distance after conditioning

The following lemma relates the distance of two states conditioned on an event to the distance between them without conditioning.

Lemma A.14. *Suppose $\rho_{XA} = \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x| \otimes \rho_{A|x}$ and $\tilde{\rho}_{XA} = \sum_{x \in \mathcal{X}} \tilde{p}(x) |x\rangle\langle x| \otimes \tilde{\rho}_{A|x}$ are classical-quantum states such that $\frac{1}{2}\|\rho_{XA} - \tilde{\rho}_{XA}\|_1 \leq \epsilon$. Then, for $x \in \mathcal{X}$ such that $p(x) > 0$, we have*

$$\frac{1}{2}\left\|\rho_{A|x} - \tilde{\rho}_{A|x}\right\|_1 \leq \frac{2\epsilon}{p(x)} \tag{A.31}$$

Proof.

$$\frac{1}{2}\|\rho_{XA} - \tilde{\rho}_{XA}\|_1 = \frac{1}{2}\sum_{x \in \mathcal{X}} \left\|p(x)\rho_{A|x} - \tilde{p}(x)\tilde{\rho}_{A|x}\right\|_1 \leq \epsilon$$

This implies that for $x \in \mathcal{X}$

$$\frac{1}{2}\left\|p(x)\rho_{A|x} - \tilde{p}(x)\tilde{\rho}_{A|x}\right\|_1 \leq \epsilon$$

and

$$\frac{1}{2}|p(x) - \tilde{p}(x)| \leq \epsilon.$$

Using these inequalities, we have

$$\begin{aligned} \frac{1}{2}\left\|\rho_{A|x} - \tilde{\rho}_{A|x}\right\|_1 &\leq \frac{1}{2}\left\|\rho_{A|x} - \frac{\tilde{p}(x)}{p(x)}\tilde{\rho}_{A|x}\right\|_1 + \frac{1}{2}\left|1 - \frac{\tilde{p}(x)}{p(x)}\right| \left\|\tilde{\rho}_{A|x}\right\|_1 \\ &= \frac{1}{p(x)}\frac{1}{2}\left\|p(x)\rho_{A|x} - \tilde{p}(x)\tilde{\rho}_{A|x}\right\|_1 + \frac{1}{p(x)}\frac{1}{2}|p(x) - \tilde{p}(x)| \\ &\leq \frac{2\epsilon}{p(x)}. \end{aligned}$$

□

A.8. Proof of approximate EAT with testing

To prove Theorem 3.18, we follow [MFSR24], which is itself based on [DF19].

Proof of Theorem 3.18. Just as in the proof of Theorem 3.12, we define

$$\mathcal{M}_k^\delta := (1-\delta)\,\mathcal{M}'_k + \delta\,\mathcal{M}_k \tag{A.32}$$

for every k and the state

$$\sigma_{A_1^n B_1^n X_1^n E} := \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho^{(0)}_{R_0 E}). \tag{A.33}$$

so that for $\beta > 1$ and $\epsilon_1 > 0$, we have

$$D_{\max}^{\epsilon_1}(\rho_{A_1^n B_1^n X_1^n E} \| \sigma_{A_1^n B_1^n X_1^n E}) \leq n z_\beta(\epsilon, \delta) + \frac{g_0(\epsilon_1)}{\beta - 1}. \tag{A.34}$$

Define $d_\beta := n z_\beta(\epsilon, \delta) + \frac{g_0(\epsilon_1)}{\beta-1}$. The bound above implies that there exists a state $\tilde{\rho}_{A_1^n B_1^n X_1^n E}$, which is also classical on X_1^n such that

$$P\left(\rho_{A_1^n B_1^n X_1^n E}, \tilde{\rho}_{A_1^n B_1^n X_1^n E}\right) \leq \epsilon_1 \tag{A.35}$$

and

$$\tilde{\rho}_{A_1^n B_1^n X_1^n E} \leq e^{d_\beta} \sigma_{A_1^n B_1^n X_1^n E}. \tag{A.36}$$

The registers X_1^n for $\tilde{\rho}$ can be chosen to be classical, since the channel measuring X_1^n only decreases the distance between $\tilde{\rho}$ and ρ, and the new state produced would also satisfy Eq. A.36. As the registers X_1^n are classical for both σ and $\tilde{\rho}$, we can condition these states on the event Ω. We will call the probability of the event Ω for the state σ and $\tilde{\rho}$ $\Pr_\sigma(\Omega)$ and $\Pr_{\tilde{\rho}}(\Omega)$ respectively. Using Lemma A.14 and the Fuchs-van de Graaf inequality, we have

$$P\left(\rho_{A_1^n B_1^n X_1^n E|\Omega}, \tilde{\rho}_{A_1^n B_1^n X_1^n E|\Omega}\right) \leq 2\sqrt{\frac{\epsilon_1}{\Pr_\rho(\Omega)}}. \tag{A.37}$$

Conditioning Eq. A.36 on Ω, we get

$$\Pr_{\tilde{\rho}}(\Omega)\tilde{\rho}_{A_1^n B_1^n X_1^n E|\Omega} \leq e^{d_\beta} \Pr_\sigma(\Omega)\sigma_{A_1^n B_1^n X_1^n E|\Omega}. \tag{A.38}$$

Together, the above two equations imply that

$$D_{\max}^{\epsilon_2}(\rho_{A_1^n B_1^n X_1^n E|\Omega} \| \sigma_{A_1^n B_1^n X_1^n E|\Omega}) \leq n z_\beta(\epsilon, \delta) + \frac{g_0(\epsilon_1)}{\beta - 1} + \log \frac{\Pr_\sigma(\Omega)}{\Pr_{\tilde{\rho}}(\Omega)} \tag{A.39}$$

for $\epsilon_2 := 2\sqrt{\frac{\epsilon_1}{\Pr_\rho(\Omega)}}$.

For $\epsilon_3 > 0$ and $\alpha \in (1, 2)$, we can plug the above in the bound provided by Lemma 3.5 to get

$$H_{\min}^{\epsilon_2+\epsilon_3}(A_1^n|B_1^n E)_{\rho|\Omega} \geq \tilde{H}_\alpha^\uparrow(A_1^n|B_1^n E)_{\sigma|\Omega} - \frac{\alpha}{\alpha - 1} n z_\beta(\epsilon, \delta)$$

$$- \frac{1}{\alpha - 1} \left(\alpha \log \frac{\Pr_\sigma(\Omega)}{\Pr_{\tilde{\rho}}(\Omega)} + g_1(\epsilon_3, \epsilon_2) + \frac{\alpha g_0(\epsilon_1)}{\beta - 1} \right). \tag{A.40}$$

Now, note that using Eq. 3.65 and [DFR20, Lemma B.7] we have

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_{\sigma_{|\Omega}} = \tilde{H}^{\uparrow}_{\alpha}(A_1^n X_1^n|B_1^n E)_{\sigma_{|\Omega}}. \tag{A.41}$$

For every k, we introduce a register D_k of dimension $|D_k| = \lceil e^{\max(f) - \min(f)} \rceil$ and a channel $\mathcal{D}_k : X_k \to X_k D_k$ as

$$\mathcal{D}_k(\omega) := \sum_x \langle x|\omega|x\rangle \, |x\rangle\langle x| \otimes \tau_x \tag{A.42}$$

where for every x, the state τ_x is a mixture between a uniform distribution on $\{1, 2, \cdots, \lfloor e^{\max(f) - f(\delta_x)} \rfloor\}$ and a uniform distribution on $\{1, 2, \cdots, \lceil e^{\max(f) - f(\delta_x)} \rceil\}$, so that

$$H(D_k)_{\tau_x} = \max(f) - f(\delta_x) \tag{A.43}$$

where δ_x is the distribution with unit weight at element x.

Define the channels $\bar{\mathcal{M}}_k := \mathcal{D}_k \circ \mathcal{M}_k$, $\bar{\mathcal{M}}'_k := \mathcal{D}_k \circ \mathcal{M}'_k$ and $\bar{\mathcal{M}}^{\delta}_k := \mathcal{D}_k \circ \mathcal{M}^{\delta}_k = (1-\delta)\bar{\mathcal{M}}'_k + \delta \bar{\mathcal{M}}_k$ and the state

$$\bar{\sigma}_{A_1^n B_1^n X_1^n D_1^n E} := \bar{\mathcal{M}}^{\delta}_n \circ \cdots \bar{\mathcal{M}}^{\delta}_1(\rho^{(0)}_{R_0 E}) \tag{A.44}$$

Note that $\bar{\sigma}_{A_1^n B_1^n X_1^n E} = \sigma_{A_1^n B_1^n X_1^n E}$. [MFSR24, Lemma 4.5] implies that this satisfies

$$\begin{aligned} \tilde{H}^{\uparrow}_{\alpha}(A_1^n X_1^n|B_1^n E)_{\sigma_{|\Omega}} &= \tilde{H}^{\uparrow}_{\alpha}(A_1^n X_1^n|B_1^n E)_{\bar{\sigma}_{|\Omega}} \\ &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n X_1^n D_1^n|B_1^n E)_{\bar{\sigma}_{|\Omega}} - \max_{x_1^n \in \Omega} H_{\alpha}(D_1^n)_{\bar{\sigma}_{|x_1^n}} \end{aligned} \tag{A.45}$$

For $x_1^n \in \Omega$, we have

$$\begin{aligned} H_{\alpha}(D_1^n)_{\bar{\sigma}_{|x_1^n}} &\leq H(D_1^n)_{\bar{\sigma}_{|x_1^n}} \\ &\leq \sum_{k=1}^n H(D_k)_{\tau_{x_k}} \\ &= \sum_{k=1}^n \max(f) - f(\delta_{x_k}) \\ &= n \max(f) - n f(\mathrm{freq}(x_1^n)) \\ &\leq n \max(f) - nh. \end{aligned} \tag{A.46}$$

We can get rid of the conditioning on the right-hand side of Eq. A.45 by using [DFR20, Lemma B.5]

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^n X_1^n D_1^n|B_1^n E)_{\bar{\sigma}_{|\Omega}} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n X_1^n D_1^n|B_1^n E)_{\bar{\sigma}} - \frac{\alpha}{\alpha - 1} \log \frac{1}{\Pr_\sigma(\Omega)}. \tag{A.47}$$

We now show that the channels $\bar{\mathcal{M}}'_k$ satisfy the second condition in Theorem 3.12. For an arbitrary $k \in [n]$ and a sequence of channels $\mathcal{N}_i \in \{\bar{\mathcal{M}}_i, \bar{\mathcal{M}}'_i\}$ for every $1 \le i < k$, let

$$\eta_{A_1^k B_1^k X_1^k D_1^k E} = \bar{\mathcal{M}}'_k \circ \mathcal{N}_{k-1} \cdots \circ \mathcal{N}_1(\rho^{(0)}_{R_0 E}).$$

For this state, we have

$$\begin{aligned} I(A_1^{k-1}D_1^{k-1} : B_k|B_1^{k-1}E)_\eta &= I(A_1^{k-1} : B_k|B_1^{k-1}E)_\eta + I(D_1^{k-1} : B_k|A_1^{k-1}B_1^{k-1}E)_\eta \\ &= 0 \end{aligned}$$

where $I(A_1^{k-1} : B_k|B_1^{k-1}E)_\eta = 0$ because of the condition in Eq. 3.74, and $I(D_1^{k-1} : B_k|A_1^{k-1}B_1^{k-1}E)_\eta = 0$ since X_1^{k-1} and hence D_1^{k-1} are determined by $A_1^{k-1}B_1^{k-1}$. This implies that for this state $A_1^{k-1}D_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$. Thus, the maps $\bar{\mathcal{M}}_k$ and $\bar{\mathcal{M}}'_k$ satisfy the conditions required for applying Theorem 3.12. Specifically, we can use the bounds in Eq. 3.59 and 3.63 for bounding α-conditional Rényi entropy in Eq. A.47

$$\begin{aligned} &\tilde{H}^{\uparrow}_\alpha(A_1^n X_1^n D_1^n|B_1^n E)_{\bar{\sigma}} \\ &\quad \ge \tilde{H}^{\uparrow}_\alpha(A_1^n D_1^n|B_1^n E)_{\bar{\sigma}} \\ &\quad \ge \sum_{k=1}^{n} \inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} \tilde{H}^{\downarrow}_\alpha(A_k D_k|B_k \tilde{R}_{k-1})_{\bar{\mathcal{M}}'_k(\omega)} - \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(e^{\frac{\alpha-1}{\alpha}2\log(|A||D||B|)} - 1\right)\right). \end{aligned} \tag{A.48}$$

The analysis in the proof of [DF19, Proposition V.3] shows that the first term above can be bounded as

$$\begin{aligned} &\inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} \tilde{H}^{\downarrow}_\alpha(A_k D_k|B_k \tilde{R}_{k-1})_{\bar{\mathcal{M}}'_k(\omega)} \\ &\qquad \ge \mathrm{Max}(f) - \frac{(\alpha-1)\log(2)}{2}\left(\log(2|A|^2+1) + \log(2)\sqrt{2+\mathrm{Var}(f)}\right)^2 - (\alpha-1)^2 K_\alpha \end{aligned} \tag{A.49}$$

Combining Eq. A.45, A.46, A.47, A.48 and A.49, we have

$$\begin{aligned} \tilde{H}^{\uparrow}_\alpha(A_1^n X_1^n|B_1^n E)_{\bar{\sigma}_{|\Omega}} \ge nh &- \frac{(\alpha-1)\log(2)}{2}\left(\log(2|A|^2+1) + \log(2)\sqrt{2+\mathrm{Var}(f)}\right)^2 - n(\alpha-1)^2 K_\alpha \\ &- \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(e^{\frac{\alpha-1}{\alpha}2\log(|A||D||B|)} - 1\right)\right) - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr_\sigma(\Omega)}. \end{aligned} \tag{A.50}$$

Plugging this into Eq. A.40, we get

$$\begin{aligned} H_{\min}^{\epsilon_2+\epsilon_3}(A_1^n|B_1^n E)_{\rho_{|\Omega}} \ge nh &- \frac{(\alpha-1)\log(2)}{2}\left(\log(2|A|^2+1) + \log(2)\sqrt{2+\mathrm{Var}(f)}\right)^2 - n(\alpha-1)^2 K_\alpha \\ &- \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(e^{\frac{\alpha-1}{\alpha}2(\log(|A||B|)+\max(f)-\min(f)+1)} - 1\right)\right) \\ &- \frac{\alpha}{\alpha-1} n z_\beta(\epsilon,\delta) - \frac{1}{\alpha-1}\left(\alpha\log\frac{1}{\Pr_\rho(\Omega)-\epsilon_1} + g_1(\epsilon_3,\epsilon_2) + \frac{\alpha g_0(\epsilon_1)}{\beta-1}\right). \end{aligned} \tag{A.51}$$

where we have used $\Pr_{\tilde{\rho}}(\Omega) \geq \Pr_{\rho}(\Omega) - \epsilon_1$ since $\frac{1}{2}\left\|\rho - \tilde{\rho}\right\|_1 \leq P(\rho,\tilde{\rho}) \leq \epsilon_1$. Note that the probability of Ω under the auxiliary state σ cancels out. $\square$

Appendix B

Appendices for Chapter 4

B.1. Proof of Theorem 4.3

In this section, we formally prove the lower bound on the smooth min-entropy required for the security of QKD in Theorem 4.3 using the entropy accumulation theorem (EAT). In Section 4.3 (Eq. 4.22 and 4.23), we showed that $\rho'_{X_1^n\Theta_1^nA_1^n} := \bar{\bar{\rho}}_{X_1^n\Theta_1^nA_1^n|\Omega}$ and $\sigma_{X_1^n\Theta_1^nA_1^n} = \left(\hat{\rho}^{(\epsilon+\delta)}_{X\Theta A}\right)^{\otimes n}$ is such that

$$\frac{1}{2}\left\|\rho'_{X_1^n\Theta_1^nA_1^n} - \bar{\rho}_{X_1^n\Theta_1^nA_1^n|\Omega}\right\|_1 \leq \frac{\epsilon_f^2}{2} \tag{B.1}$$

and

$$D_{\max}(\rho'_{X_1^n\Theta_1^nA_1^n}\|\sigma_{X_1^n\Theta_1^nA_1^n}) \leq nh(\epsilon+\delta) + \log\frac{1}{\Pr_\rho(\Omega) - \epsilon_{\text{qu}}^\delta}. \tag{B.2}$$

Fix an arbitrary strategy for Eve. Let $\Phi_{\text{QKD}} : X_1^n\Theta_1^nA_1^n \to X_1^nY_1^n\hat{X}_S\hat{C}_1^n\Theta_1^n\hat{\Theta}_1^nSTE$ be the map applied by Alice, Bob and Eve on the states produced by Alice during the QKD protocol. In order to prove security for the BB84 protocol, we need a lower bound on the following smooth min-entropy of $\Phi_{\text{QKD}}(\bar{\rho})$

$$H_{\min}^\nu(X_S|ET\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Upsilon}}$$

for some $\nu \geq 0$. In [MR22, Appendix A], it is shown that it is sufficient to show a lower bound for the smooth min-entropy of the final state of the protocol conditioned on the event Υ'' when the protocol uses perfect source states. The arguments mentioned there are also valid for our case, which is why we bound the smooth min-entropy

$$H_{\min}^\nu(X_S|ET\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Upsilon''}}$$

in Theorem 4.3[(1)].

Using the data processing inequality and Eq. B.2, we see that

$$D_{\max}(\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})\|\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})) \leq nh(\epsilon+\delta) + \log \frac{1}{\Pr_\rho(\Omega) - \epsilon_{\mathrm{qu}}^\delta}. \tag{B.3}$$

Note that $\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})$ and $\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})$ are the states that are produced at the end of the protocol if Alice's source were to produce the states $\rho'_{X_1^n\Theta_1^nA_1^n}$ and $\sigma_{X_1^n\Theta_1^nA_1^n}$ respectively. The states $\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})$ and $\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})$ also contain all the corresponding classical variables as the real protocol state $\Phi_{\mathrm{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nA_1^n|\Omega})$. In particular, the event Υ'' is well-defined (defined using classical variables) for both of these states.

Using Lemma A.14 and Eq. B.1, we have that the final states conditioned on the event Υ'' satisfy

$$\frac{1}{2}\left\|\Phi_{\mathrm{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nA_1^n})_{|\Omega\wedge\Upsilon''} - \Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}\right\|_1 \leq \frac{\epsilon_f^2}{\Pr_{\bar{\rho}}(\Upsilon''|\Omega)} \tag{B.4}$$

where $\Pr_{\bar{\rho}}(\Upsilon''|\Omega)$ is the probability for the event Υ'' for the state $\Phi_{\mathrm{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nA_1^n|\Omega})$[(2)]. Using the Fuchs-van de Graaf inequality, we can transform this to a purified distance bound

$$P(\Phi_{\mathrm{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nA_1^n})_{|\Omega\wedge\Upsilon''}, \Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}) \leq \sqrt{\frac{2}{\Pr_{\bar{\rho}}(\Upsilon''|\Omega)}}\epsilon_f. \tag{B.5}$$

Let $d := D_{\max}(\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})\|\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n}))$. We have proven an upper bound on d in Eq. B.3. By definition of $D_{\max}$, we have

$$\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n}) \leq e^d \Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n}).$$

Conditioning both sides on the event Υ'' implies that

$$\Pr_{\rho'}(\Upsilon'')\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''} \leq e^d \Pr_{\sigma}(\Upsilon'')\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}$$

where $\Pr_{\rho'}(\Upsilon'')$ and $\Pr_\sigma(\Upsilon'')$ are the probability for Υ'' for the states $\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})$ and $\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})$ respectively. Therefore, we have

$$D_{\max}(\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}\|\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}) \leq d + \log\frac{\Pr_\sigma(\Upsilon'')}{\Pr_{\rho'}(\Upsilon'')}.$$

[(1)]The arguments in [MR22, Appendix A] can also be modified to show that it is sufficient to show that $\Pr(\Upsilon'')\left\|\rho^f_{K_AE'} - \tau_{K_A}\otimes\rho^f_{E'}\right\|_1$ is small, where K_A is Alice's key and ρ^f is the state produced at the end of the protocol conditioned on not aborting, to prove the security of QKD.

[(2)]We abuse notation while writing the probability this way since the state it is evaluated on is $\Phi_{\mathrm{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nQ_1^n|\Omega})$, while we simply use the subscripts $\bar{\rho}$ for P. We also write probabilities this way for the state $\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nQ_1^n})$ and $\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nQ_1^n})$. This is done for the sake of clarity.

Together, with Eq. B.5 for $\epsilon_{\text{pa}} := \left(\frac{2}{\Pr_{\bar{\rho}}(\Upsilon''|\Omega)}\right)^{\frac{1}{2}} \epsilon_f$, we have that

$$D_{\max}^{\epsilon_{\text{pa}}}(\Phi_{\text{QKD}}(\bar{\rho}_{X_1^n\Theta_1^n A_1^n})_{|\Omega\wedge\Upsilon''}\|\Phi_{\text{QKD}}(\sigma_{X_1^n\Theta_1^n A_1^n})_{|\Upsilon''}) \leq d + \log\frac{\Pr_\sigma(\Upsilon'')}{\Pr_{\rho'}(\Upsilon'')}. \tag{B.6}$$

Let $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ be arbitrary parameters. We have

$$\begin{aligned}
&H_{\min}^{\epsilon_{\text{pa}}+\epsilon_1+2(\epsilon_2+\epsilon_3)}(X_S|E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad = H_{\min}^{\epsilon_{\text{pa}}+\epsilon_1+2(\epsilon_2+\epsilon_3)}(\bar{X}_1^n|E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad \geq H_{\min}^{\epsilon_{\text{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - 3g_0(\epsilon_3) \\
&\quad \geq H_{\min}^{\epsilon_{\text{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - \log|T| - H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - 3g_0(\epsilon_3)
\end{aligned} \tag{B.7}$$

where in the first line we have used the fact that given Θ_1^n and $\hat{\Theta}_1^n$, one can figure out the set S and then $\bar{X}_1^n = X_S(\perp)_{S^c}$ (see Table 4.1 for definition of the registers), in the second line we have used the chain rule for smooth min-entropy [**VDTR13**, Theorem 15] and in the last line we have used the dimension bound. We have used the chain rule here to reduce our proof to bounding an entropy, which in the perfect source case, can be bound using entropy accumulation [**DFR20**, Section 5.1].

Now, we can use Lemma 3.5 to derive

$$\begin{aligned}
&H_{\min}^{\epsilon_{\text{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\qquad \geq \tilde{H}_\alpha^\uparrow(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\sigma)_{|\Upsilon''}} \\
&\qquad\qquad - \frac{\alpha}{\alpha-1}D_{\max}^{\epsilon_{\text{pa}}}(\Phi_{\text{QKD}}(\bar{\rho}_{X_1^n\Theta_1^n Q_1^n})_{|\Omega\wedge\Upsilon''}\|\Phi_{\text{QKD}}(\sigma_{X_1^n\Theta_1^n Q_1^n})_{|\Upsilon''}) - \frac{g_1(\epsilon_1,\epsilon_{\text{pa}})}{\alpha-1} \\
&\qquad \geq \tilde{H}_\alpha^\uparrow(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\sigma)_{|\Upsilon''}} \\
&\qquad\qquad - \frac{\alpha}{\alpha-1}d - \frac{\alpha}{\alpha-1}\log\frac{\Pr_\sigma(\Upsilon'')}{\Pr_{\rho'}(\Upsilon'')} - \frac{g_1(\epsilon_1,\epsilon_{\text{pa}})}{\alpha-1}
\end{aligned} \tag{B.8}$$

Thus, we have reduced the problem to lower bounding α-Rényi conditional entropy for the QKD protocol in Protocol 4.1, where Alice's source produces noisy BB84 states. We can bound this conditional entropy using the entropy accumulation theorem. The only difference in the following arguments from [**DFR20**, Section 5.1] is that we need to employ entropy accumulation for α-Rényi entropies (also see [**GLvH+22**]).

Firstly, note that we can use source purification for the state $\Phi_{\text{QKD}}(\sigma)$, that is, we can imagine that the state $\Phi_{\text{QKD}}(\sigma)$ was produced by the following procedure:

(1) Alice prepares n Bell states $(\Phi^+)_{\bar{A}_1^n A_1^n}^{\otimes n}$.

(2) For each $i \in [n]$, Alice measures the qubit $\bar{A}_i$ in the basis Θ_i, which is chosen to be Z with probability $(1-\mu)$ and otherwise is chosen to be X. The measurement result is labelled X_i.

(3) She then applies the $2(\epsilon+\delta)$–depolarising channel to each of the qubits A_i for $i \in [n]$ and sends them over the channel to Bob.

We can imagine that the source state is prepared in this fashion. The initial state for EAT will be represented by the registers $\bar{A}_1^n A_1^n E$, which contain the state produced after Eve forwards the state produced above by Alice to Bob. We can now define the EAT maps $\mathcal{M}_i : \bar{A}_i^n A_i^n \to \bar{A}_{i+1}^n A_{i+1}^n \bar{X}_i \bar{Y}_i \Theta_i \hat{\Theta}_i C_i$, where the registers Θ_i and $\hat{\Theta}_i$ are produced by randomly sampling according to the probabilities in the protocol, $\bar{X}_i$ and $\bar{Y}_i$ are produced according to the measurements chosen in the protocol and the source preparation procedure above, and C_i is defined as in Table 4.1.

Note that by conditioning on the event Υ'', we are requiring that $q = \mathrm{freq}(C_1^n)$ satisfies $q(1) \leq e\mu^2$. It is shown in [**DFR20**, Proof of Claim 5.2] that there exists an affine min-tradeoff function f, such that C_1^n given Υ'' satisfies $f(\mathrm{freq}(C_1^n)) \geq (1-2\mu+\mu^2)\log(2) - h(e)$. Using the entropy accumulation theorem [**DFR20**, Proposition 4.5], we get

$$\tilde{H}_\alpha^\uparrow(\bar{X}_1^n \bar{Y}_1^n | E \Theta_1^n \hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\sigma^\delta)_{|\Upsilon''}} \geq n((1-2\mu+\mu^2)\log(2) - h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr_\sigma(\Upsilon'')} \tag{B.9}$$

where $V := 2\lceil \|\nabla f\|_\infty \rceil + 2\log(1+2|\mathcal{X}|^2) = \frac{2}{\mu^2}\log\frac{1-e}{e} + 2\log(1+2|\mathcal{X}|^2)$[(3)] and $1 < \alpha < 1 + \frac{2}{V}$. Combining Eq. B.8 and B.9, we get

$$\begin{aligned}
&H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1}(\bar{X}_1^n \bar{Y}_1^n | E \Theta_1^n \hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad \geq n((1-2\mu+\mu^2)\log(2) - h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}d - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr_{\rho'}(\Upsilon'')} - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{\alpha-1} \\
&\quad \geq n((1-2\mu+\mu^2)\log(2) - h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}nh(\epsilon+\delta) - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr_\rho(\Omega) - \epsilon_{\mathrm{qu}}^\delta} \\
&\qquad - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr_{\bar{\rho}}(\Upsilon''|\Omega) - \frac{2\epsilon_{\mathrm{qu}}^\delta}{\Pr_\rho(\Omega)}} - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{\alpha-1} \\
&\quad \geq n((1-2\mu+\mu^2)\log(2) - h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}nh(\epsilon+\delta) \\
&\qquad - \frac{\alpha}{\alpha-1}\left(\log\frac{1}{\Pr_\rho(\Omega\wedge\Upsilon'') - 2\epsilon_{\mathrm{qu}}^\delta} + 1\right) - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{\alpha-1}
\end{aligned} \tag{B.10}$$

(3) It should be noted that this term can be improved using [**DF19**].

where we have used Eq. B.1, $\epsilon_f = 2\sqrt{\frac{\epsilon_{\text{qu}}^{\delta}}{\Pr_\rho(\Omega)}}$, $\Pr_\rho(\Omega \wedge \Upsilon'') = \Pr_\rho(\Omega)\Pr_{\bar{\rho}}(\Upsilon''|\Omega)$ and $\Pr_\rho(\Omega) \geq \Pr_\rho(\Omega \wedge \Upsilon'') > 2\epsilon_{\text{qu}}^{\delta}$ to simplify the result. It should be noted that the probability $\Pr_\sigma(\Upsilon'')$ of the auxiliary state cancels out. Since, we restrict ϵ and δ to the region, where $h(\epsilon+\delta) < \frac{1}{\sqrt{2}}$, we can choose

$$\alpha := 1 + \frac{2\sqrt{2h(\epsilon+\delta)}}{V} \tag{B.11}$$

which gives us the bound

$$\begin{aligned} H_{\min}^{\epsilon_{\text{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} &\geq n((1-2\mu+\mu^2)\log(2) - h(e) - V\sqrt{2h(\epsilon+\delta)}) \\ &- \frac{V}{\sqrt{2h(\epsilon+\delta)}}\left(\log\frac{1}{\Pr_\rho(\Omega\wedge\Upsilon'') - 2\epsilon_{\text{qu}}^{\delta}} + 1\right) - \frac{g_1(\epsilon_1,\epsilon_{\text{pa}})}{2\sqrt{2h(\epsilon+\delta)}}V. \end{aligned} \tag{B.12}$$

We also need to bound $H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\text{QKD}}(\rho)_{|\Omega\wedge\Upsilon''}}$ in Eq. B.7. The bound and the proof for this bound are the same as in [DFR20, Claim 5.2]. We have for $\beta \in (1,2)$ that

$$\begin{aligned} &H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\ &\quad\leq H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\ &\quad\leq \tilde{H}_{\frac{1}{\beta}}^{\downarrow}(\bar{Y}_1^n|\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} + \frac{g_0(\epsilon_2)}{\beta-1} \\ &\quad\leq \tilde{H}_{\frac{1}{\beta}}^{\downarrow}(\bar{Y}_1^n|\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})} + \frac{\beta}{\beta-1}\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + \frac{g_0(\epsilon_2)}{\beta-1} \\ &\quad= \frac{\beta}{\beta-1}\log\sum_{\theta_1^n,\hat{\theta}_1^n} P(\theta_1^n,\hat{\theta}_1^n)e^{\left(1-\frac{1}{\beta}\right)\tilde{H}_{\frac{1}{\beta}}^{\downarrow}(\bar{Y}_1^n|\theta_1^n,\hat{\theta}_1^n)} + \frac{\beta}{\beta-1}\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + \frac{g_0(\epsilon_2)}{\beta-1} \end{aligned}$$

where the first line follows from the data processing inequality for the smooth max-entropy, second line follows from Lemma 2.23, third line using [DFR20, Lemma B.6]. Let the random variable Z denote the number of $i \in [n]$, such that $\Theta_i = \hat{\Theta}_i = 1$. Then, we have the following inequalities for the first term in the bound above

$$\begin{aligned} \frac{\beta}{\beta-1}\log\sum_{\theta_1^n,\hat{\theta}_1^n} P(\theta_1^n,\hat{\theta}_1^n)e^{\left(1-\frac{1}{\beta}\right)\tilde{H}_{\frac{1}{\beta}}^{\downarrow}(\bar{A}_1^n|\theta_1^n,\hat{\theta}_1^n)} &\leq \frac{\beta}{\beta-1}\log\sum_{\theta_1^n,\hat{\theta}_1^n} P(\theta_1^n,\hat{\theta}_1^n)e^{\left(1-\frac{1}{\beta}\right)Z_{\theta_1^n,\hat{\theta}_1^n}} \\ &= \frac{\beta}{\beta-1}\log\sum_{z=0}^{n}\binom{n}{z}\mu^{2z}(1-\mu^2)^{n-z}e^{\left(1-\frac{1}{\beta}\right)z} \\ &= n\frac{\beta}{\beta-1}\log\left(1-\mu^2+e^{\left(1-\frac{1}{\beta}\right)}\mu^2\right) \\ &= n\mu^2\frac{\beta}{(\beta-1)}\left(e^{\left(1-\frac{1}{\beta}\right)}-1\right) \\ &\leq n\mu^2\left(1+\frac{(\beta-1)}{\beta}\right) \end{aligned}$$

where we use $Z_{\theta_1^n,\hat{\theta}_1^n}$ to denote the fact that the value of random variable Z is fixed by θ_1^n and $\hat{\theta}_1^n$, in the second line we transform the expectation over θ_1^n and $\hat{\theta}_1^n$ into an expectation over Z, in the third line we use the binomial theorem, in the fourth line we use the fact that $\ln(1+x) \leq x$ for all $x > -1$, and in the last line we use the fact that $e^x \leq 1+x+x^2$ for $x \in (0,1)$ and that for $\beta > 1$ the term $\left(1 - \frac{1}{\beta}\right)$ lies in this range. Thus, we get that for $\beta \in (1,2)$,

$$H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \leq n\mu^2 + \frac{(\beta-1)}{\beta}n\mu^2 + \frac{\beta}{\beta-1}\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + \frac{g_0(\epsilon_2)}{\beta-1}.$$

Choosing $\beta = 1 + \frac{1}{\sqrt{n}}$ and using the coarse bounds $1 < \beta < 2$, gives us

$$H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \leq n\mu^2 + \sqrt{n}\left(\mu^2 + 2\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + g_0(\epsilon_2)\right). \quad \text{(B.13)}$$

Combining Eq. B.7, B.12, and B.13, we get

$$\begin{aligned}
&H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1+2(\epsilon_2+\epsilon_3)}(X_S|E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad \geq n((1-2\mu)\log(2) - h(e) - \mu^2(1-\log(2)) - V\sqrt{2h(\epsilon+\delta)}) \\
&\quad - \sqrt{n}\left(\mu^2 + 2\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + g_0(\epsilon_2)\right) - \frac{V}{\sqrt{2h(\epsilon+\delta)}}\left(\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'') - 2\epsilon_{\mathrm{qu}}^{\delta}} + 1\right) \\
&\quad - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{2\sqrt{2h(\epsilon+\delta)}}V - \log|T| - 3g_0(\epsilon_3) \qquad \text{(B.14)}
\end{aligned}$$

where the parameters $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ are arbitrary, and

$$\epsilon_{\mathrm{pa}} = 2\left(\frac{2\epsilon_{\mathrm{qu}}^{\delta}}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')}\right)^{1/2}.$$

For an arbitrary $\epsilon' > 0$, we can set $\epsilon_1 = \frac{\epsilon'}{2}$ and $\epsilon_2 = \epsilon_3 = \frac{\epsilon'}{8}$ to derive the result in the theorem.

Appendix C

Appendices for Chapter 5

C.1. Additional lemmas

The following lemma provides a tighter bound for the distance as compared to Lemma 5.3. It is proven in [DBWR14, Lemma B.3]. We could replace the use of Lemma 5.3 in Theorem 5.8 and 5.10 with the following. This would slightly improve some constants. We use Lemma 5.3 instead for notational clarity, since we do not need to keep track of the additional unitary introduced in the following lemma.

Lemma C.1. *For a normalised state ρ_{AB} and a subnormalised state $\tilde{\rho}_{AB}$ such that $P(\rho_{AB}, \tilde{\rho}_{AB}) \leq \epsilon$, there exists a unitary U_B on the register B such that the state*

$$\eta_{AB} := \rho_B^{1/2} U_B \tilde{\rho}_B^{-1/2} \tilde{\rho}_{AB} \tilde{\rho}_B^{-1/2} U_B^\dagger \rho_B^{1/2} \tag{C.1}$$

($\tilde{\rho}_B^{-1/2}$ is the Moore-Penrose pseudo-inverse) satisfies $P(\rho_{AB}, \eta_{AB}) \leq 2\epsilon$. Note that if $\tilde{\rho}_B$ is full rank, then $\eta_B = \rho_B$.

Proof. Note that since ρ_{AB} is normalised, we have $F(\tilde{\rho}_{AB}, \rho_{AB}) \geq 1 - \epsilon^2$. Let $|\tilde{\rho}\rangle_{ABR}$ be an arbitrary purification of $\tilde{\rho}_{AB}$. Let U_B be any unitary for now and let η_{AB} be defined as in Eq. C.1 above. We will choose U_B so that $F(\tilde{\rho}_{AB}, \eta_{AB})$ is large.

Observe that the pure state $|\eta\rangle_{ABR} := \rho_B^{1/2} U_B \tilde{\rho}_B^{-1/2} |\tilde{\rho}\rangle_{ABR}$ is a purification of η_{AB}. Using Uhlmann's theorem [Wat18, Theorem 3.22], we have

$$\begin{aligned} F(\tilde{\rho}_{AB}, \eta_{AB}) &\geq |\langle \tilde{\rho} | \eta \rangle|^2 \\ &= \left| \langle \tilde{\rho} | \rho_B^{1/2} U_B \tilde{\rho}_B^{-1/2} | \tilde{\rho} \rangle \right|^2 \\ &= \left| \mathrm{tr}(\rho_B^{1/2} U_B \tilde{\rho}_B^{-1/2} \tilde{\rho}_{ABR}) \right|^2 \end{aligned}$$

$$= \left|\mathrm{tr}(U_B \tilde{\rho}_B^{1/2} \rho_B^{1/2})\right|^2 .$$

Say the polar decomposition of $\tilde{\rho}_B^{1/2}\rho_B^{1/2} = V_B \left|\tilde{\rho}_B^{1/2}\rho_B^{1/2}\right|$. We can now select U_B to be $V_B^\dagger$, so that

$$\begin{aligned} F(\tilde{\rho}_{AB}, \eta_{AB}) &\geq \left(\mathrm{tr}\left|\tilde{\rho}_B^{1/2}\rho_B^{1/2}\right|\right)^2 \\ &= F(\tilde{\rho}_B, \rho_B) \\ &\geq 1 - \epsilon^2 \end{aligned} \tag{C.2}$$

where we have used $F(\tilde{\rho}_B, \rho_B) \geq F(\tilde{\rho}_{AB}, \rho_{AB})$. Further, we have

$$\begin{aligned} P(\tilde{\rho}_{AB}, \eta_{AB}) &= \sqrt{1 - F_*(\tilde{\rho}_{AB}, \eta_{AB})} \\ &\leq \sqrt{1 - F(\tilde{\rho}_{AB}, \eta_{AB})} \\ &\leq \epsilon. \end{aligned}$$

Using the triangle inequality, for this choice of U_B, we get

$$\begin{aligned} P(\rho_{AB}, \eta_{AB}) &\leq P(\rho_{AB}, \tilde{\rho}_{AB}) + P(\tilde{\rho}_{AB}, \eta_{AB}) \\ &\leq 2\epsilon. \end{aligned}$$

□

C.2. Classical approximate chain rule for the relative entropy

Lemma C.2. *Let p and p' be probability distributions over $\mathcal{X}$. Then, for a function $f : \mathcal{X} \to \mathbb{R}$, we have*

$$|\mathbb{E}_{X\sim p}[f(X)] - \mathbb{E}_{X\sim p'}[f(X)]| \leq \max_{x\in\mathcal{X}} |f(x)| \, \|p - p'\|_1 . \tag{C.3}$$

Proof.

$$\begin{aligned} |\mathbb{E}_{X\sim p}[f(X)] - \mathbb{E}_{X\sim p'}[f(X)]| &= |\sum_{x\in\mathcal{X}} p(x)f(x) - \sum_{x\in\mathcal{X}} p'(x)f(x)| \\ &= |\sum_{x\in\mathcal{X}} (p(x) - p'(x))f(x)| \\ &\leq \sum_{x\in\mathcal{X}} |p(x) - p'(x)||f(x)| \\ &\leq \max_{x\in\mathcal{X}} |f(x)| \, \|p - p'\|_1 \end{aligned}$$

□

We will call the following lemma an approximate chain rule for the relative entropy. It allows us to switch the probability distribution p_{AB} in the term $D(p_{AB}||p_B q_{A|B})$ of the chain rule for $D(p_{AB}||q_{AB})$ to $D(p'_{AB}||p'_B q_{A|B})$, where $p \approx_\epsilon p'$ while incurring a penalty which depends only on the size of the register A.

Lemma C.3. *Suppose that $\delta > 0$ and q_{AB} is a probability distribution over $\mathcal{A}\times\mathcal{B}$ such that for all $a,b \in \mathcal{A}\times\mathcal{B}$, we have $q(a|b) \geq \delta$. Then, for $0 < \epsilon < \frac{1}{2}$ and a p'_{AB} such that $\frac{1}{2}\|p_{AB} - p'_{AB}\|_1 \leq \epsilon$, we have that*

$$D(p_{AB}||q_{AB}) \leq D(p_B||q_B) + D(p'_{AB}||p'_B q_{A|B}) + z(\epsilon,\delta) \tag{C.4}$$

where $z(\epsilon,\delta) := h(2\epsilon) + 6\epsilon\log\frac{1}{\delta} + 4\epsilon\log|\mathcal{A}|$. Equivalently, we have

$$D(p_{AB}||q_{AB}) \leq D(p_B||q_B) + \inf_{\|p'_{AB}-p_{AB}\|_1 \leq 2\epsilon} D(p'_{AB}||p'_B q_{A|B}) + z(\epsilon,\delta) \tag{C.5}$$

Proof. If $\mathrm{supp}(p_B) \nsubseteq \mathrm{supp}(q_B)$, then the right-hand side is infinite and the identity is trivially true. We suppose $\mathrm{supp}(p_B) \subseteq \mathrm{supp}(q_B)$ here on.

We will show that for every p'_{AB}, which is ϵ-close to p_{AB} the right-hand side in Eq. C.4 is greater than the left-hand side. Classically, we have the chain rule

$$D(p_{AB}||q_{AB}) = D(p_B||q_B) + \mathbb{E}_{B\sim p_B}[D(p_{A|B}||q_{A|B})]. \tag{C.6}$$

Note that both the above terms are finite ($q(a|b) \geq \delta$ is given). We will bound $\max_b D(p_{A|b}||q_{A|b})$ and then use Lemma C.2 to create a bound for the expectation in terms of p'. For a given $b \in \mathcal{B}$, we have

$$\begin{aligned}
|D(p_{A|b}||q_{A|b})| &= \left|\sum_a p(a|b)\log\frac{p(a|b)}{q(a|b)}\right| \\
&\leq \left|\sum_a p(a|b)\log p(a|b)\right| + \left|\sum_a p(a|b)\log\frac{1}{q(a|b)}\right| \\
&= H(A|B=b)_p + \sum_a p(a|b)\log\frac{1}{q(a|b)} \\
&\leq \log(|\mathcal{A}|) + \log\frac{1}{\delta}.
\end{aligned}$$

Now, using Lemma C.2,

$$\mathbb{E}_{B\sim p_B}[D(p_{A|B}||q_{A|B})] \leq \mathbb{E}_{B\sim p'_B}[D(p_{A|B}||q_{A|B})] + 2\epsilon\left(\log(|\mathcal{A}|) + \log\frac{1}{\delta}\right). \tag{C.7}$$

Finally, we need to change the $p_{A|B}$ in $D(p_{A|B}\|q_{A|B})$ to $p'_{A|B}$.

$$\begin{aligned}
D(p_{A|b}\|q_{A|b}) &= -H(A|B=b)_p + \mathbb{E}_{A\sim p_{A|b}}\left[\log\frac{1}{q(A|b)}\right] \\
&\leq -H(A|B=b)_{p'} + h\left(\frac{1}{2}\left\|p_{A|b}-p'_{A|b}\right\|_1\right) + \frac{1}{2}\left\|p_{A|b}-p'_{A|b}\right\|_1 \log(|\mathcal{A}|) \\
&\quad + \mathbb{E}_{A\sim p'_{A|b}}\left[\log\frac{1}{q(A|b)}\right] + \left\|p_{A|b}-p'_{A|b}\right\|_1 \log\frac{1}{\delta} \\
&= D(p'_{A|b}\|q_{A|b}) + h\left(\frac{1}{2}\left\|p_{A|b}-p'_{A|b}\right\|_1\right) + \frac{1}{2}\left\|p_{A|b}-p'_{A|b}\right\|_1 \left(2\log\frac{1}{\delta} + \log(|\mathcal{A}|)\right)
\end{aligned}$$

where we used the Fannes-Audenaert continuity bound [**Wil13**, Theorem 11.10.2] and Lemma C.2 in the second line. Taking the expectation over p'_B, we get

$$\begin{aligned}
&\mathbb{E}_{B\sim p'_B}[D(p_{A|B}\|q_{A|B})] \\
&\leq \mathbb{E}_{B\sim p'_B}\left[D(p'_{A|B}\|q_{A|B}) + h\left(\frac{1}{2}\left\|p_{A|B}-p'_{A|B}\right\|_1\right) + \frac{1}{2}\left\|p_{A|B}-p'_{A|B}\right\|_1 \left(2\log\frac{1}{\delta} + \log(|\mathcal{A}|)\right)\right] \\
&\leq \mathbb{E}_{B\sim p'_B}\left[D(p'_{A|B}\|q_{A|B})\right] + h\left(\frac{1}{2}\left\|p'_B p_{A|B}-p'_{AB}\right\|_1\right) + \frac{1}{2}\left\|p'_B p_{A|B}-p'_{AB}\right\|_1 \left(2\log\frac{1}{\delta} + \log|\mathcal{A}|\right) \\
&\leq \mathbb{E}_{B\sim p'_B}\left[D(p'_{A|B}\|q_{A|B})\right] + h(2\epsilon) + 2\epsilon\left(2\log\frac{1}{\delta} + \log|\mathcal{A}|\right) \\
&= D(p'_{AB}\|p'_B q_{A|B}) + h(2\epsilon) + 2\epsilon\left(2\log\frac{1}{\delta} + \log|\mathcal{A}|\right)
\end{aligned}$$

where we have used the fact that if $\|p_{AB}-p'_{AB}\|_1 \leq \epsilon$, then $\left\|p'_B p_{A|B}-p'_{AB}\right\|_1 \leq 2\epsilon$. This can be derived using the triangle inequality. Putting this in Eq. C.7, we get

$$\mathbb{E}_{B\sim p_B}[D(p_{A|B}\|q_{A|B})] \leq D(p'_{AB}\|p'_B q_{A|B}) + h(2\epsilon) + 6\epsilon\log\frac{1}{\delta} + 4\epsilon\log|\mathcal{A}|.$$

Therefore, using Eq. C.6, we get

$$D(p_{AB}\|q_{AB}) \leq D(p_B\|q_B) + D(p'_{AB}\|p'_B q_{A|B}) + h(2\epsilon) + 6\epsilon\log\frac{1}{\delta} + 4\epsilon\log|\mathcal{A}|.$$

□

C.3. Markov chain condition for all input states implies independence

In the following lemma, we show that if all outputs of a channel satisfy a certain Markov chain condition with the reference registers, then under some dimension constraints the output of the channel is independent of the input. Due to this fact, we choose to state the unstructured approximate EAT using the independence condition for the side information (Eq. 5.72). We expect that this lemma can be improved further.

Lemma C.4. *Let A, B and R be registers such that $|A| = |B|$ and $|R| = |A||B|$. Let $\mathcal{M} : R \to C$ be a channel such that for all input states $\rho^{(0)}_{ABR}$ the output $\rho_{ABC} = \mathcal{M}(\rho^{(0)})$ satisfies the Markov chain $A \leftrightarrow B \leftrightarrow C$. Then, we have that $\mathcal{M}(X_R) = \mathrm{tr}(X)\omega_C$ for some state ω_C.*

Proof. Since $|R| = |A||B|$, we can view R as the registers $A'B'$, where $A \equiv A$ and $B' \equiv B$. We can construct the Choi matrix of this channel as

$$J_{ABC} = \mathcal{M}\left(|\Phi\rangle\langle\Phi|_{ABA'B'}\right) \tag{C.8}$$

where $|\Phi\rangle$ is used to denote the unnormalised maximally entangled state, i.e., $|\Phi\rangle_{ABA'B'} \coloneqq \sum_{a,b} |ab\rangle_{AB} |ab\rangle_{A'B'} = |\Phi\rangle_{AA'} \otimes |\Phi\rangle_{BB'}$ and $\mathcal{M}$ is viewed as a channel from $A'B' \to C$. Since, all outputs of $\mathcal{M}$ satisfy the Markov chain $A \leftrightarrow B \leftrightarrow C$, we have that

$$\begin{aligned} J_{ABC} &= J_{AB} J_B^{-1} J_{BC} \\ &= \mathbb{1}_A \otimes \mathbb{1}_B \cdot |A|^{-1} \mathbb{1}_B \cdot \mathcal{M}(\mathbb{1}_{A'} \otimes |\Phi\rangle\langle\Phi|_{BB'}) \\ &= \mathbb{1}_A \otimes \mathcal{M}(\tau_{A'} \otimes |\Phi\rangle\langle\Phi|_{BB'}) \end{aligned} \tag{C.9}$$

where $\tau_{A'} = |A|^{-1}\mathbb{1}_{A'}$ is the maximally mixed state on A'. Let's define $\mathcal{N} : B' \to C$ as

$$\mathcal{N}(X_{B'}) = \mathcal{M}\left(\tau_{A'} \otimes X_{B'}\right). \tag{C.10}$$

Since, the Choi matrix is unique, we can see that

$$\mathcal{M}_{A'B' \to C} = \mathcal{N}_{B' \to C} \circ \mathrm{tr}_{A'}. \tag{C.11}$$

Let $W_{A'B'}$ be the swap unitary matrix, i.e., $W_{A'B'}|ab\rangle = |ba\rangle$. Then, note that for all input states $\rho_{ABA'B'}$, the output $\mathcal{M}(W_{A'B'}\rho_{ABA'B'}W^\dagger_{A'B'})$ satisfies the Markov chain $A \leftrightarrow B \leftrightarrow C$ according to the hypthosis in the lemma statement. In particular, we can carry out the above argument using the channel $\mathcal{M}(W_{A'B'} \cdot W^\dagger_{A'B'})$ and that gives us that there exists a channel $\mathcal{N}'_{B' \to C}$ such that for all operators $X_{A'B'}$

$$\mathcal{M}(W_{A'B'}X_{A'B'}W^\dagger_{A'B'}) = \mathcal{N}'_{B' \to C}\left(\mathrm{tr}_{A'}(X_{A'B'})\right) \tag{C.12}$$

which implies that

$$\begin{aligned} \mathcal{M}(X_{A'B'}) &= \mathcal{N}'_{B' \to C}\left(\mathrm{tr}_{A'}(W_{A'B'}X_{A'B'}W^\dagger_{A'B'})\right) \\ &= \mathcal{N}'_{A' \to C}\left(\mathrm{tr}_{B'}(X_{A'B'})\right) \end{aligned} \tag{C.13}$$

Using Eq. C.11 and C.13 for $X_{A'B'} = \sigma_{A'} \otimes \sigma_{B'}$ where $\sigma_{A'}$ and $\sigma_{B'}$ are arbitrary states, we have that

$$\mathcal{N}_{B' \to C}\left(\sigma_{B'}\right) = \mathcal{N}'_{A' \to C}\left(\sigma_{A'}\right) \tag{C.14}$$

which implies that both of these must be equal to a constant state. Let's call the state ω_C. Using the fact that $\mathcal{M}$ is trace-preserving, we then have that $\mathcal{M}_{A'B'\to C}(X_{A'B'}) = \mathrm{tr}(X)\omega_C$. □

C.4. Proof of unstructured approximate EAT with testing

We follow the proof in [MFSR24], which is itself based on [DF19].

Using the orthogonal projectors defined in Eq. 5.100, further define the orthogonal projectors on the registers $A_1^k B_1^k$

$$\Pi_{A_1^k B_1^k}^{(a_1^k, b_1^k)} := \bigotimes_{i=1}^{k} \left(\Pi_{A_i}^{(a_i)} \otimes \Pi_{B_i}^{(b_i)} \right) \tag{C.15}$$

for every a_1^k, b_1^k. Together these form a measurement on the registers $A_1^k B_1^k$.

We will also need the definitions of the following simple properties of the min-tradeoff functions for our entropy accumulation theorem:

$$\max(f) := \max_{q\in\mathbb{P}} f(q) \tag{C.16}$$

$$\min(f) := \min_{q\in\mathbb{P}} f(q). \tag{C.17}$$

For an affine min-tradeoff function, and any two distributions $q_1, q_2 \in \mathbb{P}$, we have

$$\begin{aligned} |f(q_1) - f(q_2)| &\leq \|\nabla f\|_\infty \|q_1 - q_2\|_1 \\ &\leq 2\|\nabla f\|_\infty \end{aligned}$$

which implies that $\max(f) - \min(f) \leq 2\|\nabla f\|_\infty$.

Proof. We first define the sets

$$\mathcal{A}_k := \left\{ X_{A_1^k B_1^k E} : X_{A_1^k B_1^k E} = \sum_{a_1^k, b_1^k} \Pi_{A_1^k B_1^k}^{(a_1^k b_1^k)} X_{A_1^k B_1^k E} \Pi_{A_1^k B_1^k}^{(a_1^k b_1^k)} \right\} \tag{C.18}$$

for every $0 \leq k \leq n$. It should be noted that each of these sets is a *unital algebra*, i.e., a vector space over the field $\mathbb{C}$ closed under the matrix product containing identity $\mathbb{1}_{A_1^k B_1^k E}$. It is also easy to see that

$$\mathcal{A}_k \otimes \mathbb{1}_{A_{k+1}B_{k+1}} := \left\{ X_{A_1^k B_1^k E} \otimes \mathbb{1}_{A_{k+1}B_{k+1}} : X_{A_1^k B_1^k E} \in \mathcal{A}_k \right\} \subseteq \mathcal{A}_{k+1} \tag{C.19}$$

$$\mathrm{tr}_{A_k B_k}(\mathcal{A}_k) := \left\{ X_{A_1^{k-1} B_1^{k-1} E} : X_{A_1^k B_1^k E} \in \mathcal{A}_k \right\} \subseteq \mathcal{A}_{k-1}. \tag{C.20}$$

Finally, it should be noted that for all operators $X_{A_1^k B_1^k E} \in \mathcal{A}_k$, we have

$$\mathrm{tr}_{X_1^k} \circ \mathcal{T}_k \circ \cdots \circ \mathcal{T}_1 (X) = X. \tag{C.21}$$

The above equation in particular implies that one can measure a state in $\mathcal{A}_k$ using the channel $\mathcal{T}_k \circ \cdots \circ \mathcal{T}_1$ without disturbing the state. All the auxiliary states defined during this proof will be contained inside an appropriate set $\mathcal{A}_k$. Noting that these sets are, in fact, an algebra will make it easier to see this. For example, observe that each partial state $\rho_{A_1^k B_1^k E} \in \mathcal{A}_k$.

Case 1: To begin, we restrict our attention to states $\rho_{A_1^n B_1^n E}$, which have full rank. Let $\nu \in (0,1)$ be an arbitrarily chosen small parameter. For every $k \in [n]$, define the states

$$\tilde{\tilde{\rho}}^{(k,0)}_{A_1^{k-1} B_1^{k-1} E R_k} := (1-\nu) \tilde{\rho}^{(k,0)}_{A_1^{k-1} B_1^{k-1} E R_k} + \nu \tau_{A_1^{k-1} B_1^{k-1} E R_k} \tag{C.22}$$

$$\tilde{\tilde{\rho}}^{(k,1)}_{A_1^{k-1} B_1^{k-1} X_1^{k-1} E R_k} := \mathcal{T}_{k-1} \circ \cdots \circ \mathcal{T}_1 (\tilde{\tilde{\rho}}^{(k,0)}_{A_1^{k-1} B_1^{k-1} E R_k}) \tag{C.23}$$

$$\tilde{\tilde{\rho}}^{(k)}_{A_1^k B_1^k X_1^k E} := \mathcal{T}_k \circ \mathcal{M}_k \left(\tilde{\tilde{\rho}}^{(k,1)}_{A_1^{k-1} B_1^{k-1} X_1^{k-1} E R_k} \right). \tag{C.24}$$

Since the maps $\mathrm{tr}_{X_i} \circ \mathcal{T}_i$ are unital, for every k, we have

$$\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1} B_1^{k-1} E} = \tilde{\tilde{\rho}}^{(k,1)}_{A_1^{k-1} B_1^{k-1} E} \tag{C.25}$$

$$\geq \nu \tau_{A_1^{k-1} B_1^{k-1} E}. \tag{C.26}$$

Thus, the states $\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1} B_1^{k-1} E}$ are also full rank. Moreover, we have that

$$\tilde{\tilde{\rho}}^{(k)}_{A_1^k B_1^k E} \in \mathcal{A}_k. \tag{C.27}$$

For each k these states satisfy

$$\begin{aligned}
\frac{1}{2} \left\| \rho_{A_1^k B_1^k X_1^k E} - \tilde{\tilde{\rho}}^{(k)}_{A_1^k B_1^k X_1^k E} \right\| &= \frac{1}{2} \left\| \rho_{A_1^k B_1^k X_1^k E} - \mathcal{T}_k \circ \mathcal{M}_k \left(\tilde{\tilde{\rho}}^{(k,1)}_{A_1^{k-1} B_1^{k-1} X_1^{k-1} E R_k} \right) \right\|_1 \\
&= \frac{1-\nu}{2} \left\| \mathcal{T}_k \circ \cdots \circ \mathcal{T}_1 (\rho_{A_1^k B_1^k E}) - \mathcal{T}_k \circ \cdots \circ \mathcal{T}_1 \circ \mathcal{M}_k \left(\tilde{\rho}^{(k,0)}_{A_1^{k-1} B_1^{k-1} E R_k} \right) \right\|_1 + \nu \\
&\leq \frac{1-\nu}{2} \left\| \rho_{A_1^k B_1^k E} - \mathcal{M}_k \left(\tilde{\rho}^{(k,0)}_{A_1^{k-1} B_1^{k-1} E R_k} \right) \right\|_1 + \nu \\
&\leq \epsilon + \nu.
\end{aligned} \tag{C.28}$$

Now, for each $k \in [n]$, we define the normalised states

$$\omega^{(k,1)}_{A_1^{k-1} B_1^{k-1} R_k E} := \rho^{1/2}_{A_1^{k-1} B_1^{k-1} E} \left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1} B_1^{k-1} E} \right)^{-1/2} \tilde{\tilde{\rho}}^{(k,1)}_{A_1^{k-1} B_1^{k-1} R_k E} \left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1} B_1^{k-1} E} \right)^{-1/2} \rho^{1/2}_{A_1^{k-1} B_1^{k-1} E} \tag{C.29}$$

and

$$\omega^{(k)}_{A_1^k B_1^k E} := \mathcal{M}_k \left(\omega^{(k,1)}_{A_1^{k-1} B_1^{k-1} R_k E} \right) \tag{C.30}$$

$$= \rho^{1/2}_{A_1^{k-1} B_1^{k-1} E} \left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1} B_1^{k-1} E} \right)^{-1/2} \mathrm{tr}_{X_k} \circ \mathcal{T}_k \circ \mathcal{M}_k \left(\tilde{\tilde{\rho}}^{(k,1)}_{A_1^{k-1} B_1^{k-1} R_k E} \right) \left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1} B_1^{k-1} E} \right)^{-1/2} \rho^{1/2}_{A_1^{k-1} B_1^{k-1} E} \tag{C.31}$$

$$= \rho^{1/2}_{A_1^{k-1}B_1^{k-1}E} \left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}\right)^{-1/2} \tilde{\tilde{\rho}}^{(k)}_{A_1^{k}B_1^{k}E} \left(\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}\right)^{-1/2} \rho^{1/2}_{A_1^{k-1}B_1^{k-1}E}. \tag{C.32}$$

We used $\mathrm{tr}_{X_k} \circ \mathcal{T}_k \circ \mathcal{M}_k = \mathcal{M}_k$ above. Observe that $\omega^{(k)}_{A_1^k B_1^k E} \in \mathcal{A}_k$ (using Eq. C.27, C.19 and C.20). Since, we defined $\tilde{\tilde{\rho}}^{(k)}_{A_1^{k-1}B_1^{k-1}E}$ to be full rank, we have that

$$\omega^{(k)}_{A_1^{k-1}B_1^{k-1}E} = \rho_{A_1^{k-1}B_1^{k-1}E}. \tag{C.33}$$

Using Lemma 5.3, we have that

$$\begin{aligned}\frac{1}{2}\left\|\rho_{A_1^k B_1^k E} - \omega^{(k)}_{A_1^k B_1^k E}\right\|_1 &\leq (\sqrt{2}+1) P(\rho_{A_1^k B_1^k E}, \tilde{\tilde{\rho}}_{A_1^k B_1^k E}) \\ &\leq (\sqrt{2}+1)\sqrt{2(\epsilon+\nu)} \\ &\leq 4\sqrt{\epsilon+\nu}. \end{aligned} \tag{C.34}$$

Let $\delta \in (0,1)$ be a small parameter (to be set equal to ϵ later). Define the states

$$\rho^{(k,\delta)}_{A_k B_k} := (1-\delta)\rho_{A_k B_k} + \delta \tau_{A_k B_k}. \tag{C.35}$$

Finally, for every $k \in [n]$, we define the states

$$\bar{\rho}^{(k)}_{A_1^k B_1^k E} := (1-\delta)\omega^{(k)}_{A_1^k B_1^k E} + \delta \rho^{(k,\delta)}_{A_k B_k} \otimes \rho_{A_1^{k-1}B_1^{k-1}E} \tag{C.36}$$

$$= (1-\delta)\omega^{(k)}_{A_1^k B_1^k E} + \delta \rho^{(k,\delta)}_{A_k B_k} \otimes \omega^{(k)}_{A_1^{k-1}B_1^{k-1}E} \tag{C.37}$$

Also, define $\bar{\rho}^{(0)}_E := \rho_E$. For each $0 \leq k \leq n$, the state $\bar{\rho}^{(k)}_{A_1^k B_1^k E} \in \mathcal{A}_k$ using Eq. C.19. We have taken a slight diversion from the proof of Theorem 5.8 in defining the above state. This has been done to ensure we are able to bound the entropy in Eq. C.72.

Let $\Delta_k : R_k \to A_k B_k$ be the map which traces out the register R_k and outputs $\rho^{(k,\delta)}_{A_k B_k}$. Further, let

$$\mathcal{M}^{\delta}_k := (1-\delta)\,\mathcal{M}_k + \delta \Delta_k. \tag{C.38}$$

Note that similar to $\mathcal{M}_k$, $\mathcal{M}^{\delta}_k$ also satisfies

$$\mathcal{M}^{\delta}_k = \mathrm{tr}_{X_k} \circ \mathcal{T}_k \circ \mathcal{M}^{\delta}_k. \tag{C.39}$$

Then, we have that

$$\begin{aligned}\bar{\rho}^{(k)}_{A_1^k B_1^k E} &= (1-\delta)\omega^{(k)}_{A_1^k B_1^k E} + \delta \rho^{(k,\delta)}_{A_k B_k} \otimes \omega^{(k)}_{A_1^{k-1}B_1^{k-1}E} \\ &= \left((1-\delta)\,\mathcal{M}_k + \delta\Delta_k\right)\left(\omega^{(k,1)}_{A_1^{k-1}B_1^{k-1}R_k E}\right) \\ &= \mathcal{M}^{\delta}_k\left(\omega^{(k,1)}_{A_1^{k-1}B_1^{k-1}R_k E}\right). \end{aligned} \tag{C.40}$$

We also have that

$$\frac{1}{2}\left\|\bar{\rho}^{(k)}_{A_1^k B_1^k E} - \rho_{A_1^k B_1^k E}\right\|_1 \leq 4\sqrt{\epsilon+\nu} + \delta \tag{C.41}$$

and by the definition of $\bar{\rho}^{(k)}_{A_1^k B_1^k E}$

$$\delta^2 \tau_{A_k B_k} \otimes \rho_{A_1^{k-1} B_1^{k-1} E} \leq \bar{\rho}^{(k)}_{A_1^k B_1^k E} \tag{C.42}$$

Using Corollary 5.6, gives us that

$$D(\rho_{A_1^k B_1^k E} \| \bar{\rho}^{(k)}_{A_1^k B_1^k E}) \leq \frac{8\sqrt{\epsilon+\nu}+2\delta}{1-\delta^2/(|A||B|)^2} \log \frac{|A||B|}{\delta}. \tag{C.43}$$

We define the above bound as $z(\epsilon+\nu,\delta)$. Using Lemma 5.4, for the normalised auxiliary state

$$\begin{aligned}
&\sigma_{A_1^n B_1^n E} \\
&:= \int_{-\infty}^{\infty} dt \beta_0(t) \prod_{k=0}^{n-1} \left[\left(\bar{\rho}^{(k)}_{A_1^k B_1^k E} \right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(k+1)}_{A_1^k B_1^k E} \right)^{-\frac{1-it}{2}} \right] \cdot \bar{\rho}^{(n)}_{A_1^n B_1^n E} \cdot \prod_{k=n-1}^{0} \left[\left(\bar{\rho}^{(k+1)}_{A_1^k B_1^k E} \right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(k)}_{A_1^k B_1^k E} \right)^{\frac{1+it}{2}} \right]
\end{aligned} \tag{C.44}$$

we have that

$$D_m(\rho_{A_1^n B_1^n E} \| \sigma_{A_1^n B_1^n E}) \leq n z(\epsilon+\nu,\delta) \tag{C.45}$$

and

$$\begin{aligned}
&\sigma_{A_1^k B_1^k E} \\
&= \int_{-\infty}^{\infty} dt \beta_0(t) \prod_{j=0}^{k-1} \left[\left(\bar{\rho}^{(j)}_{A_1^j B_1^j E} \right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(j+1)}_{A_1^j B_1^j E} \right)^{-\frac{1-it}{2}} \right] \cdot \bar{\rho}^{(k)}_{A_1^k B_1^k E} \cdot \prod_{j=k-1}^{0} \left[\left(\bar{\rho}^{(j+1)}_{A_1^j B_1^j E} \right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(j)}_{A_1^j B_1^j E} \right)^{\frac{1+it}{2}} \right]
\end{aligned} \tag{C.46}$$

$$= \mathcal{M}_k^{\delta} \left(\int_{-\infty}^{\infty} dt \beta_0(t) \prod_{j=0}^{k-1} \left[\left(\bar{\rho}^{(j)}_{A_1^j B_1^j E} \right)^{\frac{1-it}{2}} \left(\bar{\rho}^{(j+1)}_{A_1^j B_1^j E} \right)^{-\frac{1-it}{2}} \right] \cdot \omega^{(k,1)}_{A_1^{k-1} B_1^{k-1} R_k E} \cdot \prod_{j=k-1}^{0} \left[\left(\bar{\rho}^{(j+1)}_{A_1^j B_1^j E} \right)^{-\frac{1+it}{2}} \left(\bar{\rho}^{(j)}_{A_1^j B_1^j E} \right)^{\frac{1+it}{2}} \right] \right) \tag{C.47}$$

for all $k \in [n]$. Let $\sigma^{(k,0)}_{A_1^{k-1} B_1^{k-1} R_k E}$ be the input state for $\mathcal{M}_k^{\delta}$ above, so that

$$\sigma_{A_1^k B_1^k E} = \mathcal{M}_k^{\delta} \left(\sigma^{(k,0)}_{A_1^{k-1} B_1^{k-1} R_k E} \right) \tag{C.48}$$

It is also easy to see using the properties of the algebras $\mathcal{A}_k$ that $\sigma_{A_1^k B_1^k E} \in \mathcal{A}_k$ for each k. Thus, we can extend the state σ as follows using measurements $(\mathcal{T}_i)_i$

$$\sigma_{A_1^n B_1^n X_1^n E} = \mathcal{T}_n \circ \cdots \circ \mathcal{T}_1 \left(\sigma_{A_1^n B_1^n E} \right).$$

Note that the partial state

$$\begin{aligned}
\sigma_{A_1^k B_1^k X_1^k E} &= \mathrm{tr}_{A_{k+1}^n B_{k+1}^n X_{k+1}^n} \left(\mathcal{T}_n \circ \cdots \circ \mathcal{T}_1 \left(\sigma_{A_1^n B_1^n E} \right) \right) \\
&= \mathcal{T}_k \circ \cdots \circ \mathcal{T}_1 \left(\mathrm{tr}_{A_{k+1}^n B_{k+1}^n X_{k+1}^n} \left(\mathcal{T}_n \circ \cdots \circ \mathcal{T}_{k+1} \left(\sigma_{A_1^n B_1^n E} \right) \right) \right) \\
&= \mathcal{T}_k \circ \cdots \circ \mathcal{T}_1 \left(\sigma_{A_1^k B_1^k E} \right)
\end{aligned} \tag{C.49}$$

$$= \mathcal{T}_k \circ \mathcal{M}_k^{\delta} \circ \mathcal{T}_{k-1} \circ \cdots \circ \mathcal{T}_1 \left(\sigma^{(k,0)}_{A_1^{k-1} B_1^{k-1} R_k E} \right) \tag{C.50}$$

Let's define $\mu := z(\epsilon + \nu, \delta)^{1/3}$. Using the substate theorem (Theorem 2.26), we get the following bound from the above relative entropy bound

$$D^{\mu}_{\max}(\rho_{A_1^n B_1^n E} \| \sigma_{A_1^n B_1^n E}) \leq n\mu + \frac{1}{\mu^2} + \log \frac{1}{1-\mu^2} \tag{C.51}$$

which using data processing also implies that

$$D^{\mu}_{\max}(\rho_{A_1^n B_1^n X_1^n E} \| \sigma_{A_1^n B_1^n X_1^n E}) \leq n\mu + \frac{1}{\mu^2} + \log \frac{1}{1-\mu^2} \tag{C.52}$$

The bound above implies that there exists a state $\rho'_{A_1^n B_1^n X_1^n E}$, which is also classical on X_1^n such that

$$P\left(\rho_{A_1^n B_1^n X_1^n E}, \rho'_{A_1^n B_1^n X_1^n E}\right) \leq \mu \tag{C.53}$$

and

$$\rho'_{A_1^n B_1^n X_1^n E} \leq \frac{e^{n\mu + \frac{1}{\mu^2}}}{1-\mu^2} \sigma_{A_1^n B_1^n X_1^n E}. \tag{C.54}$$

The registers X_1^n for ρ' can be chosen to be classical, since the channel measuring X_1^n only decreases the distance between ρ' and ρ, and the new state produced would also satisfy Eq. C.54. As the registers X_1^n are classical for both σ and ρ', we can condition these states on the event Ω. We will call the probability of the event Ω for the state σ and ρ' $P_\sigma(\Omega)$ and $P_{\rho'}(\Omega)$ respectively. Using Lemma A.14 and the Fuchs-van de Graaf inequality, we have

$$P\left(\rho_{A_1^n B_1^n X_1^n E|\Omega}, \rho'_{A_1^n B_1^n X_1^n E|\Omega}\right) \leq 2\sqrt{\frac{\mu}{P_\rho(\Omega)}}. \tag{C.55}$$

Conditioning Eq. C.54 on Ω, we get

$$P_{\rho'}(\Omega) \rho'_{A_1^n B_1^n X_1^n E|\Omega} \leq \frac{e^{n\mu + \frac{1}{\mu^2}}}{1-\mu^2} P_\sigma(\Omega) \sigma_{A_1^n B_1^n X_1^n E|\Omega}. \tag{C.56}$$

Together, the above two equations imply that

$$D^{\mu'}_{\max}(\rho_{A_1^n B_1^n X_1^n E|\Omega} \| \sigma_{A_1^n B_1^n X_1^n E|\Omega}) \leq n\mu + \frac{1}{\mu^2} + \log \frac{P_\sigma(\Omega)}{P_{\rho'}(\Omega)} + \log \frac{1}{1-\mu^2} \tag{C.57}$$

for $\mu' := 2\sqrt{\frac{\mu}{P_\rho(\Omega)}}$.

For $\epsilon' > 0$ such that $\mu' + \epsilon' < 1$ and $\alpha \in (1, 2]$, we can plug the above in the bound provided by Lemma 3.5 to get

$$\begin{aligned} H^{\mu'+\epsilon'}_{\min}(A_1^n | B_1^n E)_{\rho_{|\Omega}} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n | B_1^n E)_{\sigma_{|\Omega}} - \frac{\alpha}{\alpha - 1} n\mu \\ - \frac{1}{\alpha - 1} \left(\alpha \log \frac{P_\sigma(\Omega)}{P_{\rho'}(\Omega)} + \frac{\alpha}{\mu^2} + \alpha \log \frac{1}{1-\mu^2} + g_1(\epsilon', \mu') \right). \end{aligned} \tag{C.58}$$

Now, note that using Eq. 5.100 and C.49, and [DFR20, Lemma B.7] we have

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^nE)_{\sigma_{|\Omega}} = \tilde{H}^{\uparrow}_{\alpha}(A_1^nX_1^n|B_1^nE)_{\sigma_{|\Omega}}. \tag{C.59}$$

For every k, we introduce a register D_k of dimension $|D| := \lceil e^{\max(f)-\min(f)} \rceil \leq e^{2\lceil \|\nabla f\|_\infty \rceil}$ and the channels $\mathcal{D}_k : X_k \to X_kD_k$ defined as

$$\mathcal{D}_k(\omega) := \sum_x \langle x|\omega|x\rangle |x\rangle\langle x| \otimes \nu_x \tag{C.60}$$

where for every x, the state ν_x is a mixture between a uniform distribution on $\{1, 2, \cdots, \lfloor e^{\max(f)-f(\delta_x)} \rfloor\}$ and a uniform distribution on $\{1, 2, \cdots, \lceil e^{\max(f)-f(\delta_x)} \rceil\}$, so that

$$H(D_k)_{\nu_x} = \max(f) - f(\delta_x) \tag{C.61}$$

where δ_x is the distribution with unit weight at element x. Define the state

$$\bar{\sigma}_{A_1^nB_1^nX_1^nD_1^nE} := \mathcal{D}_n \circ \cdots \circ \mathcal{D}_1(\sigma_{A_1^nB_1^nX_1^nE}) \tag{C.62}$$

Now [MFSR24, Lemma 4.5] implies that this satisfies

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^nX_1^n|B_1^nE)_{\bar{\sigma}_{|\Omega}} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1^nX_1^nD_1^n|B_1^nE)_{\bar{\sigma}_{|\Omega}} - \max_{x_1^n \in \Omega} H_{\alpha}(D_1^n)_{\bar{\sigma}_{|x_1^n}} \tag{C.63}$$

For $x_1^n \in \Omega$, we have

$$\begin{aligned} H_{\alpha}(D_1^n)_{\bar{\sigma}_{|x_1^n}} &\leq H(D_1^n)_{\bar{\sigma}_{|x_1^n}} \\ &= \sum_{k=1}^n H(D_k)_{\nu_{x_k}} \\ &= \sum_{k=1}^n \max(f) - f(\delta_{x_k}) \\ &= n\max(f) - nf(\mathrm{freq}(x_1^n)) \\ &\leq n\max(f) - nh. \end{aligned} \tag{C.64}$$

We can get rid of the conditioning on the right-hand side of Eq. C.63 by using [DFR20, Lemma B.5]. This gives us

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^nX_1^nD_1^n|B_1^nE)_{\bar{\sigma}_{|\Omega}} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1^nX_1^nD_1^n|B_1^nE)_{\bar{\sigma}} - \frac{\alpha}{\alpha-1}\log\frac{1}{P_{\sigma}(\Omega)} \tag{C.65}$$

Moreover, using Eq. C.50, we can show that B_k is independent of $A_1^{k-1}X_1^{k-1}D_1^{k-1}B_1^{k-1}E$ in $\bar{\sigma}$. Firstly, for $\sigma_{A_1^{k-1}B_1^kE} = \mathrm{tr}_{A_k} \circ \mathcal{M}_k^{\delta}(\sigma^{(k,0)}_{A_1^{k-1}B_1^{k-1}R_kE})$, we have

$$\sigma_{A_1^{k-1}B_1^kE} = (1-\delta)\,\mathrm{tr}_{A_k} \circ \mathcal{M}_k(\sigma^{(k,0)}_{A_1^{k-1}B_1^{k-1}R_kE}) + \delta\rho^{(k,\delta)}_{B_k} \otimes \sigma_{A_1^{k-1}B_1^{k-1}E} \tag{C.66}$$

$$= \left((1-\delta)\theta^{(k)}_{B_k} + \delta\rho^{(k,\delta)}_{B_k}\right) \otimes \sigma_{A_1^{k-1}B_1^{k-1}E}. \tag{C.67}$$

Since, $\bar{\sigma}_{A_1^{k-1}X_1^{k-1}D_1^{k-1}B_1^kE} = \mathcal{D}_{k-1}\circ\mathcal{T}_{k-1}\cdots\circ\mathcal{D}_1\circ\mathcal{T}_1(\sigma_{A_1^{k-1}B_1^kE})$, we can easily see that B_k is independent of the other registers. In particular, $\bar{\sigma}$ satisfies the Markov chain $A_1^{k-1}X_1^{k-1}D_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k$. Let's define the channels

$$\bar{\mathcal{M}}_k := \mathcal{D}_k\circ\mathcal{T}_k\circ\mathcal{M}_k \tag{C.68}$$

$$\bar{\mathcal{M}}_k^{\delta} := \mathcal{D}_k\circ\mathcal{T}_k\circ\mathcal{M}_k^{\delta}. \tag{C.69}$$

Now we can use [DFR20, Corollary 3.5] to show that for every $k\in[n]$

$$\begin{aligned}\tilde{H}^{\downarrow}_{\alpha}(A_1^kX_1^kD_1^k|B_1^kE)_{\bar{\sigma}} &\geq \tilde{H}^{\downarrow}_{\alpha}(A_1^{k-1}X_1^{k-1}D_1^{k-1}|B_1^{k-1}E)_{\bar{\sigma}} + \inf_{\omega_{R_k\tilde{R}_k}}\tilde{H}^{\downarrow}_{\alpha}(A_kX_kD_k|B_k\tilde{R}_k)_{\bar{\mathcal{M}}_k^{\delta}(\omega)}\\ &\geq \tilde{H}^{\downarrow}_{\alpha}(A_1^{k-1}X_1^{k-1}D_1^{k-1}|B_1^{k-1}E)_{\bar{\sigma}}\\ &\quad + \min\left\{\inf_{\omega_{R_k\tilde{R}_k}}\tilde{H}^{\downarrow}_{\alpha}(A_kD_k|B_k\tilde{R}_k)_{\bar{\mathcal{M}}_k(\omega)}, \tilde{H}^{\downarrow}_{\alpha}(A_kD_k|B_k)_{\mathcal{D}_k\circ\mathcal{T}_k(\rho^{(k,\delta)}_{A_kB_k})}\right\}.\end{aligned} \tag{C.70}$$

where we have used the quasi-concavity of Rényi conditional entropies [Tom16, Pg 73] and the fact that X_k are classical in the second line.

We now lower bound the two terms in the minimum above. Using [DFR20, Lemma B.9], for a state $\nu = \bar{\mathcal{M}}_k(\omega_{R_k\tilde{R}_k})$ and $1<\alpha<\frac{1}{\log(1+2|A||D|)}$ we have that

$$\begin{aligned}\tilde{H}^{\downarrow}_{\alpha}(A_kD_k|B_k\tilde{R}_k)_{\nu} &\geq H(A_kD_k|B_k\tilde{R}_k)_{\nu} - (\alpha-1)\log^2(1+2|A||D|)\\ &= H(A_k|B_k\tilde{R}_k)_{\nu} + H(D_k|A_kB_k\tilde{R}_k)_{\nu} - (\alpha-1)\log^2(1+2|A||D|)\\ &= H(A_k|B_k\tilde{R}_k)_{\nu} + H(D_k|X_k)_{\nu} - (\alpha-1)\log^2(1+2|A||D|)\\ &= H(A_k|B_k\tilde{R}_k)_{\nu} + \sum_x \nu(x)(\max(f)-f(\delta_x)) - (\alpha-1)\log^2(1+2|A||D|)\\ &= H(A_k|B_k\tilde{R}_k)_{\nu} + \max(f) - f(\nu_X) - (\alpha-1)\log^2(1+2|A||D|)\\ &\geq \max(f) - (\alpha-1)\log^2(1+2|A||D|)\end{aligned} \tag{C.71}$$

For the second term, and $1<\alpha<\frac{1}{\log(1+2|A||D|)}$, we have

$$\begin{aligned}\tilde{H}^{\downarrow}_{\alpha}(A_kD_k|B_k)_{\mathcal{D}_k\circ\mathcal{T}_k(\rho^{(k,\delta)}_{A_kB_k})} &\geq H(A_kD_k|B_k)_{\mathcal{D}_k\circ\mathcal{T}_k(\rho^{(k,\delta)}_{A_kB_k})} - (\alpha-1)\log^2(1+2|A||D|)\\ &\geq H(A_kD_k|B_k)_{\bar{\mathcal{M}}_k(\tilde{\rho}^{(k,0)}_{R_k})} - (\alpha-1)\log^2(1+2|A||D|)\\ &\qquad - 2(\epsilon+\delta)\log|A||D| - g(\epsilon+\delta)\\ &\geq \max(f) - (\alpha-1)\log^2(1+2|A||D|) - 2(\epsilon+\delta)\log|A||D| - g_2(\epsilon+\delta)\end{aligned} \tag{C.72}$$

where we have used [DFR20, Lemma B.9] in the first line, the AFW continuity bound [Wil13, Theorem 11.10.3] in the second line to convert the entropy on $\mathcal{D}_k\circ\mathcal{T}_k(\rho^{(k,\delta)}_{A_kB_k})$

to an entropy on $\bar{\mathcal{M}}_k(\tilde{\rho}^{(k,0)}_{R_k})$ and finally we use bound in Eq. C.71 for states of this form.

Plugging these in Eq. C.70, gives us that for every $k \in [n]$

$$\begin{aligned}\tilde{H}^{\downarrow}_{\alpha}(A_1^k X_1^k D_1^k | B_1^k E)_{\bar{\sigma}} \geq \tilde{H}^{\downarrow}_{\alpha}(A_1^{k-1} X_1^{k-1} D_1^{k-1} | B_1^{k-1} E)_{\bar{\sigma}} \\ + \max(f) - (\alpha-1)\log^2(1+2|A||D|) \\ - 2(\epsilon+\delta)\log|A||D| - g_2(\epsilon+\delta)\end{aligned} \tag{C.73}$$

Consecutively using this bound gives us

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1^n X_1^n D_1^n | B_1^n E)_{\bar{\sigma}} \geq \tilde{H}^{\downarrow}_{\alpha}(A_1^n X_1^n D_1^n | B_1^n E)_{\bar{\sigma}} \\ \geq n\max(f) - n(\alpha-1)\log^2(1+2|A||D|) \\ - 2n(\epsilon+\delta)\log|A||D| - ng_2(\epsilon+\delta)\end{aligned} \tag{C.74}$$

Combining Eq. C.59, C.63, C.64, C.65 and C.74, we get

$$\begin{aligned}\tilde{H}^{\uparrow}_{\alpha}(A_1^n | B_1^n E)_{\sigma_{|\Omega}} \geq nh - n(\alpha-1)\log^2(1+2|A||D|) - 2n(\epsilon+\delta)\log|A||D| - ng_2(\epsilon+\delta) \\ - \frac{\alpha}{\alpha-1}\log\frac{1}{P_\sigma(\Omega)}.\end{aligned} \tag{C.75}$$

Plugging this into Eq. C.58, we get

$$\begin{aligned}H^{\mu'+\epsilon'}_{\min}(A_1^n | B_1^n E)_{\rho_{|\Omega}} \geq nh - n(\alpha-1)\log^2(1+2|A||D|) - \frac{\alpha}{\alpha-1}n\mu \\ - 2n(\epsilon+\delta)\log|A||D| - ng_2(\epsilon+\delta) \\ - \frac{1}{\alpha-1}\left(\alpha\log\frac{1}{P_{\rho'}(\Omega)} + \frac{\alpha}{\mu^2} + \alpha\log\frac{1}{1-\mu^2} + g_1(\epsilon',\mu')\right).\end{aligned} \tag{C.76}$$

Finally, we choose $\alpha = 1 + \frac{\sqrt{\mu}}{\log(1+2|A||D|)}$ and use the $\alpha < 2$ as an upper bound to derive

$$\begin{aligned}H^{\mu'+\epsilon'}_{\min}(A_1^n | B_1^n E)_{\rho_{|\Omega}} \geq n(h - 3\sqrt{\mu}\log(1+2|A||D|) - 2(\epsilon+\delta)\log|A||D| - g_2(\epsilon+\delta)) \\ - \frac{\log(1+2|A||D|)}{\sqrt{\mu}}\left(2\log\frac{1}{P_\rho(\Omega)-\mu} + \frac{2}{\mu^2} + 2\log\frac{1}{1-\mu^2} + g_1(\epsilon',\mu')\right).\end{aligned} \tag{C.77}$$

Note that $\log(1+2|A||D|) \leq \log(1+2|A|) + \log|D| \leq \log(1+2|A|) + 2\lceil\|\nabla f\|_\infty\rceil = V$. Recall that $\mu = z(\epsilon+\nu,\delta)^{1/3}$. Since, ν can be chosen arbitrarily greater than 0, if we let $\nu \to 0$, the bound in Eq. C.77 is still valid. Finally, choose $\delta = \epsilon$ to derive the bound in the theorem.

Case 2: Finally, for a state $\rho_{A_1^n B_1^n E}$ which satisfies the assumptions of the Theorem but is not full rank, we can consider the full rank states $\rho^{(\varepsilon)}_{A_1^n B_1^n E} := (1-\varepsilon)\rho_{A_1^n B_1^n E} + \varepsilon\tau_{A_1^n B_1^n E}$ for an arbitrarily small $\varepsilon > 0$. $\rho^{(\varepsilon)}_{A_1^n B_1^n E}$ will be full rank and satisfy the assumptions of the Theorem

and the bound in Eq. C.77 for $\epsilon \to \epsilon + \varepsilon$. One can then prove the lower bound above for the state $\rho_{A_1^n B_1^n E}$ by taking the limit $\varepsilon \to 0$. □

C.5. Proof of Lemma 5.13

The following proof was provided by user:fedja in response to a question by user:noel (pseudonym used by AM) on MathOverflow [fu23]. We reproduce it here for completeness.

Proof of Lemma 5.13. Let $\sigma = \sum_i q_i |x_i\rangle\langle x_i|$ be the eigenvalue decomposition of σ. Let $\delta_1 \in (0,1)$ be a small parameter, which will be specified later. For every $k \geq 0$, define

$$\mathcal{X}_k := \operatorname{span}\left\{|x_i\rangle : q_i \in \left((1+\delta_1)^{-(k+1)}, (1+\delta_1)^{-k}\right]\right\} \tag{C.78}$$

$$d_k := \dim(\mathcal{X}_k) \tag{C.79}$$

We have $\mathcal{X} = \bigoplus_{k=0}^{\infty} \mathcal{X}_k$ and $\sigma = \bigoplus_{k=0}^{\infty} \sigma|_{\mathcal{X}_k}$. Let P_k be the projector on the space $\mathcal{X}_k$ for every k. Note that P_k commute with σ. If we restrict σ to the space $\mathcal{X}_k$, we have

$$\frac{1}{(1+\delta_1)^{k+1}} \mathbb{1}_{\mathcal{X}_k} \leq \sigma|_{\mathcal{X}_k} \leq \frac{1}{(1+\delta_1)^k} \mathbb{1}_{\mathcal{X}_k} \tag{C.80}$$

for every k. Further, for any projector Π_k in the space $\mathcal{X}_k$, we have

$$\begin{aligned} \|\sigma|_{\mathcal{X}_k}^{-\frac{1}{2}} \Pi_k \sigma|_{\mathcal{X}_k} \Pi_k \sigma|_{\mathcal{X}_k}^{-\frac{1}{2}}\|_\infty &\leq \|\sigma|_{\mathcal{X}_k}\|_\infty \|\sigma|_{\mathcal{X}_k}^{-1}\|_\infty \\ &\leq 1 + \delta_1 \end{aligned}$$

which implies that for any projector Π_k in $\mathcal{X}_k$

$$\Pi_k \sigma|_{\mathcal{X}_k} \Pi_k \leq (1+\delta_1)\sigma|_{\mathcal{X}_k}. \tag{C.81}$$

We will choose the projector Π to satisfy the lemma to be of the form $\Pi = \bigoplus_{k=0}^{\infty} \Pi_k$, where each Π_k is a projector in the space $\mathcal{X}_k$. With such a projector, we have

$$\begin{aligned} \Pi\sigma\Pi &= \bigoplus_{k=0}^{\infty} \Pi_k \sigma|_{\mathcal{X}_k} \Pi_k \\ &\leq (1+\delta_1) \bigoplus_{k=0}^{\infty} \sigma|_{\mathcal{X}_k} \\ &= (1+\delta_1)\sigma. \end{aligned} \tag{C.82}$$

Define $\Delta := (\rho - \sigma)^+$ to be the positive part of $(\rho - \sigma)$ and observe that $\operatorname{tr}(\Delta) \leq 2\epsilon$. Note that $\rho \leq \sigma + \Delta$. Further, let $\Delta_k := P_k \Delta P_k$ and $\mu_k := \operatorname{tr}(\Delta P_k)$ for every k. Then, we have that

$$\begin{aligned} \sum_{k=0}^{\infty} \mu_k &= \sum_{k=0}^{\infty} \operatorname{tr}(\Delta P_k) \\ &= \operatorname{tr}(\Delta) \\ &\leq 2\epsilon. \end{aligned} \tag{C.83}$$

Let $\Delta_k = \sum_{i=1}^{d_k} \nu_{ki} |y_{ki}\rangle \langle y_{ki}|$ be the eigenvalue decomposition of Δ_k. Let $\delta_2 \in (0,1)$ and $K \in \mathbb{N}$ be two more parameters to be chosen later. Define the subspace $\mathcal{Y}_k$ of $\mathcal{X}_k$ as

$$\mathcal{Y}_k := \operatorname{span}\left\{|y_{ki}\rangle : \nu_{ki} \leq \delta_2(1+\delta_1)^{-(k+1)}\right\}. \tag{C.84}$$

Note that the codimension of $\mathcal{Y}_k$ in the space $\mathcal{X}_k$, is at most $\frac{\mu_k}{\delta_2}(1+\delta_1)^{k+1}$. We further restrict the subspace $\mathcal{Y}_k$ to the space

$$\mathcal{Z}_k := \mathcal{Y}_k \cap \bigcap_{j=0}^{k-K-1} \ker(P_j \Delta). \tag{C.85}$$

Since, rank of P_j is d_j, the codimension of $\mathcal{Z}_k$ in $\mathcal{Y}_k$ is at most $\sum_{j=0}^{k-K-1} d_j$. Define Π_k to be the projector on $\mathcal{Z}_k$. Note that since $\mathcal{Z}_k$ is a subspace of $\mathcal{X}_k$, Π_k is a projector in the space $\mathcal{X}_k$. We define the projector

$$\Pi := \bigoplus_{k=0}^{\infty} \Pi_k, \tag{C.86}$$

which is of the form promised and show that this satisfies the conditions of the lemma.

We begin by showing that $\Pi\Delta\Pi$ can be bounded by a small multiple of $\Pi\sigma\Pi$, specifically

$$\Pi\Delta\Pi \leq (2K+1)\delta_2 \Pi\sigma\Pi.$$

It is sufficient to show that for $|v\rangle = \bigoplus_{k=0}^{\infty} |v_k\rangle$ such that $|v_k\rangle \in \mathcal{Z}_k$ for every k, we have

$$\langle v|\Delta|v\rangle \leq (2K+1)\delta_2 \langle v|\sigma|v\rangle. \tag{C.87}$$

The left-hand side above can be expanded as

$$\begin{aligned}
\langle v|\Delta|v\rangle &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \langle v_i|\Delta|v_j\rangle \\
&= \sum_{i=0}^{\infty} \langle v_i|\Delta|v_i\rangle + 2\sum_{i=0}^{\infty}\sum_{j=0}^{i-1} \operatorname{Re}\{\langle v_i|\Delta|v_j\rangle\} \\
&\leq \sum_{i=0}^{\infty} \langle v_i|\Delta|v_i\rangle + 2\sum_{i=0}^{\infty}\sum_{j=0}^{i-1} |\langle v_i|\Delta|v_j\rangle| \\
&= \sum_{i=0}^{\infty} \langle v_i|\Delta|v_i\rangle + 2\sum_{i=0}^{\infty}\sum_{j=i-K}^{i-1} |\langle v_i|\Delta|v_j\rangle| + 2\sum_{i=0}^{\infty}\sum_{j=0}^{i-K-1} |\langle v_i|\Delta|v_j\rangle|.
\end{aligned} \tag{C.88}$$

The first summation above can be bounded as

$$\begin{aligned}
\sum_{i=0}^{\infty} \langle v_i|\Delta|v_i\rangle &\leq \delta_2 \sum_{i=0}^{\infty} \frac{1}{(1+\delta_1)^{i+1}} \langle v_i|v_i\rangle \\
&\leq \delta_2 \sum_{i=0}^{\infty} \langle v_i|\sigma|v_i\rangle
\end{aligned} \tag{C.89}$$

where we have used the fact that $|v_i\rangle \in \mathcal{Y}_i$ and the maximum eigenvalue of $P_i\Delta P_i = \Delta_i$ in this subspace is at most $\delta_2 \frac{1}{(1+\delta_1)^{i+1}}$, and the fact that the minimum eigenvalue of σ in $\mathcal{Z}_i \subseteq \mathcal{X}_i$

is at least $\frac{1}{(1+\delta_1)^{i+1}}$.

We can bound the second term as

$$\begin{aligned}
\sum_{i=0}^{\infty}\sum_{j=i-K}^{i-1} |\langle v_i|\Delta|v_j\rangle| &\leq \sum_{i=0}^{\infty}\sum_{j=i-K}^{i-1} \left\|\Delta^{\frac{1}{2}}|v_j\rangle\right\| \left\|\Delta^{\frac{1}{2}}|v_i\rangle\right\| \\
&\leq \sum_{i=0}^{\infty}\frac{1}{2}\sum_{j=i-K}^{i-1}\left(\left\|\Delta^{\frac{1}{2}}|v_j\rangle\right\|^2 + \left\|\Delta^{\frac{1}{2}}|v_i\rangle\right\|^2\right) \\
&= K\sum_{i=0}^{\infty}\left\|\Delta^{\frac{1}{2}}|v_i\rangle\right\|^2 \\
&= K\sum_{i=0}^{\infty}\langle v_i|\Delta|v_i\rangle
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality in the first line along with the fact that $\Delta \geq 0$ and the AM-GM inequality in the second line. This summation can now be bounded using Eq. C.89.

The last term in Eq. C.88 is a summation over inner products $|\langle v_i|\Delta|v_j\rangle| = |\langle v_j|P_j\Delta|v_i\rangle|$ for $j \leq i-K-1$. By our definition for the subspace $\mathcal{Z}_i$, we have that $|v_i\rangle \in \ker(P_j\Delta)$ for these choices of the index j.

Putting these together, we have the bound

$$\begin{aligned}
\langle v|\Delta|v\rangle &\leq (2K+1)\,\delta_2\sum_{i=0}^{\infty}\langle v_i|\sigma|v_i\rangle \\
&= (2K+1)\,\delta_2\,\langle v|\sigma|v\rangle
\end{aligned}$$

for every $|v\rangle = \Pi|v\rangle$. Above we have used the fact that $|v_i\rangle \in \mathcal{X}_i$ and $\sigma = \bigoplus_{i=0}^{\infty}\sigma|_{\mathcal{X}_i}$. Therefore, we have

$$\Pi\Delta\Pi \leq (2K+1)\,\delta_2\Pi\sigma\Pi \tag{C.90}$$

and also

$$\begin{aligned}
\Pi\rho\Pi &\leq \Pi\,(\sigma+\Delta)\,\Pi \\
&\leq (1+(2K+1)\,\delta_2)\,\Pi\sigma\Pi.
\end{aligned} \tag{C.91}$$

Using Eq. C.82 we can remove the sandwiching projectors from σ to get the bound

$$\Pi\rho\Pi \leq (1+(2K+1)\,\delta_2)\,(1+\delta_1)\sigma. \tag{C.92}$$

Now, we will show that $\mathrm{tr}(\Pi\sigma)$ is large. Recall that the codimension of $\mathcal{Z}_k$ in $\mathcal{X}_k$ is at most

$$\frac{\mu_k}{\delta_2}(1+\delta_1)^{k+1} + \sum_{j=0}^{k-K-1} d_j. \tag{C.93}$$

Hence,

$$\begin{aligned}
\mathrm{tr}\left((\mathbb{1}-\Pi)\sigma\right) &= \sum_{k=0}^{\infty} \mathrm{tr}\left((P_k - \Pi_k)\sigma|_{\mathcal{X}_k}\right) \\
&\leq \sum_{k=0}^{\infty} \frac{1}{(1+\delta_1)^k} \mathrm{tr}(P_k - \Pi_k) \\
&\leq \sum_{k=0}^{\infty} \frac{1}{(1+\delta_1)^k} \left(\frac{\mu_k}{\delta_2}(1+\delta_1)^{k+1} + \sum_{j=0}^{k-K-1} d_j\right) \\
&= \frac{1}{\delta_2}(1+\delta_1)\sum_{k=0}^{\infty}\mu_k + \sum_{j=0}^{\infty} d_j \sum_{k=j+K+1}^{\infty} (1+\delta_1)^{-k} \\
&= \frac{2\epsilon(1+\delta_1)}{\delta_2} + \frac{1}{\delta_1}(1+\delta_1)^{-(K-1)} \sum_{j=0}^{\infty} d_j \frac{1}{(1+\delta_1)^{j+1}} \\
&= (1+\delta_1)\left(\frac{2\epsilon}{\delta_2} + \frac{1}{\delta_1}(1+\delta_1)^{-K}\right)
\end{aligned} \tag{C.94}$$

where in the second line we have used $\sigma|_{\mathcal{X}_k} \leq (1+\delta_1)^{-k}\,\mathbb{1}_{\mathcal{X}_k}$, in the third line we have used the bound on the codimension of $\mathcal{Z}_k$ in Eq. C.93, in the fifth line we have used the bound in Eq. C.83 and in the last line we have used the fact that $\sum_{j=0}^{\infty} d_j(1+\delta_1)^{-(j+1)} \leq \mathrm{tr}\,\sigma \leq 1$.

Now all that is left to do is to select the parameters δ_1, δ_2 and K. We choose

$$\delta_1 := \epsilon^{1/3} \tag{C.95}$$

$$\delta_2 := \epsilon^{2/3} \tag{C.96}$$

$$K := \left\lfloor \frac{2}{\epsilon^{1/3}} \log \frac{1}{\epsilon^{2/3}} \right\rfloor. \tag{C.97}$$

For these parameters, we get

$$\Pi\rho\Pi \leq \left(1 + \frac{8}{3}\epsilon^{1/3}\log\frac{1}{\epsilon} + \epsilon^{2/3}\right)(1+\epsilon^{1/3})\sigma. \tag{C.98}$$

Note that $(1+\delta_1)^{-1} \leq (1-\delta_1/2) \leq e^{-\delta_1/2}$ since $\delta_1 \in (0,1)$. This implies

$$\begin{aligned}
\mathrm{tr}\left((\mathbb{1}-\Pi)\sigma\right) &\leq (1+\epsilon^{1/3})\left(2\epsilon^{1/3} + \frac{1}{\epsilon^{1/3}} e^{-K\frac{\epsilon^{1/3}}{2}}\right) \\
&\leq (1+\epsilon^{1/3})\left(2\epsilon^{1/3} + \frac{1}{\epsilon^{1/3}} e^{-\frac{\epsilon^{1/3}}{2}\left(\frac{2}{\epsilon^{1/3}}\log\frac{1}{\epsilon^{2/3}} - 1\right)}\right) \\
&\leq (1+\epsilon^{1/3})\left(2\epsilon^{1/3} + \frac{1}{\epsilon^{1/3}} 2\epsilon^{2/3}\right) \\
&\leq 4(1+\epsilon^{1/3})\epsilon^{1/3}.
\end{aligned}$$

Finally, using the fact that $\frac{1}{2}\|\rho-\sigma\|_1 \leq \epsilon$, we get

$$\begin{aligned} \operatorname{tr}\left((\mathbb{1}-\Pi)\rho\right) &\leq \operatorname{tr}\left((\mathbb{1}-\Pi)\sigma\right) + 2\epsilon \\ &\leq 4\big(1+\epsilon^{1/3}\big)\epsilon^{1/3} + 2\epsilon. \end{aligned} \tag{C.99}$$

□

Appendix D

Appendices to Chapter 6

D.1. Single-round results

Lemma D.1. *Suppose Alice and Bob use a strategy, which wins the 3CHSH$_\perp$ game with probability ω. Then, their answers A and B for the game satisfy*

$$\Pr[A = B] \geq \omega - \nu - 2\alpha. \tag{D.1}$$

Proof. Let P_{XY} represent the probability distribution of the questions for the 3CHSH$_\perp$ game. It follows that the winning probability of the 3CHSH$_\perp$ game satisfies:

$$\begin{aligned}
\omega &= \sum_{x,y} P_{XY}(xy) \sum_{a,b:V(x,y,a,b)=1} P_{AB|XY}(ab|xy) \\
&= (1-\alpha)^2(1-\nu)\Pr[A = B|X = 0, Y = 2] + (1-\alpha)^2\nu \Pr(A \oplus B = XY|X,Y \in \{0,1\}) \\
&\quad + (1-(1-\alpha)^2) \\
&\leq (1-\alpha)^2(1-\nu)\Pr[A = B|X = 0, Y = 2] + (1-\alpha)^2\nu + (1-(1-\alpha)^2) \\
&\leq \Pr[A = B] + (1-\alpha)^2\nu + (1-(1-\alpha)^2) \\
&\leq \Pr[A = B] + \nu + 2\alpha.
\end{aligned}$$

□

Lemma D.2 ([AF20, Lemma 5.3])**.** *Suppose that the quantum strategy for the 2CHSH game starting with $\rho^{(0)}_{E_A E_B E}$ wins the 2CHSH game with probability $\omega \in \left[\frac{3}{4}, \frac{2+\sqrt{2}}{4}\right]$. Let A be Alice's answer produced according to the given strategy. Let ρ_{XAE} be the state produced once Alice applies her measurements to $\rho^{(0)}$. Then, for question $x \in \{0,1\}$ we have*

$$H(A|E)_{\rho^{(x)}} \geq F(\omega) \tag{D.2}$$

where $F(\omega) = \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{3 - 16\,\omega\,(1-\omega)}\right)$ *and* $\rho^{(x)}_{AE}$ *are the* A *and* E *registers of* ρ_{XAE} *for* $X = x$.

Proof. This is proved as an intermediate step in the Proof of Lemma 5.3 (Appendix C.1) [**AF20**]. □

Lemma D.3. *Suppose that the quantum strategy for the 3CHSH game starting with* $\rho^{(0)}_{E_A E_B E}$ *wins the 3CHSH game with probability* $\omega \in \left[(1-\nu) + \frac{3}{4}\nu, (1-\nu) + \frac{2+\sqrt{2}}{4}\nu\right]$. *Let* A *be Alice's answer produced according to the strategy. Then, for the post measurement state* ρ_{XAE} *we have*

$$H(A|EX)_\rho \geq F(\hat{\omega}_\nu)$$

where $\hat{\omega}_\nu := \frac{\omega - (1-\nu)}{\nu}$ *and* F *is as in Lemma D.2*

Proof. Let $S = (\rho^{(0)}_{E_A E_B E}, \{A_x\}_{x\in\mathcal{X}}, \{B_y\}_{y\in\mathcal{Y}})$ be the strategy for the 3CHSH mentioned in the lemma hypothesis. Let S' be the strategy for the 2CHSH game, which uses the state $\rho^{(0)}_{E_A E_B E}$, the measurements $\{A_x\}_{x\in\{0,1\}}$ as Alice's measurements and the measurements $\{B_y\}_{y\in\{0,1\}}$ as Bob's measurements. Let P_{XY} be the distribution of questions in the 2CHSH game, $P_{AB|XY}$ be the conditional probability distribution of the answers in the strategy S and $Q_{AB|XY}$ in the strategy S'. Then, by definition of the strategy S', we have for all $x, y \in \{0,1\}$ that

$$Q_{AB|X=x,Y=y} = P_{AB|X=x,Y=y}.$$

Now, observe that the winning probability ω of the 3CHSH game can be written as

$$\begin{aligned}
\omega &= (1-\nu)P[A = B|X = 0, Y = 2] + \nu \sum_{x,y\in\{0,1\}} P_{XY}(xy) \sum_{a,b:a\oplus b=xy} P_{AB|XY}(ab|xy) \\
&\leq (1-\nu) + \nu \sum_{x,y\in\{0,1\}} P_{XY}(xy) \sum_{a,b:a\oplus b=xy} P_{AB|XY}(ab|xy) \\
&= (1-\nu) + \nu \sum_{x,y\in\{0,1\}} P_{XY}(xy) \sum_{a,b:a\oplus b=xy} Q_{AB|XY}(ab|xy) \\
&= (1-\nu) + \nu\omega_{S'}
\end{aligned}$$

where $\omega_{S'}$ is the winning probability for the 2CHSH game with the strategy S'. The above implies

$$\omega_{S'} \geq \frac{\omega - (1-\nu)}{\nu} = \hat{\omega}_\nu. \tag{D.3}$$

Note that $\hat{\omega}_\nu \geq 3/4$ for the range of ω in the hypothesis. Using Lemma D.2, we have for $x \in \{0,1\}$

$$H(A|E)_{\rho_x} \geq \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{3 - 16\hat{\omega}_\nu(1-\hat{\omega}_\nu)}\right)$$

where $\rho_{AE}^{(x)}$ is the state produced at the end of the strategy S' for question $X = x$. This is the same as the state produced at the end of the strategy S for question $X = x$ because of the way we have defined S'. Therefore, for the state produced at the end of S, we have

$$\begin{aligned} H(A|EX)_\rho &= \left(1 - \frac{\nu}{2}\right) H(A|E)_{\rho_{|X=0}} + \frac{\nu}{2} H(A|E)_{\rho_{|X=1}} \\ &\geq \log(2) - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{3 - 16\hat{\omega}_\nu(1-\hat{\omega}_\nu)}\right) \end{aligned}$$

which proves the lemma. □

Proof of Lemma 6.3. Let $\mathcal{X}$ and $\mathcal{Y}$ represent the questions for the 3CHSH game, and $\mathcal{X}_\perp$ and $\mathcal{Y}_\perp$ the questions for the 3CHSH$_\perp$ game. Let $S = (\rho^{(0)}_{E_A E_B E}, \{A_x\}_{x\in\mathcal{X}_\perp}, \{B_y\}_{y\in\mathcal{Y}_\perp})$ be the strategy which wins the 3CHSH$_\perp$ with probability ω. Let $\bar{S}$ be the strategy for the 3CHSH game, which uses the state $\bar{\rho}^{(0)}_{E_A E_B E} = \rho^{(0)}$, the measurements $\{A_x\}_{x\in\mathcal{X}}$ as Alice's measurements and the measurements $\{B_y\}_{y\in\mathcal{Y}}$ as Bob's measurements. Let P_{XY} be the distribution of questions for the 3CHSH$_\perp$ game, Q_{XY} be the distribution of questions for the 3CHSH game, $P_{AB|XY}$ be the conditional probability distribution of the answers in the strategy S and $Q_{AB|XY}$ in the strategy $\bar{S}$. Then, by definition of the strategy $\bar{S}$, we have for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$Q_{AB|X=x,Y=y} = P_{AB|X=x,Y=y}.$$

Now, observe that the winning probability of the 3CHSH$_\perp$ game can be written as

$$\begin{aligned} \omega &= \sum_{x\in\mathcal{X}_\perp, y\in\mathcal{Y}_\perp} P_{XY}(xy) \sum_{a,b:V(ab|xy)=1} P_{AB|XY}(ab|xy) \\ &= (1 - (1-\alpha)^2) + (1-\alpha)^2 \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} Q_{XY}(xy) \sum_{a,b:V(ab|xy)=1} P_{AB|XY}(ab|xy) \\ &= (1 - (1-\alpha)^2) + (1-\alpha)^2 \omega_{\bar{S}} \end{aligned}$$

where $\omega_{\bar{S}}$ is the winning probability for the 3CHSH game under the strategy $\bar{S}$. Thus, we have

$$\omega_{\bar{S}} = 1 - \frac{1-\omega}{(1-\alpha)^2}.$$

Observe that $\omega_{\bar{S}} \in \left[(1-\nu) + \frac{3}{4}\nu, (1-\nu) + \frac{2+\sqrt{2}}{4}\nu\right]$ for the range of ω in the hypothesis for the theorem. Let $\rho_{AE}^{(x)}$ and $\bar{\rho}_{AE}^{(x)}$ denote the states at the end of the protocol S and $\bar{S}$ when Alice receives the question $X = x$. Then, these two states are equal for $x \neq \perp$. Further, the probability distribution $P_{X|X\neq\perp} = Q_X$, which implies that

$$\begin{aligned} \rho_{XAE|X\neq\perp} &= \sum_{x\in\mathcal{X}} P_{X|X\neq\perp}(x) [\![x]\!] \otimes \rho_{AE}^{(x)} \\ &= \sum_{x\in\mathcal{X}} Q_X(x) [\![x]\!] \otimes \bar{\rho}_{AE}^{(x)} \end{aligned}$$

$$= \bar{\rho}_{XAE}.$$

We can use these to create the entropy bound as follows

$$\begin{aligned} H(AB|EXY)_\rho &\geq H(A|EXY)_\rho \\ &= H(A|EX)_\rho \\ &\geq (1-\alpha)H(A|EX)_{\rho_{|X\neq\perp}} \\ &= (1-\alpha)H(A|EX)_{\bar{\rho}} \\ &\geq (1-\alpha)\left(\log(2) - h\left(\frac{1}{2}+\frac{1}{2}\sqrt{3-16g_{\alpha,\nu}(\omega)(1-g_{\alpha,\nu}(\omega))}\right)\right) \end{aligned}$$

where the first line follows from the fact that B is classical, the second line from the no-signalling property which implies that $Y \leftrightarrow X \leftrightarrow AE$, and in the last line we have used Lemma D.3. □

D.2. Adapting statements from [BVY21] to our setting

In the parallel DIQKD setting, Eve distributes the registers E_A and E_B of the state $\psi_{E_AE_BE}$ between Alice and Bob, who then play the n anchored games parallelly using their states. In [BVY21], the situation is almost the same, except there are only two parties, Alice and Bob. We can use the results proven in [BVY21] by simply introducing a register for Eve in the state ψ in their setting and tracking it through their proofs. The objective for Alice and Bob is also different in [BVY21]. They seek to maximise the winning probability of the parallel repetition game. However, since [BVY21] considers an arbitrary strategy for parallel repetition, no modification is required on this account. Lastly, [BVY21] also considers conditioning on an event W_C, representing the two parties winning a subset C of the rounds. The results only rely on this event being determined by the variable r_{-i}, so we can simply replace W_C by the trivial event for our case.

In their proofs, [BVY21] uses the fact that the state between Alice and Bob can be assumed to be symmetric, which we cannot necessarily guarantee with three parties. However, it is straightforward to also carry out their proofs without using this assumption.

We will go through the statements considered in [BVY21] till Section 6 and briefly explain the modifications required to prove them in our setting. The numbering in the following list follows [BVY21].

(1) **Classical results:** All facts involving only classical variables follow for our setting from the same arguments, since we can consider Eve's register E as being a part of

Alice's register E_A, and Alice's measurements as $\{A^{E_A}_{x_1^n}(a_1^n) \otimes \mathbb{1}^E\}$. This reduces our setting to the one in [BVY21] for these results.

(2) **Lemma 4.6**: This is true in our case because only classical variables and their properties are used here.

(3) **Proposition 4.9**: This is true because the argument traces over all the registers in the states Φ. Thus, we can again view E as being a part of E_A here.

(4) **Eq. 33-34**: The definitions of Ξ and Λ should include the register E as part of ψ. Same for the definitions of ξ and λ.

(5) **Claim 5.13**: We instead need to prove that

$$\mathop{\mathbb{E}}_{I} \mathop{\mathbb{E}}_{R|W_C} [I(Y_i : E_A E)_{\xi_r}] = O(\delta)$$
$$\mathop{\mathbb{E}}_{I} \mathop{\mathbb{E}}_{R|W_C} [I(X_i : E_B E)_{\lambda_r}] = O(\delta)$$

[BVY21] proves the first claim and notes that the second one is similar. To prove the above results, we simply need to follow the proof in [BVY21], and note that all instances of E_A in the proof can be replaced with $E_A E$. Also, note that Eq. 38 in [BVY21] can be derived without assuming that ψ is symmetric. To prove the second claim, we would similarly consider $E_B E$ together.

(6) **Claim 5.14**: We instead need to prove that

$$\mathop{\mathbb{E}}_{I} \mathop{\mathbb{E}}_{R_{-i}|W_C} \mathop{\mathbb{E}}_{XY} \|\xi^{E_A E}_{r_{-i},x,y} - \xi^{E_A E}_{r_{-i},x,\perp}\|_1^2 = O(\delta^{1/2}/\alpha^4)$$
$$\mathop{\mathbb{E}}_{I} \mathop{\mathbb{E}}_{R_{-i}|W_C} \mathop{\mathbb{E}}_{XY} \|\lambda^{E_B E}_{r_{-i},x,y} - \lambda^{E_B E}_{r_{-i},\perp,y}\|_1^2 = O(\delta^{1/2}/\alpha^4).$$

Once again this can be done by replacing E_A by $E_A E$ in the proof for the first claim and E_B by $E_B E$ for the second claim. The proof also uses relations between several classical variables, which can be handled using the argument for classical variables mentioned above.

(7) **Proof of Lemma 5.12**:

(a) **Claim 5.15**: We need to prove that $|\tilde{\Phi}_{r_i,\perp/x,y}\rangle^{E_A E_B E}$ is a purification of the state $\xi^{E_A E}_{r_i,x,y}$ and that $|\tilde{\Phi}_{r_i,\perp,y}\rangle^{E_A E_B E}$ is a purification of the state $\xi^{E_A E}_{r_i,\perp,y}$. Once again, this can be done by noting that we don't necessarily need to use the fact that the state is symmetric for the proof. Following the same procedure as [BVY21], we can derive

$$\begin{aligned}\xi^{E_A E}_{r_i,x,y} &= \gamma^{-2}_{r_i,\perp/x,y} A_\omega(a_C)^{1/2} \operatorname{tr}_{E_B} \left(B_{\omega_{-i},y}(b_C)\Psi\right) A_\omega(a_C)^{1/2} \\ &= \tilde{\Phi}^{E_A E}_{r_i,\perp/x,y}.\end{aligned}$$

The rest of the proof remains the same.

(b) **Claim 5.16**: Also needs to be modified as Claim 5.15 has been.

The rest of the steps in the proof rely on using claims, which were proven before. Since, we have already proven them above, the steps follow in our case as well.

(8) **Lemma 5.17**: Since the definition of $\gamma_{r_{-i},s,y}$ traces over all the quantum registers, we can simply view our setting as an instance of the anchored games setting by letting Alice keep the E register. Note that this proof does not use any property about the event W_C beyond the fact that it is determined by r_{-i}.

(9) **Proof of Proposition 5.1**: All the lemmas and facts required for the proof of Proposition 5.1 have been shown to be valid in our setting. One can now simply follow the proof given in [BVY21] to prove the proposition.

D.3. Supplementary arguments for security proof

In this section, we continue the argument from the bound in Eq. 6.88 and show that it can be transformed into a lower bound for $H_{\min}^{O(\mu')}(\hat{A}_1^t|T_1^t I_1^t \Omega_1^n E X_S A_S)_{\rho_{|\neg F}}$. We also upper bound the information leakage during the information reconciliation phase.

D.3.1. Removing $\hat{B}_1^t$ from the smooth min-entropy bound

We begin by removing the $\hat{B}_1^t$ registers from the entropy in $H_{\min}^{\mu'+\epsilon'}(\hat{A}_1^t \hat{B}_1^t|\hat{X}_1^t \hat{Y}_1^t T_1^t I_1^t \Omega_{J^c} E)_{\rho_{|\neg F}}$. For this, it is sufficient to prove that the entropy

$$H_{\max}^{\epsilon'}(\hat{B}_1^t|\hat{A}_1^t \hat{X}_1^t \hat{Y}_1^t T_1^t)_{\rho_{|\neg F}} \tag{D.4}$$

is small. Intuitively, this should be true because the average winning probability is at least $\omega_{\text{th}} \geq 1-\nu$, which implies using Lemma D.1 that

$$\Pr[A = B] \geq 1 - 2\nu - 2\alpha. \tag{D.5}$$

So, we should be able to prove that

$$H_{\max}^{\epsilon'}(\hat{B}_1^t|\hat{A}_1^t \hat{X}_1^t \hat{Y}_1^t T_1^t)_{\rho_{|\neg F}} \lesssim t\ h(2(\nu+\alpha)). \tag{D.6}$$

To prove this, let $\bar{J} := \{j \in J : X_j, Y_j = (0,2)\}$ for the state ρ (unconditioned). Define the events,

$$E_1 := \left[\frac{|S|}{t} \geq \gamma - \delta_1\right] \tag{D.7}$$

$$E_2 := \left[\frac{|\bar{J}|}{t} \geq (1-\alpha)^2(1-\nu) - \delta_1\right] \tag{D.8}$$

$$E_3 := \left[\frac{1}{t}\sum_{i\in J} V(X_i, Y_i, A_i, B_i) \geq \omega_{\text{th}} - \delta_1\right] \tag{D.9}$$

for some small parameter $\delta_1 \in (0,1)$. Using the Chernoff-Hoeffding bound, we have that

$$\Pr[E_1^c] \leq e^{-\Omega(\delta_1^2 t)} \tag{D.10}$$

$$\Pr[E_2^c] \leq e^{-\Omega(\delta_1^2 t)}. \tag{D.11}$$

Following [Vid17] (which uses [TL17, Lemma 6]), we also have that

$$\Pr[\neg F \wedge E_3^c] \leq e^{-\Omega(\delta_1^2 \gamma t)}. \tag{D.12}$$

Therefore, we have that conditioned on the event $\neg F$, $E_1 \wedge E_2 \wedge E_3$ hold except with probability $\frac{e^{-\Omega(\delta_1^2\gamma t)}}{\Pr_\rho(\neg F)}$, i.e.,

$$\frac{1}{2}\left\|\rho_{T_1^t \hat{X}_1^t \hat{Y}_1^t \hat{A}_1^t \hat{B}_1^t|\neg F} - \rho_{T_1^t \hat{X}_1^t \hat{Y}_1^t \hat{A}_1^t \hat{B}_1^t, E_1\wedge E_2 \wedge E_3|\neg F}\right\|_1 \leq \frac{e^{-\Omega(\delta_1^2\gamma t)}}{\Pr_\rho(\neg F)}. \tag{D.13}$$

Let $\delta(x,y)$ be the Kronecker delta function, which is 1 if $x = y$ and 0 otherwise. Let e be the relative error between $\hat{A}_1^t$ and $\hat{B}_1^t$. If the events $E_1 \wedge E_2 \wedge E_3$ are true, then we have

$$\begin{aligned}
\omega_{\text{th}} - \delta_1 &\leq \frac{1}{t}\sum_{i\in J} V(X_i, Y_i, A_i, B_i) \\
&= \frac{1}{t}\sum_{i\in \bar{J}} V(X_i, Y_i, A_i, B_i) + \frac{1}{t}\sum_{i\in J\setminus\bar{J}} V(X_i, Y_i, A_i, B_i) \\
&\leq \frac{1}{t}\sum_{i\in \bar{J}} \delta(A_i, B_i) + \frac{1}{t}\sum_{i\in J\setminus\bar{J}} (1 + \delta(A_i, B_i)) \\
&\leq 1 - e + \frac{t - |\bar{J}|}{t} \\
&\leq 1 - e + 1 - (1-\alpha)^2(1-\nu) + \delta_1 \\
&\leq 1 - e + \nu + 2\alpha + \delta_1
\end{aligned}$$

which implies that

$$e \leq 1 - \omega_{\text{th}} + \nu + 2\alpha + 2\delta_1. \tag{D.14}$$

Further, since $\omega_{\text{th}} \geq 1 - \nu$, we have that

$$e \leq 2(\nu + \alpha + \delta_1). \tag{D.15}$$

This enables us to bound the max-entropy for the state $\rho_{E_1\wedge E_2\wedge E_3|\neg F}$:

$$H_{\max}(\hat{B}_1^t|\hat{A}_1^t)_{\rho_{E_1\wedge E_2\wedge E_3|\neg F}} \leq t \cdot h(2(\nu + \alpha + \delta_1)). \tag{D.16}$$

Combining with Eq. D.13 shows that

$$H_{\max}^{\epsilon''}(\hat{B}_1^t|\hat{A}_1^t)_{\rho_{|\neg F}} \leq t \cdot h(2(\nu + \alpha + \delta_1)) \tag{D.17}$$

for $\epsilon'' := \frac{e^{-\Omega(\delta_1^2\gamma t)}}{\sqrt{\Pr(\neg F)}} = e^{-\Omega(n)}$.

We can use this to bound the entropy of Alice's raw key alone by using the chain rule in [VDTR13, Theorem 15] and Eq. 6.88:

$$\begin{aligned}
&H_{\min}^{\mu'+5\epsilon'}(\hat{A}_1^t|\hat{X}_1^t\hat{Y}_1^t T_1^t I_1^t \Omega_{J^c} E)_{\rho_{|\neg F}} \\
&\quad \geq H_{\min}^{\mu'+\epsilon'}(\hat{A}_1^t\hat{B}_1^t|\hat{X}_1^t\hat{Y}_1^t T_1^t I_1^t \Omega_{J^c} E)_{\rho_{|\neg F}} - H_{\max}^{\epsilon'}(\hat{B}_1^t|\hat{A}_1^t\hat{X}_1^t\hat{Y}_1^t T_1^t I_1^t \Omega_{J^c} E)_{\rho_{|\neg F}} - O\left(\log\frac{1}{\epsilon'}\right) \\
&\quad \geq H_{\min}^{\mu'+\epsilon'}(\hat{A}_1^t\hat{B}_1^t|\hat{X}_1^t\hat{Y}_1^t T_1^t I_1^t \Omega_{J^c} E)_{\rho_{|\neg F}} - H_{\max}^{\epsilon'}(\hat{B}_1^t|\hat{A}_1^t)_{\rho_{|\neg F}} - O\left(\log\frac{1}{\epsilon'}\right) \\
&\quad \geq t\left((1-\alpha)F_{\alpha,\nu}(\omega_{\text{th}}) - O\left(\frac{\sqrt{\mu}}{\nu\gamma}\right) - h(2(\nu+\alpha+\delta_1))\right) - O(1). \qquad \text{(D.18)}
\end{aligned}$$

We chose the smoothing of the max-entropy above to be ϵ' as well for simplicity. Since, ϵ' is a constant greater than 0 and $\epsilon'' = e^{-\Omega(n)}$, it is valid to use the bound in Eq. D.17 for sufficiently large n.

D.3.2. Adding Ω_J to the conditioning register

Lemma D.4. *For $\rho_{\Omega XAE}$ a classical (ΩXA)-quantum (E) state which satisfies the Markov chain $\Omega \leftrightarrow X \leftrightarrow AE$ and an event F determined by X and A, i.e., $F \subseteq \mathcal{X}\times\mathcal{A}$, the conditional state $\rho_{\Omega XAE|F}$ also satisfies $\Omega \leftrightarrow X \leftrightarrow AE$.*

Proof. Since $\rho_{\Omega XAE}$ is a classical (ΩXA)-quantum (E) state and satisfies the Markov chain $\Omega \leftrightarrow X \leftrightarrow AE$, ρ is of the form

$$\rho_{\Omega XAE} = \sum_x \rho(x)[\![x]\!] \otimes \left(\sum_\omega \rho(\omega|x)[\![\omega]\!]\right) \otimes \left(\sum_a \rho(a|x)[\![a]\!] \otimes \rho_{E|a,x}\right). \qquad \text{(D.19)}$$

Let $\rho(F) = \sum_{x,a\in F}\rho(x)\rho(a|x)$ be the probability of the event F. The conditional state $\rho_{|F}$ can be written as

$$\begin{aligned}
\rho_{\Omega XAE|F} &= \frac{1}{\rho(F)}\rho_{\Omega XAE\wedge F} \\
&= \frac{1}{\rho(F)}\sum_{x,a\in F}\rho(x)\rho(a|x)[\![x,a]\!] \otimes \rho_{E|a,x} \otimes \left(\sum_\omega \rho(\omega|x)[\![\omega]\!]\right) \\
&= \sum_x \rho(x|F)[\![x]\!] \otimes \left(\sum_a \rho(a|x,F)[\![a]\!] \otimes \rho_{E|a,x}\right) \otimes \left(\sum_\omega \rho(\omega|x)[\![\omega]\!]\right)
\end{aligned}$$

which clearly satisfies the Markov chain $\Omega \leftrightarrow X \leftrightarrow AE$. □

Note that using Lemma 6.5, we have that the state ρ (unconditioned) satisfies the Markov chain

$$\Omega_J \leftrightarrow JX_JY_J \leftrightarrow A_1^n B_1^n \Omega_{J^c} E T_1^t. \qquad \text{(D.20)}$$

More precisely, according to Lemma 6.5 the above is true for every fixed J, and the above follows from using simple facts about Markov chains. Further, the event $\neg F$ is defined using the variables $JX_JY_JA_JB_JT_1^t$, so using Lemma D.4, we have that the state $\rho_{|\neg F}$ also satisfies the above Markov chain condition. Note that tracing over $A_{J^c}B_{J^c}$ does not alter this.

Therefore, there exists a channel $\Phi : JX_JY_J \to JX_JY_J\Omega_J$ such that

$$\Phi\left(\rho_{|\neg F}^{A_JB_JX_JY_JJ\Omega_{J^c}T_1^tE}\right) = \rho_{|\neg F}^{A_JB_JX_JY_JJ\Omega_1^nT_1^tE}. \tag{D.21}$$

This implies that

$$H_{\min}^{\mu'+5\epsilon'}(A_J|JX_JY_JT_1^t\Omega_1^nE)_{\rho_{|\neg F}} = H_{\min}^{\mu'+5\epsilon'}(A_J|JX_JY_JT_1^t\Omega_{J^c}E)_{\rho_{|\neg F}}. \tag{D.22}$$

We have already bounded the right-hand side above in Eq. D.18.

D.3.3. Accounting for information leakage during testing

Finally, we also need to consider the entropy loss due to Alice transmitting A_S in plaintext to Bob. Using the chain rule [VDTR13, Theorem 14] we have that

$$\begin{aligned} &H_{\min}^{\mu'+8\epsilon'}(A_J|X_JY_JT_1^tJ\Omega_1^nEA_S)_{\rho_{|\neg F}} \\ &\geq H_{\min}^{\mu'+5\epsilon'}(A_J|X_JY_JT_1^tJ\Omega_1^nE)_{\rho_{|\neg F}} - H_{\max}^{\epsilon'}(A_S|X_JY_JT_1^tJ\Omega_1^nE)_{\rho_{|\neg F}} - O\left(\log\frac{1}{\epsilon'}\right). \end{aligned} \tag{D.23}$$

We can show that the max-entropy above is small for sufficiently large n. Since, T_i are chosen in an i.i.d. fashion with probability $P(T_i = 1) = \gamma$, we have that with probability at least $1 - e^{-\Omega(\gamma^2 t)}$, the number of T_i that are 1 is at most $2\gamma t$. Let's call this event Q. We then have

$$\begin{aligned} \rho_{JX_JY_JA_JB_JT_1^t} &= \rho_{JX_JY_JA_JB_J} \otimes \rho_{T_1^t} \\ &\approx_{e^{-\Omega(\gamma^2 t)}} \rho_{JX_JY_JA_JB_J} \otimes \rho_{T_1^t \wedge Q}. \end{aligned}$$

Let $\eta_{JX_JY_JA_JB_JT_1^t} := \rho_{JX_JY_JA_JB_J} \otimes \rho_{T_1^t \wedge Q}$. Using Lemma A.14, we have that

$$\frac{1}{2}\left\|\rho_{X_JY_JA_JB_JT_1^t|\neg F} - \eta_{X_JY_JA_JB_JT_1^t|\neg F}\right\|_1 \leq \frac{e^{-\Omega(\gamma^2 t)}}{\Pr_\rho(\neg F)} \tag{D.24}$$

and hence $P(\rho_{X_JY_JA_JB_JT_1^t|\neg F}, \eta_{X_JY_JA_JB_JT_1^t|\neg F}) \leq e^{-\Omega(\gamma^2 t)}/\sqrt{\Pr_\rho(\neg F)}$. Note that for a fixed value of $T_1^t = \tau_1^t$, the state $\eta_{A_S|\tau_1^t}$ has a support of size at most $|\mathcal{A}|^{2\gamma t}$. Therefore, for $\epsilon' = \Omega(1) \geq e^{-\Omega(\gamma^2 t)}/\sqrt{\Pr_\rho(\neg F)} = e^{-\Omega(n)}$, we have that

$$H_{\max}^{\epsilon'}(A_S|JX_JY_JT_1^t\Omega_1^nE)_{\rho_{|\neg F}} \leq 2\gamma t \log|\mathcal{A}|.$$

Plugging this in Eq. D.23 and using Eq. D.18 and D.22, we get the bound

$$H_{\min}^{\mu'+8\epsilon'}(A_J|JX_JY_JT_1^t\Omega_1^nEA_S)_{\rho_{|\neg F}}$$

$$\geq t\left((1-\alpha)F_{\alpha,\nu}(\omega_{\text{th}}) - O\left(\frac{\sqrt{\mu}}{\nu\gamma}\right) - h(2(\nu+\alpha+\delta_1)) - 2\log|\mathcal{A}|\gamma\right) - O(1) \tag{D.25}$$

Note that

$$H_{\min}^{\mu'+8\epsilon'}(A_J|JT_1^t\Omega_1^n EX_S A_S)_{\rho_{|\neg F}} \geq H_{\min}^{\mu'+8\epsilon'}(A_J|JX_JY_JT_1^t\Omega_1^n EA_S)_{\rho_{|\neg F}}. \tag{D.26}$$

Using the bound in Eq. D.25 in the equation above, we have a linear lower bound for the smooth min-entropy of Alice's raw key (A_J) with respect to Eve's state ($JT_1^t\Omega_1^n EX_S A_S$):

$$\begin{aligned} &H_{\min}^{\mu'+8\epsilon'}(A_J|JT_1^t\Omega_1^n EX_S A_S)_{\rho_{|\neg F}} \\ &\quad\geq t\left((1-\alpha)F_{\alpha,\nu}(\omega_{\text{th}}) - O\left(\frac{\sqrt{\mu}}{\nu\gamma}\right) - h(2(\nu+\alpha+\delta_1)) - 2\log|\mathcal{A}|\gamma\right) - O(1) \end{aligned} \tag{D.27}$$

D.3.4. Information reconciliation cost

In the one-shot setting, the information reconciliation cost, denoted as leak_{IR}, is given by $H_{\max}^{\epsilon''}(\hat{A}_1^t|\hat{B}_1^t I_1^t)_{\rho_{|\neg F}}$ up to a constant (see [**AF20**, Section 4.2.2] for additional details).

We have already bound the entropy $H_{\max}^{\epsilon''}(\hat{B}_1^t|\hat{A}_1^t I_1^t)_{\rho_{|\neg F}}$ in Eq. D.17. The same argument and bound can also be used for the entropy above. So, we have

$$\text{leak}_{\text{IR}} \leq H_{\max}^{\epsilon''}(\hat{A}_1^t|\hat{B}_1^t I_1^t)_{\rho_{|\neg F}} + O(1) \leq t\cdot h(2(\nu+\alpha+\delta_1)) + O(1) \tag{D.28}$$

for $\epsilon'' := \frac{e^{-\Omega(\delta_1^2\gamma t)}}{\sqrt{\Pr(\neg F)}} = e^{-\Omega(n)}$.

D.3.5. Key length

Up to a constant factor the key length is given by

$$\begin{aligned} &H_{\min}^{\mu'+8\epsilon'}(A_J|JT_1^t\Omega_1^n EX_S A_S)_{\rho_{|\neg F}} - \text{leak}_{\text{IR}} \\ &\quad\geq t\left((1-\alpha)F_{\alpha,\nu}(\omega_{\text{th}}) - O\left(\frac{\sqrt{\mu}}{\nu\gamma}\right) - 2h(2(\nu+\alpha+\delta_1)) - 2\log|\mathcal{A}|\gamma\right) - O(1) \end{aligned} \tag{D.29}$$

where $\alpha,\nu \in (0,0.1)$ are parameters for the $3\text{CHSH}_\perp$ game, and the rest of the parameters are chosen as:

$$\delta \in (0,1) \tag{D.30}$$

$$t = \frac{\delta}{\log|\mathcal{A}||\mathcal{B}| + \delta} n \tag{D.31}$$

$$\epsilon = O\left(\frac{\delta^{1/16}}{\alpha^3}\right) \tag{D.32}$$

$$\mu = O\left(\epsilon^{1/6}\left(\log\frac{1}{\epsilon}\right)^{1/3}\right) \tag{D.33}$$

$$\mu' := 2\sqrt{\frac{\mu}{\Pr_\rho(\neg F)}} = O\left(\epsilon^{1/12}\left(\log\frac{1}{\epsilon}\right)^{1/6}\right) \tag{D.34}$$

$$\delta_1 \in (0,1) \tag{D.35}$$

and $\epsilon' = \Omega(1) \in (0,1)$ such that $\mu' + \epsilon' < 1$. We have assumed here that $\Pr(\neg F) \geq 2\mu$.