Smooth min-entropy lower bounds for approximation chains

Ashutosh Marwah and Frédéric Dupuis

Département d'informatique et de recherche opérationnelle,
Université de Montréal,
Montréal QC, Canada

August 2, 2023

Abstract

For a state $\rho_{A_1^n B}$, we call a sequence of states $(\sigma^{(k)}_{A_1^k B})_{k=1}^n$ an approximation chain if for every $1 \leq k \leq n$, $\rho_{A_1^k B} \approx_\epsilon \sigma^{(k)}_{A_1^k B}$. In general, it is not possible to lower bound the smooth min-entropy of such a $\rho_{A_1^n B}$, in terms of the entropies of $\sigma^{(k)}_{A_1^k B}$ without incurring very large penalty factors. In this paper, we study such approximation chains under additional assumptions. We begin by proving a simple triangle inequality, which allows us to bound the smooth min-entropy of a state in terms of the Rényi entropy of an arbitrary auxiliary state while taking into account the smooth max-relative entropy between the two. Using this triangle inequality, we create lower bounds for the smooth min-entropy of a state in terms of the entropies of its approximation chain in various scenarios. In particular, utilizing this approach, we prove an approximate version of entropy accumulation and also provide a solution to the source correlation problem in quantum key distribution.

Contents

1 Introduction

One-shot information theory investigates the behaviour of tasks in communication and cryptography under general unstructured processes, as opposed to independent and identically distributed (i.i.d) processes, where the states or the tasks themselves have a certain tensor product structure. This is crucial for information theoretically secure cryptography, where one cannot place any kind of assumption on the actions of the adversary (see, for example, [TLGR12, KWW12]). To prove security for such protocols, a common strategy is to show that some smooth min-entropy is sufficiently large. For this reason, the smooth min-entropy [Ren06, RK05] is one of the most important quantities in one-shot information theory.

The smooth min-entropy $H_{\min}^{\epsilon}(K|E)_\rho$ for the classical-quantum state $\rho = \sum_k p(k) |k\rangle\langle k| \otimes \rho_{E|k}$ characterises the amount of randomness one can extract from the classical register K independent of the adversary's register E [TRSS10]. It behaves very differently from the von Neumann conditional entropy, which characterises tasks in the i.i.d setting, and the difference between the two can be very large. 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.

An important and interesting argument, which works with the von Neumann conditional entropy but fails with the smooth min-entropy, is that of proving lower bounds on the entropy using an *approximation chain*. We call a sequence of states[(1)] $(\sigma^{(k)}_{A_1^k B})_{k=1}^n$ an ϵ-approximation chain for the state $\rho_{A_1^n B}$ if for every k, we can approximate the partial state $\rho_{A_1^k B}$ as $\|\rho_{A_1^k B} - \sigma^{(k)}_{A_1^k B}\|_1 \leq \epsilon$. If one can further prove that these states satisfy $H(A_k|A_1^{k-1}B)_{\sigma^{(k)}} \geq c$ for some $c > 0$ sufficiently large, then the following simple argument shows that $H(A_1^n|B)_\rho$ is large:

$$\begin{aligned}
H(A_1^n|B)_\rho &= \sum_{k=1}^n H(A_k|A_1^{k-1}B)_\rho \\
&\geq \sum_{k=1}^n \left(H(A_k|A_1^{k-1}B)_{\sigma^{(k)}} - g(\epsilon)\right) \\
&\geq n(c - g(\epsilon))
\end{aligned}$$

where we used continuity of the von Neumann conditional entropy in the second line ($g(\epsilon) = O(\epsilon \log \frac{|A|}{\epsilon})$ is a "small" function of ϵ). It is well known that a similar argument is not possible with the smooth min-entropy. Consequently, identities for the smooth min-entropy, like the chain rules [VDTR13], are much more restrictive. Tools like entropy accumulation [DFR20, MFSR22] also seem 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. In this paper, we consider scenarios consisting of approximation chains, similar to the above, along with additional conditions and prove lower bounds on the appropriate smooth min-entropies.

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

$$\frac{1}{2}\left\|\rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1} B}\right\|_1 \leq \epsilon. \tag{1}$$

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

(1) For n quantum registers $(X_1, X_2, \cdots, X_n)$, the notation X_i^j refers to the set of registers $(X_i, X_{i+1}, \cdots, X_j)$.

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)(2) 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 difference 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 $n \gg \frac{1}{\epsilon}$ (3). In this paper, we show how one can instead use a bound on the smooth max-relative entropy between these two states 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. 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

$$\rho_{AB} \approx_\delta \tilde{\rho}_{AB} \leq 2^d \eta_{AB} \leq 2^{-(H_{\min}(A|B)_\eta - d)} \mathbb{1}_A \otimes \sigma_B$$

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

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

We call this a *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)

$$H^{\epsilon+\delta}_{\min}(A|B)_\rho \geq \tilde{H}^\uparrow_\alpha(A|B)_\eta - \frac{\alpha}{\alpha - 1} D^\epsilon_{\max}(\rho_{AB} \| \eta_{AB}) - \frac{O(1)}{\alpha - 1} \tag{2}$$

which is valid for $\epsilon + \delta < 1$ and $1 < \alpha \leq 2$. Our general strategy for the scenarios considered in this paper 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

(2) 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 $B = 0$.

(3) 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 chosen to be equal to A_{i-1}. In this case, Q_{A_k} is the uniformly random distribution for bits and Eq. 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 $\left\| Q_{A_1^{2n} B_1^{2n}} - \prod_{i=1}^{2n} Q_{A_i} \cdot Q_{B_1^{2n}} \right\|_1 \to 2$.

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. 2 above to reduce the problem of bounding the smooth min-entropy on state ρ to that of bounding a α-Rényi entropy on the state η. Using this strategy, in Corollary 4.4, we prove that for states satisfying the approximately independent registers assumptions we have for $\delta = O(\epsilon \log \frac{|A|}{\epsilon})$ that

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

In the second scenario, we consider approximate entropy accumulation. 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 $\rho_{R_0 E}$ to produce the state $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E})$. 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. This ensures that the register B_k does not reveal any additional information about A_1^{k-1} than what was previously revealed by $B_1^{k-1}E$. The entropy accumulation theorem [DFR20], then provides a tight lower bound for the smooth min-entropy $H_{\min}^{\delta}(A_1^n|B_1^n E)$. 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 [Eke91, AFDF+18], which are 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 kth step satisfies

$$\rho_{A_1^k B_1^k E R_k} = \mathcal{M}_k(\rho_{A_1^{k-1} B_1^{k-1} E R_{k-1}}) \approx_\epsilon \mathcal{M}'_k(\rho_{A_1^{k-1} B_1^{k-1} E R_{k-1}}) := \sigma^{(k)}_{A_1^k B_1^k E R_k}. \tag{4}$$

Moreover, the assumptions on the channel $\mathcal{M}'_k$ guarantee that the state $\sigma^{(k)}_{A_1^k B_1^k E R_k}$ satisfies the Markov chain condition $A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k$, and so the chain rules and bounds used for entropy accumulation apply for it too. Roughly speaking, we use the chain rules for divergences [FF21] to show that the divergence distance between the states $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E})$ and the virtual state $\sigma_{A_1^n B_1^n E} = \mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(\rho_{R_0 E})$ is relatively small, and then reduce the problem of lower bounding the smooth min-entropy of $\rho_{A_1^n B_1^n E}$ to that of lower bounding an α-Rényi entropy of $\sigma_{A_1^n B_1^n E}$, which can be done by using the chain rules developed for entropy accumulation[(4)]. In Theorem 5.1, 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 $\delta > 0$

$$H_{\min}^{\delta}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^n \inf_{\omega_{R_k \tilde{R}_k}} H(A_k|B_k \tilde{R}_k)_{\mathcal{M}'_k(\omega_{R_k \tilde{R}_k})} - nO(\epsilon^{\frac{1}{24}}) - O\left(\frac{1}{\epsilon^{\frac{1}{24}}}\right) \tag{5}$$

(4) 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.

where the infimum is over all possible input states $\omega_{R_k \tilde{R}_k}$, and the dimensions $|A|$ and $|B|$ are assumed constant while using the asymptotic notation.

We also use the techniques developed above to provide a solution for the source correlation problem in quantum key distribution (QKD) [PCLN+22]. For security proofs of QKD protocols, it is assumed that the states produced by Alice's source are independent in each round. However, in practical implementations this is not entirely true, since physical devices have an internal quantum memory, which may cause the states across multiple rounds to be correlated with each other. The challenge is to prove security for QKD with such an imperfect and correlated source. We show that it is possible to securely implement QKD by simply measuring the output of the source in the preparation basis for a small random set of indices and conditioning on the relative deviation of the observed output being less than some small threshold ϵ from the expected output. Using the results of [BF10], this source test guarantees with high probability that the relative frequency of errors, or the average error per round, in the conditioned state is $\approx \epsilon$. We can once again show that the final state of the QKD protocol implemented on this state is only $nf(\epsilon)$ (for some small function $f(\epsilon)$) far in smooth max-relative entropy distance from the final state of the protocol if it were conducted on perfect states. This allows us to reduce the security proof under a correlated source to that of the QKD protocol which uses perfect states. In Theorem 6.3, we show that a BB84 protocol, which tests its source in the above manner is secure and the error loss due to source correlations is $\approx O(\sqrt{h(\epsilon)})$ (where h is the binary entropy) per round. We also consider the source test with imperfect measurements and demonstrate how these may be taken into account in the analysis.

Lastly, we note that the sections on approximate entropy accumulation (Sec. 5) and source correlations (Sec. 6) are independent of each other and can be read as such.

2 Background and Notation

For n quantum registers $(X_1, X_2, \cdots, X_n)$, the notation X_i^j refers to the set of registers $(X_i, X_{i+1}, \cdots, X_j)$. We use the notation [n] to denote the set $\{1, 2, \cdots, n\}$. For a 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. A quantum state refers to a positive semidefinite operator with unit trace. We will denote the set of registers a quantum state describes (equivalently, its Hilbert space) using a subscript. For example, a quantum state on the register 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. The identity operator on register A is denoted using $\mathbb{1}_A$. 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 normalized quantum states on register B.

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 write $\mathrm{supp}(X)$ to denote the support of the Hermitian operator X and use $X \ll Y$ to denote that $\mathrm{supp}(X) \subseteq \mathrm{supp}(Y)$.

The trace norm is defined as $\|X\|_1 := \mathrm{tr}\left(\left(X^\dagger X\right)^{\frac{1}{2}}\right)$. The fidelity between two positive operators P and Q is defined as $F(P,Q) = \left\|\sqrt{P}\sqrt{Q}\right\|_1^2$. The generalised fidelity between two subnormalised states ρ and σ is defined as

$$F_*(\rho,\sigma) := \left(\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 + \sqrt{(1-\mathrm{tr}\,\rho)(1-\mathrm{tr}\,\sigma)}\right)^2. \tag{6}$$

The purified distance between two subnormalised states ρ and σ is defined as

$$P(\rho,\sigma) = \sqrt{1-F_*(\rho,\sigma)}. \tag{7}$$

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

$$\|\mathcal{N}_{A\to B}\|_\diamond := \max_{X_{AR}:\|X_{AR}\|_1\leq 1} \|\mathcal{N}_{A\to B}(X_{AR})\|_1 \tag{8}$$

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

Throughout this paper, we use base 2 for both the functions log and exp. We follow the notation in Tomamichel's book [Tom16] for Rényi entropies. For $\alpha \in (0,1)\cup(1,2)$, the Petz α-Rényi relative entropy between the positive operators P and Q is defined as

$$\bar{D}_\alpha(P\|Q) = \begin{cases} \frac{1}{\alpha-1}\log\mathrm{tr}\,\frac{(P^\alpha Q^{1-\alpha})}{\mathrm{tr}(P)} & \text{if } (\alpha<1 \text{ and } P \not\perp Q) \text{ or } (P \ll Q) \\ \infty & \text{else.} \end{cases} \tag{9}$$

The sandwiched α-Rényi relative entropy for $\alpha \in (0,1)\cup(1,\infty)$ between the positive operator P and Q is defined as

$$\tilde{D}_\alpha(P\|Q) = \begin{cases} \frac{1}{\alpha-1}\log\frac{\mathrm{tr}(Q^{-\frac{\alpha'}{2}}PQ^{-\frac{\alpha'}{2}})^\alpha}{\mathrm{tr}(P)} & \text{if } (\alpha<1 \text{ and } P \not\perp Q) \text{ or } (P \ll Q) \\ \infty & \text{else.} \end{cases} \tag{10}$$

where $\alpha' = \frac{\alpha-1}{\alpha}$. In the limit $\alpha\to\infty$, the sandwiched divergence becomes equal to the max-relative entropy, $D_{\max}$, which is defined as

$$D_{\max}(P\|Q) := \inf\left\{\lambda\in\mathbb{R} : P \leq 2^\lambda Q\right\}. \tag{11}$$

In the limit of $\alpha \to 1$, both the Petz and the sandwiched relative entropies equal the quantum relative entropy, $D(P\|Q)$, which is defined as

$$D(P\|Q) := \begin{cases} \frac{\operatorname{tr}(P\log P - P\log Q)}{\operatorname{tr}(P)} & \text{if } (P \ll Q) \\ \infty & \text{else.} \end{cases} \tag{12}$$

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]

$$\mathbb{D}(\mathcal{N}\,\|\,\mathcal{M}) := \sup_{\rho_{AR}} \mathbb{D}(\mathcal{N}_{A\to B}(\rho_{AR})\|\,\mathcal{M}_{A\to B}(\rho_{AR})) \tag{13}$$

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

We can use the divergences defined above to define the following conditional entropies for the subnormalized state ρ_{AB}:

$$\begin{aligned} \bar{H}^{\uparrow}_{\alpha}(A|B)_{\rho} &:= \sup_{\sigma_B} -\bar{D}_{\alpha}(\rho_{AB}\|\,\mathbb{1}_A \otimes \sigma_B) \\ \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\rho} &:= \sup_{\sigma_B} -\tilde{D}_{\alpha}(\rho_{AB}\|\,\mathbb{1}_A \otimes \sigma_B) \\ \bar{H}^{\downarrow}_{\alpha}(A|B)_{\rho} &:= -\bar{D}_{\alpha}(\rho_{AB}\|\,\mathbb{1}_A \otimes \rho_B) \\ \tilde{H}^{\downarrow}_{\alpha}(A|B)_{\rho} &:= -\tilde{D}_{\alpha}(\rho_{AB}\|\,\mathbb{1}_A \otimes \rho_B) \end{aligned}$$

for appropriate α in the domain of the divergences. The supremum in the definition for $\bar{H}^{\uparrow}_{\alpha}$ and $\tilde{H}^{\uparrow}_{\alpha}$ is over all quantum states σ_B on register B.

For $\alpha \to 1$, all these conditional entropies are equal to the von Neumann conditional entropy $H(A|B)$. $\tilde{H}^{\uparrow}_{\infty}(A|B)_{\rho}$ is usually called the min-entropy. The min-entropy is usually denoted as $H_{\min}(A|B)_{\rho}$ and for a subnormalised state can also be defined as

$$H_{\min}(A|B)_{\rho} := \sup\left\{\lambda \in \mathbb{R} : \text{ there exists state } \sigma_B \text{ such that } \rho_{AB} \le 2^{-\lambda}\,\mathbb{1}_A \otimes \sigma_B\right\}. \tag{14}$$

For the purpose of smoothing, define the ϵ-ball around the subnormalised state ρ as the set

$$B_{\epsilon}(\rho) = \{\tilde{\rho} \ge 0 : P(\rho, \tilde{\rho}) \le \epsilon \text{ and } \operatorname{tr}\tilde{\rho} \le 1\}. \tag{15}$$

We define the smooth max-relative entropy as

$$D^{\epsilon}_{\max}(\rho\|\sigma) = \min_{\tilde{\rho}\in B_{\epsilon}(\rho)} D_{\max}(\tilde{\rho}\|\sigma) \tag{16}$$

The smooth min-entropy of ρ_{AB} is defined as

$$H^{\epsilon}_{\min}(A|B)_{\rho} = \max_{\tilde{\rho}\in B_{\epsilon}(\rho)} H_{\min}(A|B)_{\tilde{\rho}}. \tag{17}$$

3 Triangle inequality for the smooth min-entropy

In this section, we derive a simple triangle inequality (Lemma 3.5) for the smooth min-entropy of the form in Eq. 2. This Lemma is a direct consequence of the following triangle inequality for $\tilde{D}_\alpha$.

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

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

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

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

Proof. If $D_{\max}(\rho\|\eta) = \infty$, then both statements are true trivially. Otherwise, we have that $\rho \leq 2^{D_{\max}(\rho\|\eta)}\eta$ 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,

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

where we used the fact that $\mathrm{tr}(f(X))$ 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*

$$\tilde{H}^{\uparrow}_{\alpha,\epsilon}(A|B)_\rho := \max_{\tilde{\rho}_{AB}\in B_\epsilon(\rho_{AB})} \tilde{H}^{\uparrow}_\alpha(A|B)_{\tilde{\rho}}. \tag{18}$$

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

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

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

$$\tilde{D}_\alpha(\tilde{\rho}_{AB}\|\,\mathbb{1}_A \otimes \sigma_B) \leq \tilde{D}_\alpha(\eta_{AB}\|\,\mathbb{1}_A \otimes \sigma_B) + \frac{\alpha}{\alpha-1}D^{\epsilon}_{\max}(\rho_{AB}\|\eta_{AB}) + \frac{1}{\alpha-1}\log\frac{1}{1-\epsilon^2} \tag{19}$$

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 $\tilde{H}^{\uparrow}_{\alpha,\epsilon}(A|B)_\rho \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\tilde{\rho}}$. □

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*

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

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

Proof. First, note that

$$H^{\epsilon+\delta}_{\min}(A|B)_\rho \geq \sup_{\tilde{\rho}\in B_\epsilon(\rho_{AB})} H^{\delta}_{\min}(A|B)_{\tilde{\rho}}. \tag{20}$$

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. 20 is true.

Using this, we have

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

where we have used [DFR20, Lemma B.10][(5)] (originally proven in [TCR09]) 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 and η, we have*

$$H_{\min}^{\epsilon+\delta}(A|B)_\rho \geq \tilde{H}_\alpha^\uparrow(A|B)_\eta - \frac{\alpha}{\alpha-1} D_{\max}^\epsilon(\rho_{AB}\|\eta_{AB}) - \frac{g_1(\delta,\epsilon)}{\alpha-1} \tag{21}$$

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:

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

□

We can use the asymptotic equipartition theorem for smooth min-entropy and max-relative entropy [TCR09, Tom12, TH13] to derive the following novel triangle inequality for the von Neumann conditional entropy. Although, we do not use this inequality in this paper, we believe it is interesting and may prove useful in the future.

Corollary 3.6. *For $\alpha \in (1,2]$ and states ρ_{AB} and η_{AB}, we have that*

$$H(A|B)_\rho \geq \tilde{H}_\alpha^\uparrow(A|B)_\eta - \frac{\alpha}{\alpha-1} D(\rho_{AB}\|\eta_{AB}). \tag{22}$$

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. □

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

4 Approximately independent registers

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

$$\left\| \rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1} B} \right\|_1 \leq \epsilon. \tag{23}$$

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 when the states above are completely classical. 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 4.1. *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. For $S^c := \mathcal{X} \setminus S$, where S is the set defined above we have that

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

which implies that $q(S^c) \leq \epsilon^{\frac{1}{2}}$. Now, the statement of 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 $\Pr_\rho(B_k) \le 2\sqrt{\epsilon}$. 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

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

We can define the bad set $\mathcal{B} := \left\{(a_1^n, b) : L(a_1^n, b) > n\epsilon^{\frac{1}{4}}\right\}$, then 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 $P(\tilde{\rho}_{A_1^n B}, \rho_{A_1^n B}) \le \sqrt{2}\epsilon^{1/8}$. Further, note that for every $(a_1^n, b) \notin \mathcal{B}$, we have

$$\begin{aligned} \rho(a_1^n | b) &= \prod_{k=1}^n \rho(a_k | a_1^{k-1}, b) \\ &= \prod_{k:(a_1^n, b) \notin B_k} \rho(a_k | a_1^{k-1}, b) \prod_{k:(a_1^n, b) \in B_k} \rho(a_k | a_1^{k-1}, b) \\ &\le (1+\sqrt{\epsilon})^n \prod_{k:(a_1^n, b) \notin B_k} \rho_{A_k}(a_k) \\ &\le (1+\sqrt{\epsilon})^n 2^{-n(1-\epsilon^{\frac{1}{4}})H_{\min}(A_1)} \end{aligned}$$

where in the third line we have used the fact that if $(a_1^n, b) \notin B_k$, then $\rho(a_k | a_1^{k-1} b) \le (1+\sqrt{\epsilon})\rho_{A_k}(a_k)$ and in the last line we have used the fact that for $(a_1^k, b) \notin \mathcal{B}$, we have $|\{k \in [n] : (a_1^n, b) \notin B_k\}| = n - L(a_1^n, b) \ge n(1-\epsilon^{\frac{1}{4}})$, that all the states ρ_{A_k} are identical and $2^{-H_{\min}(A_k)} = \max_{a_k} \rho_{A_k}(a_k)$. 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

$$H_{\min}^{\sqrt{2}\epsilon^{\frac{1}{8}}}(A_1^n | B) \ge n(1-\epsilon^{\frac{1}{4}})H_{\min}(A_1) - n\log(1+\sqrt{\epsilon}). \tag{24}$$

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 are unable to generalise the classical argument above to the quantum case, 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 (2) and (3) 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. 23 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 above classical argument is difficult to generalise to quantum states primarily because the quantum equivalents of Lemma 4.1 are not as nice and simple. Further quantum conditional probabilities themselves are also difficult to use. Fortunately, the substate theorem serves as the perfect tool for developing a smooth max-relative entropy bound, which we can then use with the min-entropy triangle inequality. The quantum substate theorem [JRS02,JN11] provides an upper bound on the smooth max relative entropy $D^{\epsilon}_{\max}(\rho\|\sigma)$ between two states in terms of their relative entropy $D(\rho\|\sigma)$.

Theorem 4.2 (Quantum substate theorem [JN11])**.** *Let ρ and σ be two states on the same Hilbert space. Then for any $\epsilon \in (0,1)$, we have*

$$D^{\sqrt{\epsilon}}_{\max}(\rho\|\sigma) \leq \frac{D(\rho\|\sigma)+1}{\epsilon} + \log\frac{1}{1-\epsilon}. \tag{25}$$

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

$$I(X_1 : X_2 : \cdots : X_n)_\rho := D(\rho_{X_1^n}\|\rho_{X_1}\otimes\rho_{X_2}\otimes\cdots\otimes\rho_{X_n}). \tag{26}$$

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:

$$I(X_1 : X_2 : \cdots : X_n)_\rho = \sum_{k=1}^{n} H(X_k)_\rho - H(X_1 \cdots X_n)_\rho \tag{27}$$

$$= \sum_{k=2}^{n} I(X_k : X_1^{k-1}). \tag{28}$$

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. 23 implies that for every $k \in [n]$

$$I(A_k : A_1^{k-1} B)_\rho \leq \epsilon \log |A| + g_2\left(\frac{\epsilon}{2}\right) \tag{29}$$

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 4.3. *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*

$$I(A_k : A_1^{k-1} B)_\rho \leq \epsilon \tag{30}$$

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

$$\begin{aligned} H_{\min}^{\epsilon^{\frac{1}{4}}+\epsilon}(A_1^n|B)_\rho \geq &\sum_{k=1}^{n} H(A_k)_\rho - 3n\epsilon^{\frac{1}{4}} \log(1+2|A|) \\ &- \frac{2\log(1+2|A|)}{\epsilon^{3/4}} - \frac{2\log(1+2|A|)}{\epsilon^{1/4}}\left(\log(1-\sqrt{\epsilon}) + g_1(\epsilon, \epsilon^{\frac{1}{4}})\right) \end{aligned} \tag{31}$$

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

$$\begin{aligned} H_{\min}^{\epsilon^{\frac{1}{4}}+\epsilon}(A_1^n|B)_\rho \geq &\, n\left(H(A_1)_\rho - 3\epsilon^{\frac{1}{4}} \log(1+2|A|)\right) \\ &- \frac{2\log(1+2|A|)}{\epsilon^{3/4}} - \frac{2\log(1+2|A|)}{\epsilon^{1/4}}\left(\log(1-\sqrt{\epsilon}) + g_1(\epsilon, \epsilon^{\frac{1}{4}})\right). \end{aligned} \tag{32}$$

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

$$D_{\max}^{\epsilon^{\frac{1}{4}}}\left(\rho_{A_1^n B}\middle\|\bigotimes_{k=1}^{n}\rho_{A_k}\otimes\rho_B\right) \leq \frac{D(\rho_{A_1^n B}\| \bigotimes_{k=1}^{n}\rho_{A_k}\otimes\rho_B)+1}{\sqrt{\epsilon}} - \log(1-\sqrt{\epsilon})$$
$$\leq n\sqrt{\epsilon} + \frac{1}{\sqrt{\epsilon}} - \log(1-\sqrt{\epsilon}). \tag{33}$$

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:

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

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

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

□

We can now plug the bound in Eq. 29 to derive the following Corollary.

Corollary 4.4. *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*

$$\left\|\rho_{A_1^k B} - \rho_{A_k}\otimes\rho_{A_1^{k-1}B}\right\|_1 \leq \epsilon. \tag{34}$$

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

$$H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H(A_k)_\rho - 3n\delta^{\frac{1}{4}}\log(1+2|A|) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}}\left(\log(1-\sqrt{\delta}) + g_1(\delta,\delta^{\frac{1}{4}})\right) \tag{35}$$

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*

$$H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B)_\rho \geq n\left(H(A_1)_\rho - 3\delta^{\frac{1}{4}}\log(1+2|A|)\right) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}}\left(\log(1-\sqrt{\delta}) + g_1(\delta,\delta^{\frac{1}{4}})\right). \tag{36}$$

4.1 Weak approximate asymptotic equipartition

We can modify the proof of Theorem 4.3 to prove a *weak* approximate asymptotic equipartition property (AEP).

Theorem 4.5. *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*

$$\left\|\rho_{A_1^k B_1^k E} - \rho_{A_k B_k} \otimes \rho_{A_1^{k-1} B_1^{k-1} E}\right\|_1 \leq \epsilon. \tag{37}$$

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

$$H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^{n} H(A_k|B_k)_\rho - 3n\delta^{\frac{1}{4}}\log(1+2|A|) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}}\left(\log(1-\sqrt{\delta}) + g_1(\delta,\delta^{\frac{1}{4}})\right) \tag{38}$$

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*

$$H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B_1^n E)_\rho \geq n\left(H(A_1|B_1)_\rho - 3\delta^{\frac{1}{4}}\log(1+2|A|)\right) \\ - \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}}\left(\log(1-\sqrt{\delta}) + g_1(\delta,\delta^{\frac{1}{4}})\right). \tag{39}$$

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

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

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 4.3, only difference being that now we have $\tilde{H}^{\uparrow}_{\alpha}(A_1^n | B_1^n E)_\eta = \sum_{k=1}^{n} \tilde{H}^{\uparrow}_{\alpha}(A_k | B_k)_\rho$. □

We call this generalisation *weak* because the smoothing term (δ) depends on size of the side information $|B|$. In Sec. E, 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.

4.2 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 will briefly sketch a proof for sequential device independent quantum key distribution (DIQKD) to demonstrate this idea. The protocol for sequential DIQKD used in [AF20] is presented as Protocol 1.

We consider a simple model for DIQKD, where 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 1. The kth 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 kth round, the random variables produced in the previous $k-1$ rounds are $r_{k-1} \coloneqq x_1^{k-1}, y_1^{k-1}, t_1^{k-1}, a_1^{k-1}, b_1^{k-1}$ 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 kth round. Note that Alice's device cannot distinguish whether the CHSH game is played in a round or it is used for key generation. We can further take an average over all the previous round's

Sequential DIQKD protocol

Parameters:

- ω_{exp} is the expected winning probability for the honest implementation of the device
- $n \geq 1$ is the number of rounds in the protocol
- $\gamma \in (0, 1]$ is the fraction of test rounds

Protocol:

1. For every $0 \leq i \leq n$ perform the following steps:
   (a) Alice chooses a random $T_i \in \{0, 1\}$ with $\Pr[T_i = 1] = \gamma$.
   (b) Alice sends T_i to Bob.
   (c) 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 $\{0, 1\}$.
   (d) 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. **Error correction:** Alice and Bob use an error correction procedure, which lets Bob obtain the raw key $\tilde{A}_1^n$ (if the error correction protocol succeeds, then $A_1^n = \tilde{A}_1^n$). In case the error correction 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 1

random variables to derive the probability of winning the kth game

$$\Pr[W_k] = \mathbb{E}_{r_{k-1}}\left[\Pr[W_k|r_{k-1}]\right]. \tag{40}$$

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 $\rho^{(0)}_{E_AE_BE}$ distributed between Alice and Bob at the start of the protocol by Eve has an average winning probability at least ω_{exp}, that is,

$$\frac{1}{n}\sum_{k=1}^{n}\Pr[W_k] \geq \omega_{\text{exp}} - \delta \tag{41}$$

for some small $\delta > 0$. Using standard sampling arguments it can be argued that either this is true or the protocol aborts with high probability.

For any shared state $\sigma_{E_AE_BE}$ (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 [PAB+09] (see [AF20, Lemma 5.3] for the following form)

$$H(A|XYE) \geq f(\omega) = \begin{cases} 1 - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16\omega(\omega-1)+3}\right) & \text{if } \omega \in [\frac{3}{4}, \frac{2+\sqrt{2}}{4}] \\ 0 & \text{if } \omega \in [0, \frac{3}{4}) \end{cases} \tag{42}$$

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 Figure 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

$$1 - f(\omega_{\text{exp}} - \delta) = h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16(\omega_{\text{exp}} - \delta)(\omega_{\text{exp}} - \delta - 1) + 3}\right) \leq \epsilon^4. \tag{43}$$

We will now use Eq. 42 to bound the von Neumann entropy of the answers given Eve's

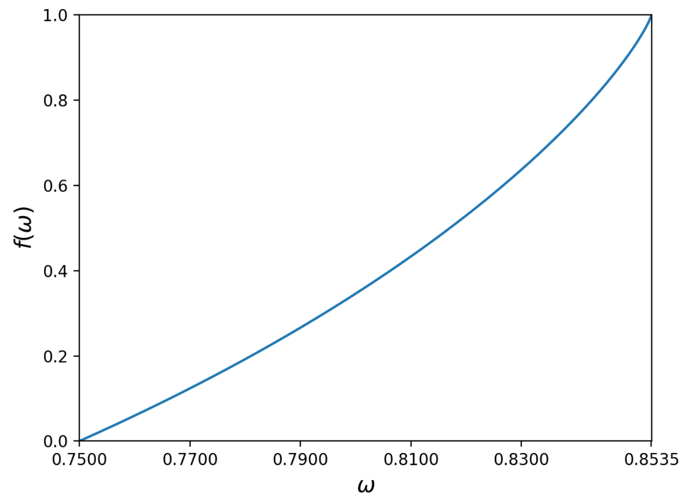

Figure 1: The lower bound in Eq. 42 for the interval $[\frac{3}{4}, \frac{2+\sqrt{2}}{4}]$

information for the sequential DIQKD protocol. We have

$$\begin{aligned}
H(A_1^n|X_1^nY_1^nT_1^nE) &= \sum_{k=1}^n H(A_k|A_1^{k-1}X_1^nY_1^nT_1^nE) \\
&\overset{(1)}{=} \sum_{k=1}^n H(A_k|A_1^{k-1}X_1^kY_1^kT_1^kE) \\
&\overset{(2)}{=} \sum_{k=1}^n H(A_k|X_kY_kR_{k-1}E) \\
&= \sum_{k=1}^n \mathbb{E}_{r_{k-1}\sim R_{k-1}}\left[H(A_k|X_kY_kE)_{\rho^{(k)}_{|r_{k-1}}}\right] \\
&\overset{(3)}{\geq} \sum_{k=1}^n \mathbb{E}_{r_{k-1}\sim R_{k-1}}\left[f\left(\Pr[W_k|r_{k-1}]\right)\right] \\
&\geq \sum_{k=1}^n f\left(\Pr[W_k]\right) \\
&\geq nf\left(\frac{1}{n}\sum_{k=1}^n \Pr[W_k]\right) \\
&\geq nf(\omega_{\text{exp}} - \delta) \geq n(1-\epsilon^4)
\end{aligned}$$

where in (1) we have used the fact that the questions sampled in the rounds after the kth 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. 42, 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, MFSR22] 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.

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

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

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 4.3. We can simply use the smooth min-entropy bound derived there here as well. This gives us the bound

$$\begin{aligned} H_{\min}^{2\epsilon}(A_1^n | X_1^n Y_1^n T_1^n E) &\geq \sum_{k=1}^{n} H(A_k) - 3n\epsilon \log 5 - O\left(\frac{1}{\epsilon^3}\right) \\ &= n(1 - 3\epsilon \log 5) - O\left(\frac{1}{\epsilon^3}\right) \end{aligned} \tag{44}$$

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.

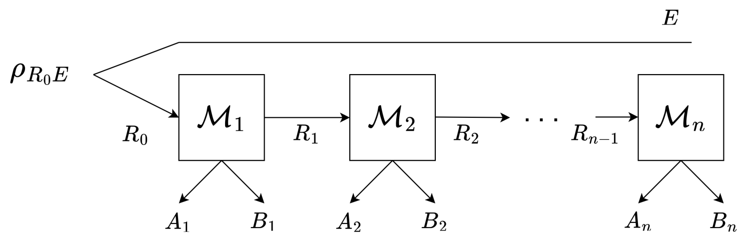

Figure 2: The setting for entropy accumulation and Theorem 5.1. For $k \in [n]$, the channels $\mathcal{M}_k$ are repeatedly applied to the registers R_{k-1} to produce the "secret" information A_k and the side information B_k.

5 Approximate entropy accumulation

In general, it is very difficult to estimate the smooth min-entropy of states produced during cryptographic protocols. 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 Figure 2. The parties implementing the protocol begin 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 $k \in [n]$ 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. 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 kth 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 the following lower bound for the smooth min-entropy

$$H^{\epsilon}_{\min}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^{n} \inf_{\omega_{R_{k-1}R}} H(A_k|B_k R)_{\mathcal{M}_k(\omega_{R_{k-1}R})} - c\sqrt{n} \tag{45}$$

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 Figure 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 a chain of approximations. For each $k \in [n]$, we have

$$\rho_{A_1^k B_1^k E} = \mathrm{tr}_{R_k} \circ \mathcal{M}_k(\rho_{A_1^{k-1} B_1^{k-1} E}) \approx_\epsilon \mathrm{tr}_{R_k} \circ \mathcal{M}'_k(\rho_{A_1^{k-1} B_1^{k-1} E}) =: \sigma^{(k)}_{A_1^k B_1^k E}.$$

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 $H_{\min}^{\epsilon'}(A_1^n | B_1^n E)_\rho$ 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. For example, the entropy accumulation theorem can be used to prove the security of sequential device independent quantum key distribution (DIQKD) protocols [AFDF+18]. 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 some physical effect 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 kth 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 the applications of this theorem for future work.

Theorem 5.1. *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*

$$\rho_{A_1^n B_1^n E} = \mathrm{tr}_{R_n} \circ \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}) \tag{46}$$

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:*

$$\frac{1}{2} \left\| \mathcal{M}_k - \mathcal{M}'_k \right\|_\diamond \leq \epsilon \tag{47}$$

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 $\mathcal{M}'_k \circ \mathcal{N}_{k-1} \circ \cdots \circ \mathcal{N}_1(\rho_{R_0 E})$ satisfies the Markov chain*

$$A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k. \tag{48}$$

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*

$$\begin{aligned} H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^n \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}'_k(\omega_{R_k\tilde{R}_k})} - n(\alpha-1)\log^2(1+2|A|) \\ - \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(2^{\frac{\alpha-1}{\alpha} 2\log(|A||B|)} - 1\right)\right) \\ - \frac{\alpha}{\alpha-1} n z_\beta(\epsilon,\delta) - \frac{1}{\alpha-1}\left(g_1(\epsilon_2,\epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta-1}\right). \end{aligned} \tag{49}$$

where

$$z_\beta(\epsilon,\delta) := \frac{\beta+1}{\beta-1}\log\left(\left(1+\sqrt{(1-\delta)\epsilon}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\beta}\right)^{\frac{1}{\beta+1}}\right) \tag{50}$$

and $g_1(x,y) = -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$ *and the infimum in Eq. 49 is taken over all input states to the channels* $\mathcal{M}'_k$.

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

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

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

$$\begin{aligned} H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n|B_1^n E)_\rho \geq \sum_{k=1}^n \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}'_k(\omega_{R_k\tilde{R}_k})} \\ - n\sqrt{\epsilon_r}(\log^2(1+2|A|)+2) - \frac{1}{\sqrt{\epsilon_r}}\left(g_1(\epsilon_2,\epsilon_1) + 2g_0(\epsilon_1)\right) \end{aligned} \tag{51}$$

The entropy loss per round in the above bound behaves as $\sim \epsilon^{\frac{1}{24}}$. 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 F.1). 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. 51, 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. 5.4.

In order to prove this theorem, we will use a channel divergence based chain rule. 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

$$\tilde{D}_\alpha(\mathcal{E}_{A\to B}(\rho_A)\|\mathcal{F}_{A\to B}(\sigma_A)) \leq \tilde{D}_\alpha(\rho_A\|\sigma_A) + \tilde{D}_\alpha^{\mathrm{reg}}(\mathcal{E}_{A\to B}\|\mathcal{F}_{A\to B}) \tag{52}$$

where $\tilde{D}_\alpha^{\mathrm{reg}}(\mathcal{E}_{A\to B}\|\mathcal{F}_{A\to B}) \coloneqq \lim_{n\to\infty}\frac{1}{n}\tilde{D}_\alpha(\mathcal{E}_{A\to B}^{\otimes n}\|\mathcal{F}_{A\to B}^{\otimes n})$ and $\tilde{D}_\alpha(\cdot\|\cdot)$ is the channel divergence.

Now observe that if we were guaranteed that for the maps in Theorem 5.1 above, $\tilde{D}_\alpha^{\mathrm{reg}}(\mathcal{M}_k\,\|\,\mathcal{M}'_k) \leq \epsilon$ for every k for some $\alpha > 1$. Then, we could use the chain rule in Eq. 52 as follows

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

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

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

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

In order to prove Theorem 5.1, 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 $\tilde{D}_\alpha^{\mathrm{reg}}$ or any other channel divergence. We will get around this issue by instead using mixed channels $\mathcal{M}_k^\delta \coloneqq (1-\delta)\,\mathcal{M}'_k + \delta\,\mathcal{M}_k$. Also, instead of trying to bound channel divergence in terms of $\tilde{D}_\alpha^{\mathrm{reg}}$, 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 5.3.

5.1 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

$$D^{\#}_{\alpha}(P\|Q) := \min_{A \geq P} \hat{D}_{\alpha}(A\|Q) \tag{53}$$

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

$$\hat{D}_{\alpha}(A\|Q) = \begin{cases} \frac{1}{\alpha-1} \log \operatorname{tr}\left(Q\left(Q^{-\frac{1}{2}} A Q^{-\frac{1}{2}}\right)^{\alpha}\right) & \text{if } A \ll Q \\ \infty & \text{otherwise.} \end{cases} \tag{54}$$

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, Section 2.6.3]).

Lemma 5.2 (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*

$$X \leq (1+t)\Pi X \Pi + \left(1 + \frac{1}{t}\right) \Pi_{\perp} X \Pi_{\perp}. \tag{55}$$

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

$$D_\alpha^{\#}(\rho\|\sigma) \le \frac{\alpha+1}{\alpha-1}\log\left((1+\sqrt{\epsilon})^{\frac{\alpha}{\alpha+1}} + \left(2^{\alpha d}\sqrt{\epsilon}\right)^{\frac{1}{\alpha+1}}\right). \tag{56}$$

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

Further, let

$$\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}} = \sum_{i=1}^n \lambda_i |x_i\rangle\langle x_i| \tag{57}$$

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

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

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

$$S := \{i \in [n] : \lambda_i \le \sqrt{\epsilon}\}. \tag{58}$$

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

$$\begin{aligned}\operatorname{tr}(\sigma\Pi_\perp) &= \sum_{i\in S^c} \langle x_i|\sigma|x_i\rangle \\ &= \Pr_q(I \in S^c) \\ &\leq \sqrt{\epsilon}.\end{aligned} \tag{59}$$

Moreover, by the definition of set S (Eq. 58) we have

$$\begin{aligned}\Pi\sigma^{-\frac{1}{2}}P\sigma^{-\frac{1}{2}}\Pi &= \sum_{i\in S} \lambda_i |x_i\rangle\langle x_i| \\ &\leq \sqrt{\epsilon}\Pi\end{aligned} \tag{60}$$

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

$$\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} \leq 2^d\,\mathbb{1}\,. \tag{61}$$

Now, observe that since $\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} \geq 0$, for an arbitrary $t > 0$, using Lemma 5.2 we have

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

where we have used $\rho \leq \sigma + P$ to bound the first term and Eq. 61 to bound the second term in the second line, and Eq. 60 to bound $\Pi\sigma^{-\frac{1}{2}}P\sigma^{-\frac{1}{2}}\Pi$ in the last step.

We will define $A_t := (1+t)(1+\sqrt{\epsilon})\sigma^{\frac{1}{2}}\Pi\sigma^{\frac{1}{2}} + \left(1+\frac{1}{t}\right)2^d\sigma^{\frac{1}{2}}\Pi_\perp\sigma^{\frac{1}{2}}$. 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:

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

where in the last line we use $\operatorname{tr}(\sigma\Pi)\leq 1$ and $\operatorname{tr}(\sigma\Pi_\perp)\leq\sqrt{\epsilon}$ (Eq. 59). Finally, since $t>0$ was arbitrary, we can choose the $t>0$ which minimizes the right-hand side. For this choice of $t_{\min} = \left(\frac{2^{\alpha d}\sqrt{\epsilon}}{(1+\sqrt{\epsilon})^\alpha}\right)^{\frac{1}{\alpha+1}}$, we get

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

which proves the required bound. □

5.2 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:

$$\begin{aligned}
\frac{1}{2}\|\mathcal{M}_\delta-\mathcal{N}\|_\diamond &= \frac{1}{2}\|(1-\delta)\mathcal{M}+\delta\mathcal{N}-\mathcal{N}\|_\diamond \\
&= (1-\delta)\frac{1}{2}\|\mathcal{M}-\mathcal{N}\|_\diamond \\
&\leq (1-\delta)\epsilon. \qquad (62)
\end{aligned}$$

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. 62 have that

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

Using Lemma 5.3, 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 5.4. *Let $\epsilon > 0$. Suppose channels $\mathcal{N}$ and $\mathcal{M}$ from register A to B are such that $\frac{1}{2}\left\|\mathcal{N} - \mathcal{M}\right\|_\diamond \leq \epsilon$. 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*

$$D_\alpha^{\#}(\mathcal{N} \| \mathcal{M}_\delta) \leq \frac{\alpha+1}{\alpha-1} \log\left(\left(1+\sqrt{(1-\delta)\epsilon}\right)^{\frac{\alpha}{\alpha+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\alpha}\right)^{\frac{1}{\alpha+1}}\right). \tag{63}$$

5.3 Proof of the approximate entropy accumulation theorem

We use the mixed channels defined in the previous section to define the auxiliary state $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E})$ for our proof. It is easy to show using the divergence bounds in Sec. 5.2 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 $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E})$ 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 5.1. Using Lemma 5.4, 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

$$\begin{aligned} D_\beta^\#(\mathcal{M}_k \,\|\, \mathcal{M}_k^\delta) &\leq \frac{\beta+1}{\beta-1} \log\left(\left(1+\sqrt{(1-\delta)\epsilon}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\beta}\right)^{\frac{1}{\beta+1}}\right) \\ &:= z_\beta(\epsilon,\delta) \end{aligned} \tag{64}$$

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

$$\sigma_{A_1^nB_1^nE} := \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E}). \tag{65}$$

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

$$\begin{aligned} &D_{\max}^{\epsilon_1}(\rho_{A_1^nB_1^nE} \| \sigma_{A_1^nB_1^nE}) \\ &\quad\leq \tilde{D}_\beta(\rho_{A_1^nB_1^nE} \| \sigma_{A_1^nB_1^nE}) + \frac{g_0(\epsilon_1)}{\beta-1} \\ &\quad\leq D_\beta^\#(\rho_{A_1^nB_1^nE} \| \sigma_{A_1^nB_1^nE}) + \frac{g_0(\epsilon_1)}{\beta-1} \\ &\quad= D_\beta^\#(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0E}) \| \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E})) + \frac{g_0(\epsilon_1)}{\beta-1} \\ &\quad\leq D_\beta^\#(\mathcal{M}_{n-1} \circ \cdots \circ \mathcal{M}_1(\rho_{R_0E}) \| \mathcal{M}_{n-1}^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0E})) + D_\beta^\#(\mathcal{M}_n \,\|\, \mathcal{M}_n^\delta) + \frac{g_0(\epsilon_1)}{\beta-1} \\ &\quad\leq \cdots \\ &\quad\leq \sum_{k=1}^n D_\beta^\#(\mathcal{M}_k \,\|\, \mathcal{M}_k^\delta) + \frac{g_0(\epsilon_1)}{\beta-1} \\ &\quad\leq n z_\beta(\epsilon,\delta) + \frac{g_0(\epsilon_1)}{\beta-1} \end{aligned} \tag{66}$$

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. 64.

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

$$\begin{aligned} H^{\epsilon_1+\epsilon_2}_{\min}(A_1^n|B_1^n E)_\rho \geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_\sigma - \frac{\alpha}{\alpha-1} n z_\beta(\epsilon, \delta) \\ - \frac{1}{\alpha-1}\left(g_1(\epsilon_2, \epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta - 1}\right). \end{aligned} \tag{67}$$

We have now reduced our problem to lower bounding $\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_\sigma$. Note that we cannot directly use the entropy accumulation here, since the mixed maps $\mathcal{M}^{\delta}_k = (1-\delta)\,\mathcal{M}'_k + \delta\,\mathcal{M}_k$, 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 $\mathcal{M}^{\delta}_k$ 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 $\theta_{A_1^n B_1^n C_1^n E} = \mathcal{N}_n \circ \cdots \circ \mathcal{N}_1(\rho_{R_0 E})$. 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

$$\begin{aligned} \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_\sigma &= \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_\theta \\ &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n C_1^n E)_\theta. \end{aligned} \tag{68}$$

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 entropy 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 5.5. *Define* $h_k := \inf_{\omega_{R_k \tilde{R}_k}} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_k\tilde{R}_k)_{\mathcal{M}'_k(\omega_{R_k\tilde{R}_k})}$ *where the infimum is over all states* $\omega_{R_k\tilde{R}_k}$ *for a register* $\tilde{R}_k$*, which is isomorphic to* R_k*, and* $s := \log(|A||B|^2)$*. Then, we have*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_{\theta_{|c_1^n}} \geq \sum_{k=1}^{n} \left(\delta(c_k,1)h_k - \delta(c_k,0)s\right) \tag{69}$$

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 D.1, in the second line we have used the dimension bound in Lemma D.3 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 $\theta_{R_{k-1}A_1^{k-1}B_1^{k-1}E|c_1^k} = \Phi_{k-1} \circ \Phi_{k-2} \cdots \circ \Phi_1(\rho_{R_0 E})$ 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 C.5 (the $\tilde{H}^{\uparrow}_{\alpha}$ counterpart for [DFR20, Corollary 3.5], which is the main chain rule used for proving entropy accumulation), we have

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

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

$$\tilde{H}_\alpha^\uparrow(A_1^k|B_1^kE)_{\theta_{|c_1^k}} \geq \tilde{H}_\alpha^\uparrow(A_1^{k-1}|B_1^{k-1}E)_{\theta_{|c_1^{k-1}}} + (\delta(c_k,1)h_k - \delta(c_k,0)s)\,. \tag{70}$$

Using this bound n times starting with $\tilde{H}_\alpha^\uparrow(A_1^n|B_1^nE)_{\theta_{|c_1^n}}$ 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

$$\begin{aligned}
\sum_{c_1^n}\theta(c_1^n)\exp\left(\frac{1-\alpha}{\alpha}\tilde{H}_\alpha^\uparrow(A_1^n|B_1^nE)_{\theta_{|c_1^n}}\right) &\leq \sum_{c_1^n}\theta(c_1^n)\exp\left(\frac{1-\alpha}{\alpha}\sum_{k=1}^n l_k(c_k)\right)\\
&= \sum_{c_1^n}\prod_{k=1}^n\theta(c_k)2^{\frac{1-\alpha}{\alpha}l_k(c_k)}\\
&= \prod_{k=1}^n\sum_{c_k}\theta(c_k)2^{\frac{1-\alpha}{\alpha}l_k(c_k)}.
\end{aligned} \tag{71}$$

Then, we have

$$\begin{aligned}
\tilde{H}_\alpha^\uparrow(A_1^n|B_1^nC_1^nE)_\theta &= \frac{\alpha}{1-\alpha}\log\sum_{c_1^n}\theta(c_1^n)\exp_2\left(\frac{1-\alpha}{\alpha}\tilde{H}_\alpha^\uparrow(A_1^n|B_1^nE)_{\theta_{|c_1^n}}\right).\\
&\geq \frac{\alpha}{1-\alpha}\sum_{k=1}^n\log\sum_{c_k}\theta(c_k)2^{\frac{1-\alpha}{\alpha}l_k(c_k)}\\
&= \frac{\alpha}{1-\alpha}\sum_{k=1}^n\log\left((1-\delta)2^{\frac{1-\alpha}{\alpha}h_k} + \delta2^{-\frac{1-\alpha}{\alpha}s}\right)\\
&= \sum_{k=1}^n h_k - \frac{\alpha}{\alpha-1}\sum_{k=1}^n\log\left(1-\delta+\delta2^{\frac{\alpha-1}{\alpha}(s+h_k)}\right)\\
&\geq \sum_{k=1}^n h_k - \frac{\alpha}{\alpha-1}n\log\left(1+\delta\left(2^{\frac{\alpha-1}{\alpha}(s+\log|A|)}-1\right)\right)
\end{aligned}$$

where in the second line we have used Eq. 71 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

$$\begin{aligned}
\tilde{H}_\alpha^\uparrow(A_1^n|B_1^nC_1^nE)_\theta \geq \sum_{k=1}^n\inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}'_k(\omega_{R_k\tilde{R}_k})} - n(\alpha-1)\log^2(1+2|A|)\\
- \frac{\alpha}{\alpha-1}n\log\left(1+\delta\left(2^{\frac{\alpha-1}{\alpha}2\log(|A||B|)}-1\right)\right).
\end{aligned} \tag{72}$$

Putting Eq. 67, Eq. 68, and Eq. 72 together, we have

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

□

5.4 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 this theorem. The classical version of this theorem has a much better dependence of $O(\sqrt{\epsilon})$ on ϵ (see Theorem F.1). The reason for the poor performance of the quantum version is that our bound on the channel divergence (Lemma 5.4) is very weak compared to the bound we can use classically. It should be noted, however, that if Lemma 5.4 were to be improved in the future, one could simply plug the new bound into our proof and derive an improvement for Theorem 5.1.

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 $(1-\delta)\mathcal{M}'_k + \delta\tau_{A_kB_k}\otimes \mathrm{tr}_{A_kB_k}\circ\mathcal{M}_k$ (which is what is done classically), where $\tau_{A_kB_k}$ is the completely mixed state on registers A_kB_k. Here, instead of mixing the channel $\mathcal{M}'_k$ with $\mathcal{M}_k$, we mix it with $\tau_{A_kB_k}\otimes\mathrm{tr}_{A_kB_k}\circ\mathcal{M}_k$, 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 $\mathcal{M}_n\circ\cdots\circ\mathcal{M}_1(\rho_{R_0E})$ and the auxiliary state $\mathcal{M}_n^\delta\circ\cdots\circ\mathcal{M}_1^\delta(\rho_{R_0E})$ 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 5.3 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 F.1) 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 E.

6 Source Correlations

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 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 $(k+1)$th round, then it is plausible that the state produced in the $(k+1)$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.

We will consider the BB84-QKD protocol here. An extensive line of research has led to techniques for the proof of such a QKD protocol [SP00, Ren06, TL17, DFR20, MR22]. However, almost all of these techniques rely on *source purification*[(6)]- 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 the above methods. In this section, we will use the triangle inequality of Lemma 3.5 to reduce the security of the BB84-QKD protocol with source correlations to that of the BB84 protocol with a perfect source. Then, one can simply use the security analysis methods developed for such protocols to complete the security proof. Although, we focus on the

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

BB84-QKD protocol here, our technique is quite general, and we believe it can also be applied to other cryptographic protocols.

Our method relies on randomly testing the output of the source on a small sample by measuring it in the preparation basis and conditioning on the relative deviation of the observed output being less than some small threshold ϵ from the expected output. We do not have to place any assumptions on the source, except that it passes this source test with a non-negligible probability. We also demonstrate how our analysis can be modified to incorporate imperfect measurements. 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 kth 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 can not be guaranteed in practical setups. Importantly, it is also not possible to estimate the parameter ℓ_c experimentally. In contrast, we provide a simpler and more general technique, which reduces the security proof to that of a noisy version of the underlying protocol(7), and itself only requires assumptions for the measurements employed for testing, which are used at a far smaller rate than the source itself during the protocol.

6.1 Security proof for BB84 with source correlations

The BB84 QKD protocol has been described in Protocol 2. In Table 1, we list all the variables, we use for our proof and their definitions.

(7) Following this reduction, any security proof technique for QKD which can bound $\tilde{H}^{\uparrow}_{\alpha}$ 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 of required for this security proof and the assumptions used during our testing procedure.

BB84 QKD protocol

Parameters:

- n is the number of qubits sent by Alice.
- $\mu \in (0,1)$ is the probability of encoding and measurement in the X basis $\{|+\rangle, |-\rangle\}$.
- $e \in (0, \frac{1}{2})$ is the maximum error tolerated.
- $r \in (0,1)$ is the 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 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. **Error correction:** Alice and Bob use an error correction procedure, which lets Bob obtain a guess $\hat{X}_S$ for Alice's raw key X_S. In case the error correction protocol aborts, they abort the QKD protocol too.
4. **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 computes $\hat{e} = \frac{1}{|\mathcal{X}|}\sum_{i \in S_X} \hat{X}_i \oplus Y_i$. They abort if $\hat{e} > e$.
5. **Privacy Amplification:** Alice chooses a random function F from a set of two 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 2

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 Q_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 error correction step.
T	Transcript for error correction
$\bar{X}_1^n$	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., $\mathrm{freq}(\hat{C}_1^n)(1) \leq e$ and $\mathrm{hash}(X_S) = \mathrm{hash}(\hat{X}_S)$.
Υ'	The event that $X_S = \hat{X}_S$
Υ''	The event that $\mathrm{freq}(C_1^n)(1) \leq e$.

Table 1: Definition of variables for QKD

For the protocol above, we use X_i to denote the random bit selected by Alice, Θ_i for the choice of the basis, and A_i to represent the qubit encoded and sent by Alice to Bob in round i. At the beginning of every round Alice prepares the registers X_i, Θ_i, and A_i. A perfect BB84 source produces the following state in every round,

$$\hat{\rho}_{X\Theta A} \coloneqq \sum_{x\in\mathcal{X},\theta\in\Theta} p(x,\theta) \left|x,\theta\right\rangle \left\langle x,\theta\right|_{X\Theta} \otimes H^{\theta} \left|x\right\rangle \left\langle x\right|_{A} H^{\theta} \tag{73}$$

where $\mathcal{X} = \{0,1\}$, $\Theta = \{0,1\}$ and H is the Hadamard gate and

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

Suppose, there are n rounds in the BB84 QKD protocol. To implement the QKD protocol with an imperfect source, we require Alice to perform the preparation and testing protocol

Source preparation and testing

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 subsequent protocols.

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 and 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. Output the registers corresponding to the remaining rounds.

Protocol 3

given in Protocol 3 before using her source for QKD (Note that Alice does not need a quantum memory to perform this testing procedure). Alice produces $(n+m)$ BB84 states and then randomly measure m of them in the preparation basis. We assume that the classical randomness used by Alice is perfect, though this assumption can also be relaxed.

Let the imperfect source produce the state $\rho_{X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$. Using the analysis of quantum sampling strategies given in [BF10], we prove that the state conditioned on passing the test in Protocol 3 has a relatively small smooth max-relative entropy with a depolarised version of the perfect source. The main result of [BF10] has been summarised as the Theorem below.

Theorem 6.1. *The relative weight of a string x_1^k is defined as $\omega(x_1^k) := \frac{1}{k}|\{i \in [k] : x_i \neq 0\}|$. Let Ψ be a sampling strategy which takes a string a_1^n, selects a random subset $\Gamma \subseteq [n]$ with probability p_Γ, and 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*

$$B^\delta_{\gamma\kappa}(\Psi) := \{a_1^n : |\omega(a_{\bar{\gamma}}) - f(\gamma, a_\gamma, \kappa)| < \delta\}\,. \tag{74}$$

The classical maximum error probability for this strategy Ψ is defined as

$$\epsilon_{cl}^{\delta} := \max_{a_1^n} \Pr_{\Gamma K}[a_1^n \notin B_{\Gamma K}^{\delta}(\Psi)]. \tag{75}$$

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

$$\tilde{\rho}_{\Gamma K A_1^n E} := \sum_{\gamma\kappa} p(\gamma\kappa) |\gamma\kappa\rangle\langle\gamma\kappa|_{\Gamma K} \otimes \frac{\Pi_{A_1^n}^{\delta|\gamma\kappa} \rho_{A_1^n E} \Pi_{A_1^n}^{\delta|\gamma\kappa}}{\mathrm{tr}\left(\Pi_{A_1^n}^{\delta|\gamma\kappa} \rho_{A_1^n E}\right)} \tag{76}$$

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.*

Let Ω denote the event that Protocol 3 does not abort and define the unitaries,

$$V_A^{x,\theta} := H^{\theta} X^{x} \tag{77}$$

$$V_{X\Theta A} := \sum_{x,\theta} |x,\theta\rangle\langle x,\theta|_{X\Theta} \otimes V_A^{x,\theta} \tag{78}$$

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

$$\nu_{X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} := \bigotimes_{i=1}^{n+m} V_{X_i\Theta_i A_i}^{\dagger}\ \rho_{X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} \bigotimes_{i=1}^{n+m} V_{X_i\Theta_i A_i}. \tag{79}$$

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 6.1. 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. 89). 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 6.1, 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

$$\frac{1}{2}\left\|\nu_{\Gamma A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} - \eta_{\Gamma A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}}\right\|_1 \leq \epsilon_{\mathrm{qu}}^{\delta} \tag{80}$$

and

$$\eta_{\Gamma A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} := \sum_{\gamma} p(\gamma) |\gamma\rangle\langle\gamma| \otimes \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} \tag{81}$$

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

$$\eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} = \Pi^{\delta|\gamma}_{A_1^{n+m}} \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}} \Pi^{\delta|\gamma}_{A_1^{n+m}} \tag{82}$$

for the projectors $\Pi^{\delta|\gamma}_{A_1^{n+m}}$ defined as in Theorem 6.1 (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 $\epsilon^{\delta}_{\mathrm{qu}} = \sqrt{2}\exp(-\frac{n\delta^2}{2(n+2)}m)$. We can also define the extended state $\eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}$ as

$$\begin{aligned}\eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} := \sum_{\gamma,x_1^{n+m},\theta_1^{n+m}} & p(\gamma)p(x_1^{n+m},\theta_1^{n+m}) \\ & |\gamma, x_1^{n+m},\theta_1^{n+m}\rangle\langle\gamma, x_1^{n+m},\theta_1^{n+m}| \otimes \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m},\theta_1^{n+m}}\end{aligned} \tag{83}$$

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

$$\frac{1}{2}\left\|\nu_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} - \eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}\right\| \le \epsilon^{\delta}_{\mathrm{qu}}. \tag{84}$$

Define Ω' to be the event that 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 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 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

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

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

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

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

$$\bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'} = \sum_{a_1^n, \bar{a}_1^n : \in \{x_1^n :\ \omega(x_1^n) < \epsilon + \delta\}} \bar{\sigma}^{(\gamma)}(a_1^n, \bar{a}_1^n) \left|a_1^n\right\rangle \left\langle \bar{a}_1^n\right|. \tag{85}$$

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

$$\bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'} = Q^{\epsilon+\delta}_{A_1^n} \bar{\sigma}^{(\gamma)}_{A_1^n \wedge \Omega'} Q^{\epsilon+\delta}_{A_1^n} \tag{86}$$

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

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

Using the data processing inequality, we also have that

$$\frac{1}{2} \left\| \bar{\nu}_{\Gamma X_1^n \Theta_1^n A_1^n \wedge \Omega'} - \bar{\eta}_{\Gamma X_1^n \Theta_1^n A_1^n \wedge \Omega'} \right\|_1 \leq \epsilon^\delta_{\mathrm{qu}} \tag{87}$$

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

$$\min_{z_1^n : \omega(z_1^n) < \epsilon+\delta} \left\langle z_1^n\right| (\hat{\eta}^{(\epsilon+\delta)}_A)^{\otimes n} \left|z_1^n\right\rangle \geq 2^{-nh(\epsilon+\delta)}$$

which implies that

$$Q^{\epsilon+\delta}_{A_1^n} \leq 2^{nh(\epsilon+\delta)} (\hat{\eta}^{(\epsilon+\delta)}_A)^{\otimes n}.$$

Thus, we have

$$\bar{\eta}_{X_1^n \Theta_1^n A_1^n \wedge \Omega'} \leq 2^{nh(\epsilon+\delta)} \left(\rho_{X\Theta} \otimes \hat{\eta}^{(\epsilon+\delta)}_A \right)^{\otimes n}. \tag{88}$$

As noted earlier, the state produced by measuring the registers A_γ of ν in the computational basis is the same as 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 $\bigotimes_{k=1}^{n} V^{\dagger}_{X_i\Theta_i A_i}$. Under this correspondence, we have

$$\bar{\rho}_{X_1^n\Theta_1^n A_1^n\wedge\Omega} = \bigotimes_{i=1}^{n} V_{X_i\Theta_i A_i}\bar{\nu}_{X_1^n\Theta_1^n A_1^n\wedge\Omega'}\bigotimes_{i=1}^{n} V^{\dagger}_{X_i\Theta_i A_i}. \tag{89}$$

Further, for the state defined as

$$\begin{aligned}\bar{\bar{\rho}}_{X_1^n\Theta_1^n A_1^n\wedge\Omega} &\coloneqq \bigotimes_{i=1}^{n} V_{X_i\Theta_i A_i}\bar{\eta}_{X_1^n\Theta_1^n A_1^n\wedge\Omega'}\bigotimes_{i=1}^{n} V^{\dagger}_{X_i\Theta_i A_i} \\ &\le 2^{nh(\epsilon+\delta)}\bigotimes_{i=1}^{n}\left(V_{X_i\Theta_i A_i}\ \rho_{X_i\Theta_i}\otimes\hat{\eta}^{(\epsilon+\delta)}_{A_i}V^{\dagger}_{X_i\Theta_i A_i}\right) \\ &= 2^{nh(\epsilon+\delta)}\left(\hat{\rho}^{(\epsilon+\delta)}_{X\Theta A}\right)^{\otimes n}\end{aligned} \tag{90}$$

where $\hat{\rho}^{(\epsilon+\delta)}_{X\Theta A} \coloneqq (1-2(\epsilon+\delta))\hat{\rho}_{X\Theta A} + 2(\epsilon+\delta)\hat{\rho}_{X\Theta}\otimes\tau_A$ for the mixed state τ_A on register A. Using Eq. 87, we also have

$$\frac{1}{2}\left\|\bar{\rho}_{X_1^n\Theta_1^n A_1^n\wedge\Omega} - \bar{\bar{\rho}}_{X_1^n\Theta_1^n A_1^n\wedge\Omega}\right\|_1 \le \epsilon^{\delta}_{\mathrm{qu}}. \tag{91}$$

The following Lemma helps use relate distances while conditioning states.

Lemma 6.2. *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 $\frac{1}{2}\|\rho_{XA}-\tilde{\rho}_{XA}\|_1 \le \epsilon$. Then, for $x\in\mathcal{X}$ such that $p(x)>0$, we have*

$$\frac{1}{2}\left\|\rho_{A|x} - \tilde{\rho}_{A|x}\right\|_1 \le \frac{2\epsilon}{p(x)} \tag{92}$$

Proof.

$$\frac{1}{2}\|\rho_{XA}-\tilde{\rho}_{XA}\|_1 = \frac{1}{2}\sum_{x\in\mathcal{X}}\left\|p(x)\rho_{A|x} - \tilde{p}(x)\tilde{\rho}_{A|x}\right\|_1 \le \epsilon$$

This implies that for $x\in\mathcal{X}$

$$\frac{1}{2}\left\|p(x)\rho_{A|x} - \tilde{p}(x)\tilde{\rho}_{A|x}\right\|_1 \le \epsilon$$

and

$$\frac{1}{2}|p(x)-\tilde{p}(x)| \le \epsilon.$$

Using these inequalities, we have

$$\begin{aligned}\frac{1}{2}\left\|\rho_{A|x}-\tilde{\rho}_{A|x}\right\|_1 &\leq \frac{1}{2}\left\|\rho_{A|x}-\frac{\tilde{p}(x)}{p(x)}\tilde{\rho}_{A|x}\right\|_1+\frac{1}{2}\left|1-\frac{\tilde{p}(x)}{p(x)}\right|\left\|\tilde{\rho}_{A|x}\right\|_1\\ &=\frac{1}{p(x)}\frac{1}{2}\left\|p(x)\rho_{A|x}-\tilde{p}(x)\tilde{\rho}_{A|x}\right\|_1+\frac{1}{p(x)}\frac{1}{2}\left|p(x)-\tilde{p}(x)\right|\\ &\leq\frac{2\epsilon}{p(x)}.\end{aligned}$$

□

Using the Lemma above, we have

$$\frac{1}{2}\left\|\bar{\rho}_{X_1^n\Theta_1^nA_1^n|\Omega}-\bar{\tilde{\rho}}_{X_1^n\Theta_1^nA_1^n|\Omega}\right\|\leq\frac{2\epsilon_{\mathrm{qu}}^{\delta}}{\Pr_\rho(\Omega)} \tag{93}$$

where $\Pr_\rho(\Omega)\coloneqq\mathrm{tr}\left(\bar{\rho}_{X_1^n\Theta_1^nA_1^n\wedge\Omega}\right)$ is the probability of the event Ω when the testing procedure is applied to the state ρ, and

$$\begin{aligned}\bar{\tilde{\rho}}_{X_1^n\Theta_1^nA_1^n|\Omega} &\leq \frac{2^{nh(\epsilon+\delta)}}{\Pr_{\tilde{\rho}}(\Omega)}\left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)}\right)^{\otimes n}\\ &\leq\frac{2^{nh(\epsilon+\delta)}}{\Pr_\rho(\Omega)-\epsilon_{\mathrm{qu}}^{\delta}}\left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)}\right)^{\otimes n}\end{aligned} \tag{94}$$

where $\Pr_{\tilde{\rho}}(\Omega)\coloneqq\mathrm{tr}\left(\bar{\tilde{\rho}}_{X_1^n\Theta_1^nA_1^n\wedge\Omega}\right)$ is defined similar to $\Pr_\rho(\Omega)$. Together these imply that

$$D_{\max}^{\epsilon_f}(\bar{\rho}_{X_1^n\Theta_1^nA_1^n|\Omega}||\left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)}\right)^{\otimes n})\leq nh(\epsilon+\delta)+\log\frac{1}{\Pr_\rho(\Omega)-\epsilon_{\mathrm{qu}}^{\delta}} \tag{95}$$

where $\epsilon_f=2\sqrt{\frac{\epsilon_{\mathrm{qu}}^{\delta}}{\Pr_\rho(\Omega)}}$.

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 Section G. 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[8]

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

[8] We also need to condition on the QKD protocol not aborting. We do this in Sec. G

Let us define the virtual state $\sigma_{X_1^n \Theta_1^n A_1^n} := \left(\hat{\rho}_{X\Theta A}^{(\epsilon+\delta)}\right)^{\otimes n}$. 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

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

Thus, it is sufficient to bound the α-Rényi conditional entropy $\tilde{H}_\alpha^\uparrow(X_S|ET\Theta_1^n\hat{\Theta}_1^n)$ 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 σ. In Sec. G, we show how one can modify the security proof for BB84-QKD based on entropy accumulation to get the following bound.

Theorem 6.3. *Suppose Alice uses the output of Protocol 3 (with error threshold ϵ) as her source for the BB84 QKD protocol. Let $\delta > 0$ and assume that $h(\epsilon+\delta) < \frac{1}{\sqrt{2}}$. Then, for*

$$\epsilon_{qu}^\delta = \sqrt{2}\exp\left(-\frac{n\delta^2}{2(n+2)}m\right) \tag{96}$$

$$\epsilon_{pa} = 2\left(\frac{2\epsilon_{qu}^\delta}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'')}\right)^{1/2} \tag{97}$$

and $\epsilon' > 0$, the following lower bound on the smooth min-entropy

$$\begin{aligned} &H_{\min}^{\epsilon_{pa}+\epsilon'}(X_S|E\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{QKD}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\ &\quad\geq n(1-2\mu-h(e)-V\sqrt{2h(\epsilon+\delta)}) - \sqrt{n}\left(\mu^2\ln(2) + 2\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + g_0\left(\frac{\epsilon'}{8}\right)\right) \\ &\quad - \frac{V}{\sqrt{2h(\epsilon+\delta)}}\left(\log\frac{1}{\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'')-2\epsilon_{qu}^\delta}+1\right) - \frac{g_1(\frac{\epsilon'}{2},\epsilon_{pa})}{2\sqrt{2h(\epsilon+\delta)}}V - \log|T| - 3g_0\left(\frac{\epsilon'}{8}\right) \end{aligned} \tag{98}$$

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 $\Pr_{\bar{\rho}}(\Omega\wedge\Upsilon'') > 2\epsilon_{qu}^\delta$, $g_0(x) = -\log(1-\sqrt{1-x^2})$ and $g_1(x,y) = -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$.

We see above that the asymptotic rate for QKD using an imperfect source is $\approx O(\sqrt{h(\epsilon)})$ lesser than a protocol, which uses a perfect source.

6.2 Imperfect measurements

In our analysis above, we assumed that the measurements used in the source testing procedure above are perfect. It should be noted that if the source produces states at a rate r, then the measurement device is only used at an average rate $\frac{m}{n+m}r$, which is much smaller than r. 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, 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 assumption, define

$$\hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma} := \bigotimes_{i\in\gamma} V^{x_i,\theta_i}_{A_i}\left(\sum_{a_\gamma:\omega(a_\gamma)\leq\epsilon+\epsilon_m} |a_\gamma\rangle\langle a_\gamma|_{A_\gamma}\right)\bigotimes_{j\in\gamma}\left(V^{x_j,\theta_j}_{A_j}\right)^\dagger \tag{99}$$

$$\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma} := \mathbb{1}_{A_1^n} - \hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma} \tag{100}$$

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_1^m in the basis θ_1^m. 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 $\hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma}$ 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 $\xi > 0$, for every collection of states $\{\sigma^{(\gamma)}_{A_\gamma|x_\gamma\theta_\gamma}\}_{\gamma,x_\gamma,\theta_\gamma}$ the measurement elements satisfy

$$\sum_\gamma p(\gamma)\sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma)\,\mathrm{tr}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\sigma^{(\gamma)}_{A_\gamma|x_\gamma\theta_\gamma}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right)\leq\xi. \tag{101}$$

Stated in words, we require that for any collection of states $\{\sigma^{(\gamma)}_{A_\gamma|x_\gamma\theta_\gamma}\}_{\gamma,x_\gamma,\theta_\gamma}$ with a relative weight larger than $\epsilon+\epsilon_m$ (lying in the subspace corresponding to the projector $\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}$), 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, we will again derive a smooth max-relative entropy bound similar to Eq. 95. The smoothing parameter of this relative entropy, 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 3) 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. We leave the details for the specific measurement model for future work.

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 6.1, so that we can compare the relative weight with the string x_1^{n+m} in the basis given by θ_1^{n+m}.

$$\bar{\Pi}^{\delta|\gamma, x_1^{n+m}, \theta_1^{n+m}}_{A_1^{n+m}} := \bigotimes_{i=1}^{n+m} V^{x_i,\theta_i}_{A_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{j=1}^{n+m} (V^{x_i,\theta_i}_{A_i})^\dagger \tag{102}$$

We use the state η defined in the previous section to define the state

$$\tilde{\rho}_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} := \bigotimes_{i=1}^{n+m} V_{X_i\Theta_i A_i}\eta_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} \bigotimes_{i=1}^{n+m} V^\dagger_{X_i\Theta_i A_i}. \tag{103}$$

Using the distance bound proven in Eq. 84 and the definition of ν in Eq. 79, we have

$$\frac{1}{2}\left\|\rho_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}} - \tilde{\rho}_{\Gamma X_1^{n+m}\Theta_1^{n+m}A_1^{n+m}}\right\|_1 \leq \epsilon^\delta_{\text{qu}}$$

The conditional states $\tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}}$ satisfy

$$\begin{aligned}
&\bar{\Pi}^{\delta|\gamma,x_1^{n+m},\theta_1^{n+m}}_{A_1^{n+m}} \tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \bar{\Pi}^{\delta|\gamma,x_1^{n+m},\theta_1^{n+m}}_{A_1^{n+m}} \\
&\quad = \bigotimes_{i=1}^{n+m} V^{x_i,\theta_i}_{A_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{j=1}^{n+m} (V^{x_i,\theta_i}_{A_i})^\dagger \tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \bigotimes_{i=1}^{n+m} V^{x_i,\theta_i}_{A_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{j=1}^{n+m} (V^{x_i,\theta_i}_{A_i})^\dagger \\
&\quad = \bigotimes_{i=1}^{n+m} V^{x_i,\theta_i}_{A_i} \Pi^{\delta|\gamma}_{A_1^{n+m}} \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \Pi^{\delta|\gamma}_{A_1^{n+m}} \bigotimes_{j=1}^{n+m} (V^{x_i,\theta_i}_{A_i})^\dagger \\
&\quad = \bigotimes_{i=1}^{n+m} V^{x_i,\theta_i}_{A_i} \eta^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}} \bigotimes_{j=1}^{n+m} (V^{x_i,\theta_i}_{A_i})^\dagger \\
&\quad = \tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}}
\end{aligned}$$

where we have used the definition of $\tilde{\rho}^{(\gamma)}_{A_1^{n+m}|x_1^{n+m}\theta_1^{n+m}}$ (Eq. 103) in the second equality, and Eq. 82 for the fourth line.

Let the event that the imperfect measurements measure a relative weight less than ϵ be denoted by Ω_{im}. We call the 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 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 \\
&\qquad\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} \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 \\
&\qquad\qquad \Bigg[\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\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] \\
&\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\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\qquad + 2\xi \mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}}
\end{aligned}$$

where we have used the pinching inequality (Lemma 5.2 with $t = 1$) in the second line and defined the state $\mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}}$ as the normalization of the state

$$\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle \left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\ \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma)\,\mathrm{tr}_{A_\gamma}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right).$$

Note that

$$\begin{aligned}
&\mathrm{tr}\Bigg(\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle \left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma)\,\mathrm{tr}_{A_\gamma}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right)\Bigg) \\
&= \sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma)\,\mathrm{tr}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right) \\
&= \sum_{\gamma} p(\gamma) \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma)\,\mathrm{tr}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma}\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\left(\sum_{x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(x_{\bar{\gamma}},\theta_{\bar{\gamma}})\tilde{\rho}^{(\gamma)}_{A_\gamma|x_1^{n+m}\theta_1^{n+m}}\right)\hat{P}^{\perp|x_\gamma,\theta_\gamma}_{A_\gamma}\right) \\
&\leq \xi,
\end{aligned}$$

which follows from our assumption about the measurements (Eq. 101). Therefore, we have

$$\begin{aligned}
\tilde{\rho}'_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega_{\mathrm{im}}} &\leq 2\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle \left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma)\,\mathrm{tr}_{A_\gamma}\left(\Lambda\left(\leq\epsilon|\gamma,x_\gamma,\theta_\gamma\right)_{A_\gamma}\hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma}\tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}}\hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma}\right) \\
&\qquad + 2\xi\mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}} \\
&\leq 2\sum_{\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}} p(\gamma)p(x_{\bar{\gamma}},\theta_{\bar{\gamma}}) \left|\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right\rangle \left\langle\gamma,x_{\bar{\gamma}},\theta_{\bar{\gamma}}\right| \otimes \\
&\qquad \sum_{x_\gamma,\theta_\gamma} p(x_\gamma,\theta_\gamma)\,\mathrm{tr}_{A_\gamma}\left(\hat{P}^{x_\gamma,\theta_\gamma}_{A_\gamma}\tilde{\rho}^{(\gamma)}_{A_\gamma A_{\bar{\gamma}}|x_1^{n+m}\theta_1^{n+m}}\right) + 2\xi\mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}} \\
&\leq 2\bar{\bar{\rho}}^{(\epsilon+\epsilon_m)}_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega_{\mathrm{im}}} + 2\xi\mu_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}}
\end{aligned}$$

where the state $\bar{\bar{\rho}}^{(\epsilon+\epsilon_m)}_{\Gamma X_{\bar{\Gamma}}\Theta_{\bar{\Gamma}}A_{\bar{\Gamma}}\wedge\Omega_{\mathrm{im}}}$ 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 threshold for the relative weight of the perfect measurement in last section was ϵ. Thus, we can use the previously derived bound in Eq. 90 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. 90, we get

$$\begin{aligned}\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}} &\le 2\bar{\tilde{\rho}}^{(\epsilon+\epsilon_m)}_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}} + 2\xi\mu_{X_1^n\Theta_1^nA_1^n}\\ &\le 2^{nh(\epsilon+\epsilon_m+\delta)+1}\left(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A}\right)^{\otimes n} + 2\xi\mu_{X_1^n\Theta_1^nA_1^n} \end{aligned}\tag{104}$$

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

$$\frac{1}{2}\left\|\rho'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}} - \tilde{\rho}'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}}\right\|_1 \le \epsilon^\delta_{\text{qu}}.\tag{105}$$

Using Lemma 6.2, the conditional states satisfy

$$\frac{1}{2}\left\|\rho'_{X_1^n\Theta_1^nA_1^n|\Omega_{\text{im}}} - \tilde{\rho}'_{X_1^n\Theta_1^nA_1^n|\Omega_{\text{im}}}\right\|_1 \le \frac{2\epsilon^\delta_{\text{qu}}}{P_\rho(\Omega_{\text{im}})}\tag{106}$$

for $P_\rho(\Omega_{\text{im}}) := \operatorname{tr}\left(\rho'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}}\right)$, defined as the probability that the Protocol 3 does not abort with the imperfect measurements and

$$\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n|\Omega_m} \le \frac{2^{nh(\epsilon+\epsilon_m+\delta)+1}}{P_{\tilde{\rho}}(\Omega_m)}\left(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A}\right)^{\otimes n} + \frac{4\xi}{P_{\tilde{\rho}}(\Omega_m)}\frac{\mu_{X_1^n\Theta_1^nA_1^n}}{2}.\tag{107}$$

where $P_{\tilde{\rho}}(\Omega_{\text{im}}) := \operatorname{tr}\left(\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n\wedge\Omega_{\text{im}}}\right)$. For $0<\mu<1$, the hypothesis testing relative entropy [WR12] is defined as

$$D^\mu_h(\rho\|\sigma) := -\inf\left\{\log\operatorname{tr}(\sigma Q) : 0\le \mu Q\le \mathbb{1}, \text{ and } \operatorname{tr}(\rho Q)\ge 1\right\}.\tag{108}$$

Equivalently, using semidefinite programming duality (see [Wat20]) it can be shown that

$$D^\mu_h(\rho\|\sigma) = -\sup\left\{\log(\lambda - \operatorname{tr}(Y)) : Y\ge 0, \lambda\ge 0, \text{ and } \lambda\rho\le\sigma+\mu Y\right\}\tag{109}$$

$$= \inf\left\{\log\lambda' - \log(1-\operatorname{tr}(Z)) : Z\ge 0, \lambda'\ge 0, \text{ and } \rho\le\lambda'\sigma+\mu Z\right\}.\tag{110}$$

Thus, Eq. 107 implies

$$D^\mu_h(\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n|\Omega_m}\|(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A})^{\otimes n}) \le nh(\epsilon+\epsilon_m+\delta) + 2 + \log\frac{1}{P_{\tilde{\rho}}(\Omega_m)}\tag{111}$$

for $\mu := \frac{4\xi}{P_{\tilde{\rho}}(\Omega_m)}$. Using [Wat20, Theorem 5.11] (originally proven in [ABJT19]), this implies that[(9)]

$$D^{\sqrt{\mu}}_{\max}(\tilde{\rho}'_{X_1^n\Theta_1^nA_1^n|\Omega_m}\|(\hat{\rho}^{(\epsilon+\epsilon_m+\delta)}_{X\Theta A})^{\otimes n}) \le nh(\epsilon+\epsilon_m+\delta) + 2 + \log\frac{1}{P_{\tilde{\rho}}(\Omega_m)} + \log\frac{1}{\mu(1-\mu)}\tag{112}$$

(9) The smoothing for $D^\epsilon_{\max}(\rho\|\sigma)$ 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.

Using the triangle inequality, we can state this in terms of the real state $\tilde{\rho}'_{X_1^n\Theta_1^n A_1^n|\Omega_m}$

$$D_{\max}^{\epsilon_f}(\rho'_{X_1^n\Theta_1^n A_1^n|\Omega_m}||\left(\hat{\rho}_{X\Theta A}^{(\epsilon+\epsilon_m+\delta)}\right)^{\otimes n}) \leq nh(\epsilon+\epsilon_m+\delta)+2+\log\frac{1}{P_\rho(\Omega_m)-\epsilon_{\mathrm{qu}}^\delta} + \log\frac{1}{4\xi(P_\rho(\Omega_m)-\epsilon_{\mathrm{qu}}^\delta-4\xi)} \tag{113}$$

for $\epsilon_f := \frac{2\xi^{1/2}}{\sqrt{P_\rho(\Omega_m)-\epsilon_{\mathrm{qu}}^\delta}} + 2\sqrt{\frac{\epsilon_{\mathrm{qu}}^\delta}{P_\rho(\Omega_m)}}$. 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 seems that it should be possible to improve this dependence). One can use the above bound in place of Eq. 95 to prove a smooth min-entropy lower bound for the QKD protocol.

6.3 Discussion and future work

Theorem 6.3 gives a simple bound on the smooth min-entropy relevant for the QKD protocol in Protocol 2. With a source error of ϵ, the rate of the QKD protocol decreases by $\sim O((\epsilon\log\frac{1}{\epsilon})^{1/2})$ and the privacy amplification error can be made arbitrarily small assuming perfect measurements are used for the source test. For imperfect measurements, under a very broad assumption, we showed that the rate decrease is similar to the perfect case and the privacy amplification error depends on the error of the measurements. The measurement error too can be made arbitrarily small under further reasonable physical assumptions, like independence of the measurement errors. We leave the details of such a physical model and its relation to our assumption on the measurements for future work. It should also be noted that if the source is known to pass the source test with a high probability (which can be made arbitrarily close to 1), say $1-\epsilon_s$, then the source test need not even be performed for the QKD protocol. The error ϵ_s can simply be added to QKD error.

Acknowledgments

We would like to thank Omar Fawzi for interesting discussions and for pointing out Lemma B.2. We also thank Ernest Tan and Amir Arqand, whose observations helped improve Theorem 5.1. We are also grateful to Ernest Tan and Shlok Ashok Nahar, who explained the source correlation problem for QKD to us during and after the QKD Security Proof Workshop at the Institute of Quantum Computing, Waterloo and also for their comments on a manuscript of this paper. AM was supported by the J.A. DeSève Foundation.

APPENDICES

A 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

$$H_{\min}^{\epsilon'}(A|B)_\rho \geq H_{\min}^{\epsilon}(A|B)_\eta - O(D_{\max}^{\epsilon''}(\rho||\eta)) \tag{114}$$

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_1^n is chosen to be a random n-bit string if $B = 1$ otherwise A_1^n 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 $H_{\min}^{\epsilon}(A|B)_p = n$ (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. 114, were true then we would have

$$\begin{aligned} n - O\left(\log \frac{1}{\epsilon}\right) &\leq H_{\min}^{\epsilon}(A|B)_P - O(D_{\max}^{\epsilon''}(p_{AB|E}||p_{AB})) \\ &\leq H_{\min}^{\epsilon'}(A|B)_{p_{|E}} = O(1) \end{aligned}$$

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 3.6 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

$$0 \geq (1 - \epsilon)n - \log \frac{1}{\epsilon}$$

which is not true for large n.

B Bounds for $D_\alpha^\#$ of the form in Lemma 5.3 necessarily diverge in the limit $\alpha = 1$

Classically, we have the following bound for Rényi entropies.

Lemma B.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*

$$D_\alpha(p\|q) \leq \frac{1}{\alpha-1}\log\Big((1+\sqrt{\epsilon})^{\alpha-1}(1-2\sqrt{\epsilon}) + 2^{d(\alpha-1)+1}\sqrt{\epsilon}\Big). \tag{115}$$

In the limit, $\alpha \to 1$, we get the bound

$$D(p\|q) \leq (1-2\sqrt{\epsilon})\log(1+\sqrt{\epsilon}) + 2\sqrt{\epsilon}d. \tag{116}$$

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 4.1. Thus, for $\alpha > 1$ we have

$$\begin{aligned}
\sum_{x\in\mathcal{X}} p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} &= \sum_{x\in S} p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} + \sum_{x\notin S} p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} \\
&\leq \sum_{x\in S}(1+\sqrt{\epsilon})^{\alpha-1}p(x) + \sum_{x\notin S} 2^{d(\alpha-1)}p(x) \\
&= (1+\sqrt{\epsilon})^{\alpha-1}p(S) + 2^{d(\alpha-1)}p(S^c) \\
&\leq (1+\sqrt{\epsilon})^{\alpha-1}(1-2\sqrt{\epsilon}) + 2^{d(\alpha-1)+1}\sqrt{\epsilon}
\end{aligned}$$

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 Section 5.1 that the bound in Lemma 5.3 for $D_\alpha^\#$ tends to ∞ as $\alpha \to 1$ for a fixed $\epsilon > 0$. One may wonder if a bound like Eq. 116 exists for $\lim_{\alpha\to 1} D_\alpha^\#(\rho\|\sigma) = \hat{D}(\rho\|\sigma)$ [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 $\frac{1}{2}\|\rho-\sigma\|_1 \leq \epsilon$ and $\rho \leq 2^d\sigma$, the following bound holds

$$\hat{D}(\rho\|\sigma) \leq f(\epsilon, d) \tag{117}$$

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. 116 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 B.2. [10] *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

$$\begin{aligned}\rho &:= |0\rangle\langle 0| = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\\ \sigma'_\epsilon &:= (\sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|1\rangle)(\sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|1\rangle)^\dagger\\ &= \begin{pmatrix}1-\epsilon & \sqrt{\epsilon(1-\epsilon)}\\ \sqrt{\epsilon(1-\epsilon)} & \epsilon\end{pmatrix}\\ \sigma_\epsilon &:= (1-\delta)\sigma'_\epsilon + \delta\rho\\ &= \begin{pmatrix}(1-\epsilon)(1-\delta)+\delta & (1-\delta)\sqrt{\epsilon(1-\epsilon)}\\ (1-\delta)\sqrt{\epsilon(1-\epsilon)} & (1-\delta)\epsilon\end{pmatrix}\end{aligned}$$

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

$$\sigma_\epsilon^{-1} = \frac{1}{(1-\delta)\delta\epsilon}\begin{pmatrix}(1-\delta)\epsilon & -(1-\delta)\sqrt{\epsilon(1-\epsilon)}\\ -(1-\delta)\sqrt{\epsilon(1-\epsilon)} & (1-\epsilon)(1-\delta)+\delta\end{pmatrix}$$

[10] This Lemma was pointed out to us by Omar Fawzi.

which implies that $\hat{D}(\rho\|\sigma_\epsilon) = \log\frac{1}{\delta}$ using Lemma B.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. 117 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. 117 (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. 117 as $\alpha \to 1$, for example, the bound in Eq. 115, 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 5.3, 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 $D^{\#}_\alpha$).

C Transforming lemmas for EAT from $\tilde{H}^{\downarrow}_\alpha$ to $\tilde{H}^{\uparrow}_\alpha$

We have to redo the Lemmas used in [DFR20] using $\tilde{H}^{\uparrow}_\alpha$ because we were only able to prove the dimension bound we need ($\tilde{H}^{\uparrow}_\alpha(A|BC) \geq \tilde{H}^{\uparrow}_\alpha(A|B) - 2\log|C|$) in terms of $\tilde{H}^{\uparrow}_\alpha$

Lemma C.1 ([DFR20, Lemma 3.1])**.** *For $\rho_{A_1A_2B}$ and σ_B be states and $\alpha \in (0,\infty)$, we have the chain rule*

$$\tilde{D}_\alpha(\rho_{A_1B}\| \mathbb{1}_{A_1}\otimes\sigma_B) - \tilde{D}_\alpha(\rho_{A_1A_2B}\| \mathbb{1}_{A_1A_2}\otimes\sigma_B) = \tilde{H}^{\downarrow}_\alpha(A_2|A_1B)_\nu \tag{118}$$

where the state $\nu_{A_1A_2B}$ is defined as

$$\nu_{A_1B} := \frac{\left(\rho^{\frac{1}{2}}_{A_1B}\sigma_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^\alpha}{\mathrm{tr}\left(\rho^{\frac{1}{2}}_{A_1B}\sigma_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^\alpha}$$
$$\nu_{A_1A_2B} := \nu^{\frac{1}{2}}_{A_1B}\rho_{A_2|A_1B}\nu^{\frac{1}{2}}_{A_1B}$$

and $\alpha' := \frac{\alpha-1}{\alpha}$.

Corollary C.2 (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*

$$\tilde{H}^{\downarrow}_\alpha(A_1A_2|B)_\rho = \tilde{H}^{\downarrow}_\alpha(A_1|B)_\rho + \tilde{H}^{\downarrow}_\alpha(A_2|A_1B)_\nu \tag{119}$$

where the state $\nu_{A_1A_2B}$ is defined as

$$\nu_{A_1B} := \frac{\left(\rho^{\frac{1}{2}}_{A_1B}\rho_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^\alpha}{\mathrm{tr}\left(\rho^{\frac{1}{2}}_{A_1B}\rho_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^\alpha}$$
$$\nu_{A_1A_2B} := \nu^{\frac{1}{2}}_{A_1B}\rho_{A_2|A_1B}\nu^{\frac{1}{2}}_{A_1B}$$

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 C.3 (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*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_\rho \geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B)_\rho + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_\nu \tag{120}$$

where the state $\nu_{A_1A_2B}$ *is defined as*

$$\nu_{A_1B} := \frac{\left(\rho^{\frac{1}{2}}_{A_1B}\sigma_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^{\alpha}}{\mathrm{tr}\left(\rho^{\frac{1}{2}}_{A_1B}\sigma_B^{-\alpha'}\rho^{\frac{1}{2}}_{A_1B}\right)^{\alpha}}$$

$$\nu_{A_1A_2B} := \nu^{\frac{1}{2}}_{A_1B}\rho_{A_2|A_1B}\nu^{\frac{1}{2}}_{A_1B}$$

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*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_\rho \leq \tilde{H}^{\uparrow}_{\alpha}(A_1|B)_\rho + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_\nu \tag{121}$$

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 C.1, 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 C.4. *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*

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B_1B_2)_\rho \geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1)_\rho + \inf_{\nu} \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B_1B_2)_\nu \tag{122}$$

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

$$\rho_{A_1B_1B_2} = \bigoplus_{j \in J} p(j) \rho_{A_1 a_j} \otimes \rho_{c_j B_2}. \tag{123}$$

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

$$\mathcal{R}_{B_1 \to B_1B_2}(X) := \bigoplus_{j \in J} \mathrm{tr}_{c_j}\left(\Pi_{a_j} \otimes \Pi_{c_j} X \Pi_{a_j} \otimes \Pi_{c_j}\right) \otimes \rho_{c_j B_2} \tag{124}$$

where $\Pi_{a_j} \otimes \Pi_{c_j}$ is the projector on the subspace $a_j \otimes c_j$. This recovery channel satisfies

$$\mathcal{R}_{B_1 \to B_1B_2}(\rho_{A_1B_1}) = \rho_{A_1B_1B_2}. \tag{125}$$

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

$$\begin{aligned}
\tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_\rho &= \sup_{\sigma_{B_1B_2}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2} \| \mathbb{1}_{A_1} \otimes \sigma_{B_1B_2}) \\
&\leq \sup_{\sigma_{B_1B_2}} -\tilde{D}_\alpha(\mathcal{R}_{B_1 \to B_1B_2} \circ \mathrm{tr}_{B_2}(\rho_{A_1B_1B_2}) \| \mathcal{R}_{B_1 \to B_1B_2} \circ \mathrm{tr}_{B_2}(\mathbb{1}_{A_1} \otimes \sigma_{B_1B_2})) \\
&= \sup_{\sigma_{B_1}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2} \| \mathbb{1}_{A_1} \otimes \mathcal{R}_{B_1 \to B_1B_2}(\sigma_{B_1})) \\
&\leq \sup_{\sigma_{B_1B_2}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2} \| \mathbb{1}_{A_1} \otimes \sigma_{B_1B_2}) \\
&= \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_\rho
\end{aligned}$$

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

$$\tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_\rho = \sup_{\sigma_{B_1}} -\tilde{D}_\alpha(\rho_{A_1B_1B_2} \| \mathbb{1}_{A_1} \otimes \mathcal{R}_{B_1 \to B_1B_2}(\sigma_{B_1})) \tag{126}$$

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 C.3, for this choice of $\sigma_{B_1B_2}$, we have that

$$\tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B_1B_2)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A_1|B_1B_2)_{\rho} + \tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B_1B_2)_{\nu} \tag{127}$$

where the state $\nu_{A_1A_2B_1B_2}$ is defined as

$$\nu_{A_1B_1B_2} := \frac{\left(\rho^{\frac{1}{2}}_{A_1B_1B_2}\sigma^{-\alpha'}_{B_1B_2}\rho^{\frac{1}{2}}_{A_1B_1B_2}\right)^{\alpha}}{\mathrm{tr}\left(\rho^{\frac{1}{2}}_{A_1B_1B_2}\sigma^{-\alpha'}_{B_1B_2}\rho^{\frac{1}{2}}_{A_1B_1B_2}\right)^{\alpha}}$$

$$\nu_{A_1A_2B_1B_2} := \nu^{\frac{1}{2}}_{A_1B_1B_2}\rho_{A_2|A_1B_1B_2}\nu^{\frac{1}{2}}_{A_1B_1B_2}.$$

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

$$\begin{aligned}
\sigma_{B_1B_2} &= \mathcal{R}_{B_1\to B_1B_2}(\eta_{B_1}) \\
&= \bigoplus_{j\in J}\mathrm{tr}_{c_j}\left(\Pi_{a_j}\otimes\Pi_{c_j}\eta_{B_1}\Pi_{a_j}\otimes\Pi_{c_j}\right)\otimes\rho_{c_jB_2} \\
&= \bigoplus_{j\in J} q(j)\omega_{a_j}\otimes\rho_{c_jB_2}
\end{aligned}$$

where we have defined the probability distribution $q(j) := \mathrm{tr}(\Pi_{a_j}\otimes\Pi_{c_j}\eta_{B_1})$ and states $\omega_{a_j} = \frac{1}{q(j)}\Pi_{a_j}\,\mathrm{tr}_{c_j}\left(\Pi_{c_j}\eta_{B_1}\Pi_{c_j}\right)\Pi_{a_j}$ 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

$$\begin{aligned}
&\rho_{A_1B_1B_2} \ll \mathbb{1}_{A_1}\otimes\sigma_{B_1B_2} \\
\Rightarrow &\bigoplus_{j\in J'} p(j)\rho_{A_1a_j}\otimes\rho_{c_jB_2} \ll \mathbb{1}_{A_1}\otimes\bigoplus_{j\in J} q(j)\omega_{a_j}\otimes\rho_{c_jB_2} \\
\Rightarrow &\text{for every } j\in J' : \rho_{A_1a_j} \ll \mathbb{1}_{A_1}\otimes\omega_{a_j} \text{ and } q(j) > 0.
\end{aligned} \tag{128}$$

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 $N := \operatorname{tr}\left(\rho_{A_1B_1B_2}^{\frac{1}{2}}\sigma_{B_1B_2}^{-\alpha'}\rho_{A_1B_1B_2}^{\frac{1}{2}}\right)^{\alpha}$. Further, we have

$$
\begin{aligned}
&\nu_{A_1B_1}^{-\frac{1}{2}}\nu_{A_1B_1B_2}^{\frac{1}{2}} \\
&\qquad = \frac{1}{N^{-\frac{1}{2}}}\bigoplus_{j\in J'} p(j)^{-\frac{\alpha}{2}} q(j)^{-\frac{1-\alpha}{2}}\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right)^{-\frac{\alpha}{2}}\otimes\rho_{c_j}^{-\frac{1}{2}} \\
&\qquad\qquad\qquad \cdot\frac{1}{N^{\frac{1}{2}}}\bigoplus_{j\in J'} p(j)^{\frac{\alpha}{2}} q(j)^{\frac{1-\alpha}{2}}\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right)^{\frac{\alpha}{2}}\otimes\rho_{c_jB_2}^{\frac{1}{2}} \\
&\qquad = \bigoplus_{j\in J'}\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right)^{0}\otimes\rho_{c_j}^{-\frac{1}{2}}\rho_{c_jB_2}^{\frac{1}{2}} \\
&\qquad = \bigoplus_{j\in J'}\rho_{A_1a_j}^{0}\otimes\rho_{c_j}^{-\frac{1}{2}}\rho_{c_jB_2}^{\frac{1}{2}}
\end{aligned}
$$

where in the last line we have used that the projector $\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right)^{0}$ is equal to the projector $\rho_{A_1a_j}^{0}$ 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

$$
\operatorname{im}\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right)\subseteq\operatorname{im}\left(\rho_{A_1a_j}\right). \tag{129}
$$

Second, we have that Eq. 128 above implies that $\omega_{a_j}^{0}\rho_{Aa_j}^{0} = \rho_{Aa_j}^{0}$ for every $j\in J'$. Now, for $j\in J'$ we have the following inequality

$$
\begin{aligned}
\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right) &\geq m\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{0}\rho_{A_1a_j}^{\frac{1}{2}}\right) \\
&= m\rho_{A_1a_j}
\end{aligned}
$$

where $m>0$ is the minimum non-zero eigenvalue of $\omega_{a_j}^{-\alpha'}$. Finally, raising the above to the power of 0 (this action is operator monotone)

$$
\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right)^{0}\geq\rho_{A_1a_j}^{0}. \tag{130}
$$

Eq. 129 and 130 together imply that for $j\in J'$

$$
\left(\rho_{A_1a_j}^{\frac{1}{2}}\omega_{a_j}^{-\alpha'}\rho_{A_1a_j}^{\frac{1}{2}}\right)^{0}=\rho_{A_1a_j}^{0}.
$$

Finally, we have that

$$
\begin{aligned}
\rho_{A_1B_1}^{-\frac{1}{2}}\rho_{A_1B_1B_2}^{\frac{1}{2}} &= \bigoplus_{j\in J'} p(j)^{-\frac{1}{2}}\rho_{A_1a_j}^{-\frac{1}{2}}\otimes\rho_{c_j}^{-\frac{1}{2}}\bigoplus_{j\in J'} p(j)^{\frac{1}{2}}\rho_{A_1a_j}^{\frac{1}{2}}\otimes\rho_{c_jB_2}^{\frac{1}{2}} \\
&= \bigoplus_{j\in J'}\rho_{A_1a_j}^{0}\otimes\rho_{c_j}^{-\frac{1}{2}}\rho_{c_jB_2}^{\frac{1}{2}}.
\end{aligned}
$$

This proves that

$$\nu_{A_1B_1}^{-\frac{1}{2}}\nu_{A_1B_1B_2}^{\frac{1}{2}} = \rho_{A_1B_1}^{-\frac{1}{2}}\rho_{A_1B_1B_2}^{\frac{1}{2}} \tag{131}$$

and hence

$$\begin{aligned}\nu_{A_2B_2|A_1B_1} &= \nu_{A_1B_1}^{-\frac{1}{2}}\nu_{A_1B_1B_2}^{\frac{1}{2}}\nu_{A_2|A_1B_1B_2}\nu_{A_1B_1B_2}^{\frac{1}{2}}\nu_{A_1B_1}^{-\frac{1}{2}} \\ &= \rho_{A_1B_1}^{-\frac{1}{2}}\rho_{A_1B_1B_2}^{\frac{1}{2}}\rho_{A_2|A_1B_1B_2}\rho_{A_1B_1B_2}^{\frac{1}{2}}\rho_{A_1B_1}^{-\frac{1}{2}} \\ &= \rho_{A_2B_2|A_1B_1}\end{aligned}$$

where we have used the fact that $\nu_{A_2|A_1B_1B_2} = \rho_{A_2|A_1B_1B_2}$ and Eq. 131. We can now modify Eq. 127 to get

$$\tilde{H}_\alpha^\uparrow(A_1A_2|B_1B_2)_\rho \geq \tilde{H}_\alpha^\uparrow(A_1|B_1B_2)_\rho + \inf_\nu \tilde{H}_\alpha^\downarrow(A_2|A_1B_1B_2)_\nu$$

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

$$\begin{aligned}\tilde{H}_\alpha^\uparrow(A_1|B_1B_2)_\rho &= \tilde{H}_\alpha^\uparrow(A_1|B_1B_2)_{\mathcal{R}_{B_1\to B_1B_2}(\rho_{AB_1})} \\ &\geq \tilde{H}_\alpha^\uparrow(A_1|B_1)_\rho.\end{aligned}$$

Together with the above inequality this proves the Lemma. □

We will use the following modification of [DFR20, Corollary 3.5].

Corollary C.5. *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*

$$\tilde{H}_\alpha^\uparrow(A_1A_2|B_1B_2)_\rho \geq \tilde{H}_\alpha^\uparrow(A_1|B_1)_\rho + \inf_\omega \tilde{H}_\alpha^\downarrow(A_2|A_1B_1B_2)_{\mathcal{M}(\omega)} \tag{132}$$

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 $\nu_{A_1A_2B_1B_2} = \mathcal{M}(\omega)$. □

D Dimension bounds for conditional Rényi entropies

Lemma D.1 (Dimension bound)**.** *For $\alpha \in [\frac{1}{2}, \infty]$, a state $\rho_{A_1A_2B}$, the following bounds hold for the sandwiched conditional entropies*

$$\tilde{H}^{\downarrow}_{\alpha}(A_1|B)_\rho - \log|A_2| \le \tilde{H}^{\downarrow}_{\alpha}(A_1A_2|B)_\rho \le \tilde{H}^{\downarrow}_{\alpha}(A_1|B)_\rho + \log|A_2|$$
$$\tilde{H}^{\uparrow}_{\alpha}(A_1|B)_\rho - \log|A_2| \le \tilde{H}^{\uparrow}_{\alpha}(A_1A_2|B)_\rho \le \tilde{H}^{\uparrow}_{\alpha}(A_1|B)_\rho + \log|A_2|.$$

For $\alpha \in [0, 2]$ and a state $\rho_{A_1A_2B}$, the following bounds hold for the Petz conditional entropies

$$\bar{H}^{\downarrow}_{\alpha}(A_1A_2|B)_\rho \le \bar{H}^{\downarrow}_{\alpha}(A_1|B)_\rho + \log|A_2|$$
$$\bar{H}^{\uparrow}_{\alpha}(A_1A_2|B)_\rho \le \bar{H}^{\uparrow}_{\alpha}(A_1|B)_\rho + \log|A_2|.$$

Proof. For the sandwiched conditional entropies, we simply use the corresponding chain rules (Corollary C.2 or Corollary C.3) along with the fact that for all states ν, $\tilde{H}^{\downarrow}_{\alpha}(A_2|A_1B)_\nu \in [-\log|A_2|, \log|A_2|]$ [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 $\alpha \in [0, 1]$

$$\begin{aligned} \mathrm{tr}_X\, P_{XY}^{\alpha} &= \sum_{i=1}^{|X|} \mathbb{1}_Y \otimes \langle e_i|_X\, P_{XY}^{\alpha}\, \mathbb{1}_Y \otimes |e_i\rangle_X \\ &\le |X| \left(\sum_{i=1}^{|X|} \frac{1}{|X|} \mathbb{1}_Y \otimes \langle e_i|_X\, P_{XY}\, \mathbb{1}_Y \otimes |e_i\rangle_X \right)^{\alpha} \\ &= |X|^{1-\alpha} P_Y^{\alpha} \end{aligned} \tag{133}$$

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

$$\mathrm{tr}_X\, P_{XY}^{\alpha} \ge |X|^{1-\alpha} P_Y^{\alpha}. \tag{134}$$

To prove the dimension bound, observe that for a positive state σ_B and $\alpha \in [0,2]$, we have

$$\begin{aligned}
-\bar{D}_\alpha(\rho_{A_1A_2B}\| \mathbb{1}_{A_1A_2} \otimes \sigma_B) &= \frac{1}{1-\alpha}\log \operatorname{tr}\left(\rho^\alpha_{A_1A_2B}\sigma_B^{1-\alpha}\right) \\
&= \frac{1}{1-\alpha}\log \operatorname{tr}\left(\operatorname{tr}_{A_2}\left(\rho^\alpha_{A_1A_2B}\right)\sigma_B^{1-\alpha}\right) \\
&\leq \frac{1}{1-\alpha}\log \operatorname{tr}\left(|A_2|^{1-\alpha}\rho^\alpha_{A_1B}\sigma_B^{1-\alpha}\right) \\
&= -\bar{D}_\alpha(\rho_{A_1B}\| \mathbb{1}_{A_1} \otimes \sigma_B) + \log|A_2|.
\end{aligned}$$

We can now take a supremum over σ_B to prove the dimension bound for $\bar{H}^\uparrow_\alpha$ or choose $\sigma_B = \rho_B$ to prove the dimension bound for $\bar{H}^\downarrow_\alpha$. □

The following Lemma was originally proven in [MLDS+13, Proposition 8]. We reproduce the proof argument here.

Lemma D.2. *For $\alpha \in [\frac{1}{2}, \infty]$, a state ρ_{ABC}, we have*

$$\tilde{H}^\uparrow_\alpha(A|BC)_\rho \geq \tilde{H}^\uparrow_\alpha(AC|B)_\rho - \log|C| \tag{135}$$

and for $\alpha \in [0,2]$

$$\bar{H}^\uparrow_\alpha(A|BC)_\rho \geq \bar{H}^\uparrow_\alpha(AC|B)_\rho - \log|C| \tag{136}$$

Proof. By the definition of the sandwiched conditional entropy, we have

$$\begin{aligned}
\tilde{H}^\uparrow_\alpha(A|BC) &= \sup_{\eta_{BC}\in D(BC)} -\tilde{D}_\alpha(\rho_{ABC}\| \mathbb{1}_A \otimes \eta_{BC}) \\
&\geq \sup_{\eta_B\in D(B)} -\tilde{D}_\alpha\left(\rho_{ABC}\| \mathbb{1}_A \otimes \frac{\mathbb{1}_C}{|C|} \otimes \eta_B\right) \\
&= \sup_{\eta_B\in D(B)} -\tilde{D}_\alpha\left(\rho_{ABC}\| \mathbb{1}_{AC} \otimes \eta_B\right) - \log|C| \\
&= \tilde{H}^\uparrow_\alpha(AC|B) - \log|C|
\end{aligned}$$

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 $\bar{H}^\uparrow_\alpha$ entropy. □

The following lemma was originally proven in [Led16, Proposition 3.3.5].

Lemma D.3 (Dimension bound for conditioning register)**.** *For $\alpha \in [\frac{1}{2}, \infty]$ and a state ρ_{ABC} we have*

$$\tilde{H}^\uparrow_\alpha(A|BC)_\rho \geq \tilde{H}^\uparrow_\alpha(A|B)_\rho - 2\log|C|. \tag{137}$$

Further, if the register C is classical, then we have

$$\tilde{H}^{\uparrow}_{\alpha}(A|BC)_{\rho} \geq \tilde{H}^{\uparrow}_{\alpha}(A|B)_{\rho} - \log|C|. \tag{138}$$

Proof. This bound can be proven by combining Lemma D.1 and Lemma D.2. 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]. □

E Bounds on the size of the side information are necessary for the approximate entropy accumulation theorem

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 5.1. The following classical example demonstrates this.

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 $H_{\min}(A_1^n|B_1^n C_1^n)_{\mathcal{M}'_n \circ \cdots \circ \mathcal{M}'_1(1)} = n$ whereas $H_{\min}(A_1^n|B_1^n C_1^n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(1)} = O(1)$ 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 E.1. *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*

$$\left\|\Phi(\rho_{RR'}) - \Phi'(\rho_{RR'})\right\|_1 \leq \left\|P^{\Phi}_{AC} - P^{\Phi'}_{AC}\right\|_1 \tag{139}$$

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 $p(c) := \mathrm{tr}\left(|c\rangle\langle c|_R\,\rho_R\right)$ and $\rho_{R'|c} := \frac{1}{p(c)}\mathrm{tr}_R\left(|c\rangle\langle c|_R\,\rho_{RR'}\right)$. 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

$$\begin{aligned}\rho^{\Phi}_{ACR'} &:= \sum_{ac} p^{\Phi}_{A|C}(a|c)p(c)\,|a,c\rangle\langle a,c| \otimes \rho_{R'|c}\\ \rho^{\Phi'}_{ACR'} &:= \sum_{ac} p^{\Phi'}_{A|C}(a|c)p(c)\,|a,c\rangle\langle a,c| \otimes \rho_{R'|c}.\end{aligned}$$

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

$$\begin{aligned}\left\|\Phi(\rho_{RR'}) - \Phi'(\rho_{RR'})\right\|_1 &\leq \left\|\rho^{\Phi}_{ACR'} - \rho^{\Phi'}_{ACR'}\right\|_1\\ &\leq \left\|\rho^{\Phi}_{AC} - \rho^{\Phi'}_{AC}\right\|_1\\ &= \left\|P^{\Phi}_{AC} - P^{\Phi'}_{AC}\right\|_1.\end{aligned}$$

□

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

$$\begin{aligned}\left\|P_{A_1^k B_k C_k} - Q_{A_1^k B_k C_k}\right\|_1 &= \sum_{a_1^k, c_k} P(a_1^{k-1})P(a_k)P(c_k)\left\|P_{B_k|a_1^k,c_k} - Q_{B_k}\right\|_1\\ &= \sum_{a_1^k} P(a_1^{k-1})P(a_k)\left(\left(1-\frac{\epsilon}{2}\right)\left\|P_{B_k|a_1^k,c_k=1} - Q_{B_k}\right\|_1 + \frac{\epsilon}{2}\left\|P_{B_k|a_1^k,c_k=0} - Q_{B_k}\right\|_1\right)\\ &\leq \sum_{a_1^k} P(a_1^{k-1})P(a_k)\epsilon\\ &= \epsilon\end{aligned}$$

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

$$H_{\min}(A_1^n | B_1^n C_1^n)_S = n.$$

We will show that $H_{\min}^{\epsilon'}(A_1^n | B_1^n C_1^n)_R = O(1)$ 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 E.2. *Let P_{AB} be a subnormalized probability distribution such that $A = f(B)$ for some function f (that is, $P(a,b) > 0$ only if $a = f(b)$). Then, $H_{\min}^{\epsilon}(A|B)_P \leq \log \frac{1}{\mathrm{tr}(P) - \sqrt{2\epsilon}}$.*

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

$$\begin{aligned}
2^{-H_{\min}(A|B)_{P'}} &= P'_{\text{guess}}(A|B) \\
&\geq \sum_b P'_{AB}(f(b), b) \\
&\geq \sum_b P_{AB}(f(b), b) - \sqrt{2\epsilon} \\
&= \mathrm{tr}(P) - \sqrt{2\epsilon}
\end{aligned}$$

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 $H_{\min}^{\epsilon}(A|B)_P$. □

We then have that

$$\begin{aligned}
H_{\min}^{\epsilon'}(A_1^n | B_1^n C_1^n)_R &\leq H_{\min}^{\epsilon'}(A_1^n | B_1^n C_1^n \wedge E)_R \\
&\leq H_{\min}^{\epsilon'}(A_1^{n-l} | B_1^n C_1^n \wedge E)_R + l \\
&\leq \log \frac{1}{p(E) - \sqrt{2\epsilon'}} + l \\
&\leq \log \frac{1}{1 - \epsilon' - \sqrt{2\epsilon'}} + l \\
&\leq l + \log 8/3 = O(1)
\end{aligned}$$

where in the first line we have used [TL17, Lemma 10] in the first line, dimension bound (can be proven using Lemma D.1) in the second line, Lemma E.2 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 4.5, 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 [MFSR22] 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.

F Classical approximate entropy accumulation

We present a simple proof for the approximate entropy accumulation theorem for classical distributions. This result also requires a much weaker assumption than Theorem 5.1.

Theorem F.1. *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*

$$\left\| p_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} - q^{(k)}_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} \right\|_\infty \leq \epsilon \tag{140}$$

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

$$\begin{aligned} H_{\min}^{\epsilon'}(A_1^n | B_1^n E)_p \geq \sum_{k=1}^n \inf_q H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}} \\ - n(\alpha - 1)\log^2(2|A| + 1) - \frac{\alpha}{\alpha - 1} n \log(1 + \epsilon |A||B|) - \frac{g_0(\epsilon')}{\alpha - 1}. \end{aligned} \tag{141}$$

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 $\sqrt{\epsilon} \leq 1 + \frac{1}{\log(1+2|A|)}$), and using $\alpha \leq 2$ and $\log(1+x) \leq x$ as long as $x \geq 0$, the above bound gives us

$$\begin{aligned} H_{\min}^{\epsilon'}(A_1^n|B_1^n E)_p \geq \sum_{k=1}^{n} \inf_q H(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k|A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}} \\ - n\sqrt{\epsilon}\left(\log^2(2|A|+1) - 2|A||B|\right) - \frac{g_0(\epsilon')}{\alpha - 1} \end{aligned} \tag{142}$$

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

$$\begin{aligned} &\left\| p_{A_k B_k|a_1^{k-1}, b_1^{k-1}, e} - q^{(k)}_{A_k B_k|a_1^{k-1}, b_1^{k-1}, e} \right\|_\infty \leq \epsilon \\ \Rightarrow\ & p_{A_k B_k|a_1^{k-1}, b_1^{k-1}, e} \leq q^{(k)}_{A_k B_k|a_1^{k-1}, b_1^{k-1}, e} + \epsilon \mathbb{1}_{A_k B_k} \\ \Rightarrow\ & p_{A_k B_k|a_1^{k-1}, b_1^{k-1}, e} \leq q^{(k)}_{A_k B_k|a_1^{k-1}, b_1^{k-1}, e} + \epsilon |A||B| u_{A_k B_k} \\ \Rightarrow\ & p_{A_k B_k|a_1^{k-1}, b_1^{k-1}, e} \leq (1 + |A||B|\epsilon) r^{(k)}_{A_k B_k|A_1^{k-1} B_1^{k-1} E} \end{aligned}$$

Define the distribution

$$r_{A_1^n B_1^n E} = \prod_{k=1}^{n} r^{(k)}_{A_k B_k|A_1^{k-1} B_1^{k-1} E} p_E. \tag{143}$$

Note that for every k, a_1^{k-1}, b_1^k, and e, we have

$$\begin{aligned} r_{B_k|A_1^{k-1} B_1^{k-1} E}(b_k|a_1^{k-1} b_1^{k-1} e) &= \frac{1}{1+|A||B|\epsilon} q^{(k)}_{B_k|A_1^{k-1} B_1^{k-1} E}(b_k|a_1^{k-1} b_1^{k-1} e) + \frac{\epsilon}{1+|A||B|\epsilon} \\ &= \frac{1}{1+|A||B|\epsilon} q^{(k)}_{B_k|B_1^{k-1} E}(b_k|b_1^{k-1} e) + \frac{\epsilon}{1+|A||B|\epsilon}, \end{aligned}$$

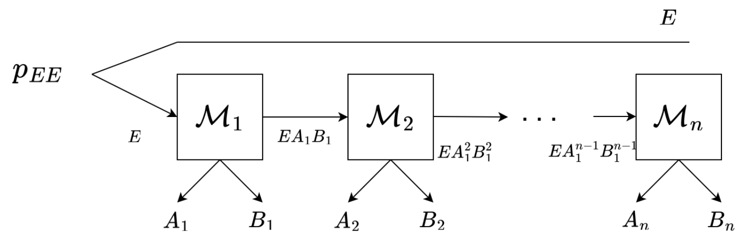

Figure 3: Setting for classical EAT

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

$$
\begin{aligned}
p_{A_1^n B_1^n E}(a_1^n, b_1^n, e) &= \prod_{k=1}^n p_{A_k B_k|A_1^{k-1}, B_1^{k-1}, E}(a_k, b_k|a_1^{k-1}, b_1^{k-1}, e) p_E(e) \\
&\leq (1+\epsilon|A||B|)^n \prod_{k=1}^n r^{(k)}_{A_k B_k|A_1^{k-1}, B_1^{k-1}, E}(a_k, b_k|a_1^{k-1}, b_1^{k-1}, e) p_E(e) \\
&= (1+\epsilon|A||B|)^n r_{A_1^n B_1^n E}(a_1^n, b_1^n, e)
\end{aligned}
$$

which shows that $D_{\max}(p_{A_1^n B_1^n E} \| r_{A_1^n B_1^n E}) \leq n \log(1+\epsilon|A||B|)$.

The distribution $r_{A_1^n B_1^n E}$ can be viewed as the result of a series of maps as in Fig. 3. 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

$$
\begin{aligned}
\tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_r &\geq \sum_{k=1}^{n} \inf_{q_{A_1^{k-1}B_1^{k-1}E}} \tilde{H}^{\downarrow}_{\alpha}(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{r^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} \\
&\geq \sum_{k=1}^{n} \inf_{q_{A_1^{k-1}B_1^{k-1}E}} H(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{r^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} - n(\alpha-1)\log^2(2|A|+1) \\
&\geq \sum_{k=1}^{n} \left(\inf_{q_{A_1^{k-1}B_1^{k-1}E}} \frac{1}{1+|A||B|\epsilon} H(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} \right. \\
&\qquad \left. + \frac{\epsilon}{1+|A||B|\epsilon} \log|A| \right) - n(\alpha-1)\log^2(2|A|+1) \\
&\geq \sum_{k=1}^{n} \inf_{q_{A_1^{k-1}B_1^{k-1}E}} H(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} - n(\alpha-1)\log^2(2|A|+1)
\end{aligned}
$$

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

$$
\begin{aligned}
H^{\epsilon'}_{\min}(A_1^n|B_1^n E)_p &\geq \tilde{H}^{\uparrow}_{\alpha}(A_1^n|B_1^n E)_r - \frac{\alpha}{\alpha-1} D_{\max}(p_{A_1^n B_1^n E} \| r_{A_1^n B_1^n E}) - \frac{g_1(\epsilon', 0)}{\alpha-1} \\
&\geq \sum_{k=1}^{n} \inf_{q} H(A_k|B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k|A_1^{k-1}B_1^{k-1}E} q_{A_1^{k-1}B_1^{k-1}E}} \\
&\qquad - n(\alpha-1)\log^2(2|A|+1) - \frac{\alpha}{\alpha-1} n \log(1+\epsilon|A||B|) - \frac{g_0(\epsilon')}{\alpha-1}.
\end{aligned}
$$

□

G Proof of Theorem 6.3

In this section, we formally prove the lower bound on the smooth min-entropy required for the security of QKD in Theorem 6.3 using the entropy accumulation theorem (EAT). In Section 6.1 (Eq. 93 and 94), we showed that $\rho'_{X_1^n \Theta_1^n A_1^n} := \bar{\bar{\rho}}_{X_1^n \Theta_1^n A_1^n|\Omega}$ and $\sigma_{X_1^n \Theta_1^n A_1^n} = \left(\hat{\rho}^{(\epsilon+\delta)}_{X\Theta A}\right)^{\otimes n}$ is such that

$$
\frac{1}{2} \left\| \rho'_{X_1^n \Theta_1^n A_1^n} - \bar{\rho}_{X_1^n \Theta_1^n A_1^n|\Omega} \right\|_1 \leq \frac{\epsilon_f^2}{2} \tag{144}
$$

and

$$
D_{\max}(\rho'_{X_1^n \Theta_1^n A_1^n} \| \sigma_{X_1^n \Theta_1^n A_1^n}) \leq nh(\epsilon+\delta) + \log \frac{1}{\Pr_\rho(\Omega) - \epsilon^{\delta}_{\text{qu}}}. \tag{145}
$$

Fix an arbitrary strategy for Eve. Let $\Phi_{\text{QKD}} : X_1^n \Theta_1^n A_1^n \to X_1^n Y_1^n \hat{X}_S \hat{C}_1^n \Theta_1^n \hat{\Theta}_1^n STE$ 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 $\Phi_{\text{QKD}}(\bar{\rho})$

$$H_{\min}^{\nu}(X_S | E T \Theta_1^n \hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Upsilon}}$$

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

$$H_{\min}^{\nu}(X_S | E T \Theta_1^n \hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Upsilon''}}$$

in Theorem 6.3[(11)].

Using the data processing inequality and Eq. 145, we see that

$$D_{\max}(\Phi_{\text{QKD}}(\rho'_{X_1^n \Theta_1^n A_1^n}) \| \Phi_{\text{QKD}}(\sigma_{X_1^n \Theta_1^n A_1^n})) \leq n h(\epsilon + \delta) + \log \frac{1}{\Pr_{\rho}(\Omega) - \epsilon_{\text{qu}}^{\delta}}. \tag{146}$$

Note that $\Phi_{\text{QKD}}(\rho'_{X_1^n \Theta_1^n A_1^n})$ and $\Phi_{\text{QKD}}(\sigma_{X_1^n \Theta_1^n A_1^n})$ are the states that are produced at the end of the protocol if Alice's source were to produce the states $\rho'_{X_1^n \Theta_1^n A_1^n}$ and $\sigma_{X_1^n \Theta_1^n A_1^n}$ respectively. The states $\Phi_{\text{QKD}}(\rho'_{X_1^n \Theta_1^n A_1^n})$ and $\Phi_{\text{QKD}}(\sigma_{X_1^n \Theta_1^n A_1^n})$ also contain all the corresponding classical variables as the real protocol state $\Phi_{\text{QKD}}(\bar{\rho}_{X_1^n \Theta_1^n A_1^n | \Omega})$. In particular, the event Υ'' is well-defined (defined using classical variables) for both of these states.

Using Lemma 6.2 and Eq. 144, we have that the final states conditioned on the event Υ'' satisfy

$$\left\| \Phi_{\text{QKD}}(\bar{\rho}_{X_1^n \Theta_1^n A_1^n})_{|\Omega \wedge \Upsilon''} - \Phi_{\text{QKD}}(\rho'_{X_1^n \Theta_1^n A_1^n})_{|\Upsilon''} \right\|_1 \leq \frac{\epsilon_f^2}{\Pr_{\bar{\rho}}(\Upsilon'' | \Omega)} \tag{147}$$

where $\Pr_{\bar{\rho}}(\Upsilon''|\Omega)$ is the probability for the event Υ'' for the state $\Phi_{\text{QKD}}(\bar{\rho}_{X_1^n \Theta_1^n A_1^n | \Omega})$[(12)]. Using the Fuchs- van de Graaf inequality [Tom16, Lemma 3.5], we can transform this to a

[(11)]The arguments in [MR22, Appendix A] can also be modified to show that it is sufficient to show that $P(\Upsilon'') \left\| \rho^f_{K_A E'} - \tau_{K_A} \otimes \rho^f_{E'} \right\|_1$ 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.

[(12)]We abuse notation while writing the probability this way since the state it is evaluated on is $\Phi_{\text{QKD}}(\bar{\rho}_{X_1^n \Theta_1^n Q_1^n | \Omega})$, while we simply use the subscripts $\bar{\rho}$ for P. We also write probabilities this way for the state $\Phi_{\text{QKD}}(\rho'_{X_1^n \Theta_1^n Q_1^n})$ and $\Phi_{\text{QKD}}(\sigma_{X_1^n \Theta_1^n Q_1^n})$. This is done for the sake of clarity.

purified distance bound

$$P(\Phi_{\mathrm{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nA_1^n})_{|\Omega\wedge\Upsilon''}, \Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}) \leq \sqrt{\frac{2}{P_{\bar{\rho}}(\Upsilon''|\Omega)}}\epsilon_f. \tag{148}$$

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. 146. By definition of $D_{\max}$, we have

$$\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n}) \leq 2^d\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n}).$$

Conditioning both sides on the event Υ'' implies that

$$P_{\rho'}(\Upsilon'')\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''} \leq 2^dP_{\sigma}(\Upsilon'')\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}$$

where $P_{\rho'}(\Upsilon'')$ and $P_{\sigma}(\Upsilon'')$ 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

$$D_{\max}(\Phi_{\mathrm{QKD}}(\rho'_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}\|\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}) \leq d + \log\frac{P_{\sigma}(\Upsilon'')}{P_{\rho'}(\Upsilon'')}.$$

Together, with Eq. 148 for $\epsilon_{\mathrm{pa}} := \left(\frac{2}{P_{\bar{\rho}}(\Upsilon''|\Omega)}\right)^{\frac{1}{2}}\epsilon_f$, we have that

$$D_{\max}^{\epsilon_{\mathrm{pa}}}(\Phi_{\mathrm{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nA_1^n})_{|\Omega\wedge\Upsilon''}\|\Phi_{\mathrm{QKD}}(\sigma_{X_1^n\Theta_1^nA_1^n})_{|\Upsilon''}) \leq d + \log\frac{P_{\sigma}(\Upsilon'')}{P_{\rho'}(\Upsilon'')}. \tag{149}$$

Let $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ be arbitrary parameters. We have

$$\begin{aligned}
&H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1+2(\epsilon_2+\epsilon_3)}(X_S|E\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}}\\
&\quad= H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1+2(\epsilon_2+\epsilon_3)}(\bar{X}_1^n|E\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}}\\
&\quad\geq H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^nE\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - 3g_0(\epsilon_3)\\
&\quad\geq H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - \log|T| - H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^nE\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} - 3g_0(\epsilon_3)
\end{aligned} \tag{150}$$

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 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

$$\begin{aligned}
&H_{\min}^{\epsilon_{\text{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad\geq \tilde{H}_\alpha^\uparrow(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\sigma)_{|\Upsilon''}} \\
&\qquad - \frac{\alpha}{\alpha-1}D_{\max}^{\epsilon_{\text{pa}}}(\Phi_{\text{QKD}}(\bar{\rho}_{X_1^n\Theta_1^nQ_1^n})_{|\Omega\wedge\Upsilon''}\|\Phi_{\text{QKD}}(\sigma_{X_1^n\Theta_1^nQ_1^n})_{|\Upsilon''}) - \frac{g_1(\epsilon_1,\epsilon_{\text{pa}})}{\alpha-1} \\
&\quad\geq \tilde{H}_\alpha^\uparrow(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\text{QKD}}(\sigma)_{|\Upsilon''}} \\
&\qquad - \frac{\alpha}{\alpha-1}d - \frac{\alpha}{\alpha-1}\log\frac{P_\sigma(\Upsilon'')}{P_{\rho'}(\Upsilon'')} - \frac{g_1(\epsilon_1,\epsilon_{\text{pa}})}{\alpha-1} \qquad (151)
\end{aligned}$$

Thus, we have reduced the problem to lower bounding α-Rényi conditional entropy for the QKD protocol in Protocol 2, 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 $(\Phi^+)^{\otimes n}_{\bar{A}_1^nA_1^n}$.
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 $\bar{A}_1^nA_1^nE$, which contain the state produced after Eve forwards the state produced above by Alice to Bob. We can now define the EAT maps $\mathcal{M}_i : \bar{A}_i^nA_i^n \to \bar{A}_{i+1}^nA_{i+1}^n\bar{X}_i\bar{Y}_i\Theta_i\hat{\Theta}_iC_i$, 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 1.

Note that by conditioning on the event Υ'', we are requiring that $q = \text{freq}(C_1^n)$ 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 $f(\text{freq}(C_1^n)) \geq 1-2\mu+\mu^2-h(e)$. Using the entropy accumulation theorem [DFR20, Proposition 4.5], we get

$$\tilde{H}^{\uparrow}_{\alpha}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\sigma^\delta)_{|\Upsilon''}} \geq n(1-2\mu+\mu^2-h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}\log\frac{1}{P_\sigma(\Upsilon'')} \tag{152}$$

where $V := 2\lceil\|\nabla f\|_\infty\rceil + 2\log(1+2|\mathcal{X}|^2) = \frac{2}{\mu^2}\log\frac{1-e}{e} + 2\log(1+2|\mathcal{X}|^2)$ and $1 < \alpha < 1+\frac{2}{V}$. Combining Eq. 151 and 152, we get

$$\begin{aligned}
&H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\qquad \geq n(1-2\mu+\mu^2-h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}d - \frac{\alpha}{\alpha-1}\log\frac{1}{P_{\rho'}(\Upsilon'')} - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{\alpha-1} \\
&\qquad \geq n(1-2\mu+\mu^2-h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}nh(\epsilon+\delta) - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr_\rho(\Omega)-\epsilon_{\mathrm{qu}}^\delta} \\
&\qquad\quad - \frac{\alpha}{\alpha-1}\log\frac{1}{P_{\bar{\rho}}(\Upsilon''|\Omega) - \frac{2\epsilon_{\mathrm{qu}}^\delta}{P_\rho(\Omega)}} - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{\alpha-1} \\
&\qquad \geq n(1-2\mu+\mu^2-h(e)) - n\frac{\alpha-1}{4}V^2 - \frac{\alpha}{\alpha-1}nh(\epsilon+\delta) \\
&\qquad\quad - \frac{\alpha}{\alpha-1}\left(\log\frac{1}{P_\rho(\Omega\wedge\Upsilon'')-2\epsilon_{\mathrm{qu}}^\delta}+1\right) - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{\alpha-1}
\end{aligned} \tag{153}$$

where we have used Eq. 144, $\epsilon_f = 2\sqrt{\frac{\epsilon_{\mathrm{qu}}^\delta}{P_\rho(\Omega)}}$, $P_\rho(\Omega\wedge\Upsilon'') = P_\rho(\Omega)P_{\bar{\rho}}(\Upsilon''|\Omega)$ and $P_\rho(\Omega) \geq P_\rho(\Omega\wedge\Upsilon'') > 2\epsilon_{\mathrm{qu}}^\delta$ to simplify the result. It should be noted that the probability $P_\sigma(\Upsilon'')$ 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

$$\alpha := 1 + \frac{2\sqrt{2h(\epsilon+\delta)}}{V} \tag{154}$$

which gives us the bound

$$\begin{aligned}
H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1}(\bar{X}_1^n\bar{Y}_1^n|E\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} &\geq n(1-2\mu+\mu^2-h(e)-V\sqrt{2h(\epsilon+\delta)}) \\
&\quad - \frac{V}{\sqrt{2h(\epsilon+\delta)}}\left(\log\frac{1}{P_\rho(\Omega\wedge\Upsilon'')-2\epsilon_{\mathrm{qu}}^\delta}+1\right) - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{2\sqrt{2h(\epsilon+\delta)}}V.
\end{aligned} \tag{155}$$

We also need to bound $H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^nE\Theta_1^n\hat{\Theta}_1^nT)_{\Phi_{\mathrm{QKD}}(\rho)_{|\Omega\wedge\Upsilon''}}$ in Eq. 150. 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

$$
\begin{aligned}
&H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad\le H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad\le \tilde{H}^{\downarrow}_{\frac{1}{\beta}}(\bar{Y}_1^n|\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} + \frac{g_0(\epsilon_2)}{\beta-1} \\
&\quad\le \tilde{H}^{\downarrow}_{\frac{1}{\beta}}(\bar{Y}_1^n|\Theta_1^n\hat{\Theta}_1^n)_{\Phi_{\mathrm{QKD}}(\bar{\rho})} + \frac{\beta}{\beta-1}\log\frac{1}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + \frac{g_0(\epsilon_2)}{\beta-1} \\
&\quad= \frac{\beta}{\beta-1}\log\sum_{\theta_1^n,\hat{\theta}_1^n} P(\theta_1^n,\hat{\theta}_1^n)2^{\left(1-\frac{1}{\beta}\right)\tilde{H}^{\downarrow}_{\frac{1}{\beta}}(\bar{Y}_1^n|\theta_1^n,\hat{\theta}_1^n)} + \frac{\beta}{\beta-1}\log\frac{1}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + \frac{g_0(\epsilon_2)}{\beta-1}
\end{aligned}
$$

where the first line follows from the data processing inequality for the smooth max-entropy, second line follows from [DFR20, Lemma B.10], 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

$$
\begin{aligned}
\frac{\beta}{\beta-1}\log\sum_{\theta_1^n,\hat{\theta}_1^n} P(\theta_1^n,\hat{\theta}_1^n)2^{\left(1-\frac{1}{\beta}\right)\tilde{H}^{\downarrow}_{\frac{1}{\beta}}(\bar{A}_1^n|\theta_1^n,\hat{\theta}_1^n)} &\le \frac{\beta}{\beta-1}\log\sum_{\theta_1^n,\hat{\theta}_1^n} P(\theta_1^n,\hat{\theta}_1^n)2^{\left(1-\frac{1}{\beta}\right)Z_{\theta_1^n,\hat{\theta}_1^n}} \\
&= \frac{\beta}{\beta-1}\log\sum_{z=0}^{n}\binom{n}{z}\mu^{2z}(1-\mu^2)^{n-z}2^{\left(1-\frac{1}{\beta}\right)z} \\
&= n\frac{\beta}{\beta-1}\log\left(1-\mu^2+2^{\left(1-\frac{1}{\beta}\right)}\mu^2\right) \\
&= n\mu^2\frac{\beta}{(\beta-1)\ln(2)}\left(2^{\left(1-\frac{1}{\beta}\right)}-1\right) \\
&\le n\mu^2\left(1+\frac{(\beta-1)\ln(2)}{\beta}\right)
\end{aligned}
$$

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 $\ln(1+x) \le x$ for all $x > -1$, and in the last line we use the fact that $e^x \le 1 + x + x^2$ for $x \in (0,1)$ and that for $\beta > 1$ the term $\ln(2)\left(1-\frac{1}{\beta}\right)$ lies in this range. Thus, we get that for $\beta \in (1,2)$,

$$
H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \le n\mu^2 + \frac{(\beta-1)\ln(2)}{\beta}n\mu^2 + \frac{\beta}{\beta-1}\log\frac{1}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + \frac{g_0(\epsilon_2)}{\beta-1}.
$$

Choosing $\beta = 1 + \frac{1}{\sqrt{n}}$ and using the coarse bounds $1 < \beta < 2$, gives us

$$H_{\max}^{\epsilon_2}(\bar{Y}_1^n|\bar{X}_1^n E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \leq n\mu^2 + \sqrt{n}\left(\mu^2\ln(2) + 2\log\frac{1}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + g_0(\epsilon_2)\right). \tag{156}$$

Combining Eq. 150, 155, and 156, we get

$$\begin{aligned}
&H_{\min}^{\epsilon_{\mathrm{pa}}+\epsilon_1+2(\epsilon_2+\epsilon_3)}(X_S|E\Theta_1^n\hat{\Theta}_1^n T)_{\Phi_{\mathrm{QKD}}(\bar{\rho})_{|\Omega\wedge\Upsilon''}} \\
&\quad\geq n(1-2\mu-h(e)-V\sqrt{2h(\epsilon+\delta)}) - \sqrt{n}\left(\mu^2\ln(2) + 2\log\frac{1}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'')} + g_0(\epsilon_2)\right) \\
&\quad -\frac{V}{\sqrt{2h(\epsilon+\delta)}}\left(\log\frac{1}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'') - 2\epsilon_{\mathrm{qu}}^{\delta}} + 1\right) - \frac{g_1(\epsilon_1,\epsilon_{\mathrm{pa}})}{2\sqrt{2h(\epsilon+\delta)}}V - \log|T| - 3g_0(\epsilon_3)
\end{aligned} \tag{157}$$

where the parameters $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ are arbitrary, and

$$\epsilon_{\mathrm{pa}} = 2\left(\frac{2\epsilon_{\mathrm{qu}}^{\delta}}{P_{\bar{\rho}}(\Omega\wedge\Upsilon'')}\right)^{1/2}.$$

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.

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.

[AF04] R Alicki and M Fannes. Continuity of quantum conditional information. *Journal of Physics A: Mathematical and General*, 37(5):L55L57, 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.

[BACGPH22] Andreas Bluhm, Ángela Capel, Paul Gondolf, and Antonio Pérez-Hernández. Continuity of quantum entropic quantities via almost convexity, 2022, 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.

[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.

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997.

[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.

[CLPK+23] Guillermo Currs-Lorenzo, Margarida Pereira, Go Kato, Marcos Curty, and Kiyoshi Tamaki. A security framework for quantum key distribution implementations, 2023, 2305.05930.

[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.

[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.

[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.

[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, 2203.06554. doi:10.48550/ARXIV.2203.06554.

[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.

[JK23] Rahul Jain and Srijita Kundu. A direct product theorem for quantum communication complexity with applications to device-independent cryptography, 2023, 2106.04299.

[JN11] Rahul Jain and Ashwin Nayak. Short proofs of the quantum substate theorem, 2011, 1103.6067. doi:10.48550/ARXIV.1103.6067.

[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.

[KWW12] Robert Konig, Stephanie Wehner, and Jrg Wullschleger. Unconditional security from noisy quantum storage. *IEEE Transactions on Information Theory*, 58(3):1962–1984, 2012. doi:10.1109/TIT.2011.2177772.

[Led16] Felix Leditzky. *Relative entropies and their use in quantum information theory*. PhD thesis, 2016, 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.

[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.

[MFSR22] Tony Metger, Omar Fawzi, David Sutter, and Renato Renner. Generalised entropy accumulation, 2022, 2203.04989. doi:10.48550/ARXIV.2203.04989.

[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, 2203.04993. doi:10.48550/ARXIV.2203.04993.

[PAB+09] Stefano Pironio, Antonio Acn, 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 Currs-Lorenzo, lvaro 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.

[Ren06] Renato Renner. *Security of Quantum Key Distribution*. PhD thesis, 2006, 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.

[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.

[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.

[Tan23] Ernest Y. Z. Tan. Robustness of implemented device-independent protocols against constrained leakage, 2023, 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, 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.

[Wat60] Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.*, 4(1):6682, jan 1960. doi:10.1147/rd.41.0066.

[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.

[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.