Smooth Min-entropy Lower Bounds for Approximation Chains

Ashutosh Marwah[ORCID]**, Frédéric Dupuis**

Département d'informatique et de Recherche Opérationnelle, Université de Montréal,
2900 boul. Édouard-Montpetit, Montreal, QC H3T 1J4, Canada. E-mail: ashutosh.marwah@outlook.com

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 entropic 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, utilising this approach, we prove approximate versions of the asymptotic equipartition property and entropy accumulation. In the companion paper (Marwah and Dupuis in Proving Security of BB84 Under Source Correlations, 2024. arXiv:2402.12346 [quant-ph]), we show that the techniques developed in this paper can be used to prove the security of quantum key distribution in the presence of source correlations.

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, [2,3]). 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 [4,5] is one of the most important quantities in one-shot information theory.

The smooth min-entropy $H^\epsilon_{\min}(K|E)_\rho$ 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 [6]. 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] $(\sigma^{(k)}_{A^k_1 B})^n_{k=1}$ an ϵ-approximation chain for the state $\rho_{A^n_1 B}$ if for every k, we can approximate the partial state $\rho_{A^k_1 B}$ as $\|\rho_{A^k_1 B} - \sigma^{(k)}_{A^k_1 B}\|_1 \leq \epsilon$. If one can further prove that these states satisfy $H(A_k|A^{k-1}_1 B)_{\sigma^{(k)}} \geq c$ for some $c > 0$ sufficiently large, then the following simple argument shows that $H(A^n_1|B)_\rho$ is large:

$$H(A^n_1|B)_\rho = \sum_{k=1}^n H(A_k|A^{k-1}_1 B)_\rho$$

$$\geq \sum_{k=1}^n \left(H(A_k|A^{k-1}_1 B)_{\sigma^{(k)}} - g(\epsilon) \right)$$

$$\geq n(c - g(\epsilon))$$

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 [7], are much more restrictive. Tools like entropy accumulation [8,9] 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 $\rho_{A^n_1 B}$, which for every $1 \leq k \leq n$ satisfies

$$\frac{1}{2} \left\| \rho_{A^k_1 B} - \rho_{A_k} \otimes \rho_{A^{k-1}_1 B} \right\|_1 \leq \epsilon. \tag{1}$$

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 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

---

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

[2] The smoothing parameter must depend on ϵ in such a scenario. This can be seen by considering the probability distribution $P_{A^n_1 B}$ such that B is 0 with probability ϵ and 1 otherwise and A^n_1 is a random n-bit string if $B = 1$ and constant if $B = 0$.

function $g'(\epsilon)$). However, it is not possible to prove this result using techniques, which rely only on the triangle inequality and smoothing. The triangle inequality, in general, can only be used to bound the trace distance between $\rho_{A_1^n B}$ and $\otimes_{k=1}^n \rho_{A_k} \otimes \rho_B$ by $n\epsilon$, which will result in a trivial bound when $n \gg \frac{1}{\epsilon}$.[3] Instead, we show that a bound on the entropic distance given by the smooth max-relative entropy between these two states can be used to prove a lower bound for the smooth min-entropy in this scenario.

While an upper bound of $n\epsilon$ is trivial and meaningless for the trace distance for large n, it is still a meaningful bound for the relative entropy between two states, which is unbounded in general. We can show that the above approximation conditions (Eq. 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 [10] 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 $d := D_{\max}^\delta(\rho_{AB}||\eta_{AB})$, we can easily bound the smooth min-entropy of ρ in terms of the min-entropy of η by observing that

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

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

$$H_{\min}^\delta(A|B)_\rho \geq H_{\min}(A|B)_\eta - D_{\max}^\delta(\rho_{AB}||\eta_{AB}) \qquad (3)$$

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

$$H_{\min}^{\epsilon+\delta}(A|B)_\rho \geq \tilde{H}_\alpha^\uparrow(A|B)_\eta - \frac{\alpha}{\alpha-1} D_{\max}^\epsilon(\rho_{AB}||\eta_{AB}) - \frac{g_1(\delta,\epsilon)}{\alpha-1} \qquad (4)$$

for some function g_1, $\epsilon + \delta < 1$ 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 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. 4 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 $\delta = O\left(\epsilon \log \frac{|A|}{\epsilon}\right)$ that

$$H_{\min}^{\delta^{\frac{1}{4}}}(A_1^n|B)_\rho \geq n\left(H(A_1)_\rho - O(\delta^{\frac{1}{4}})\right) - O\left(\frac{1}{\delta^{3/4}}\right). \qquad (5)$$

Another scenario we consider here is that of approximate entropy accumulation. In the setting for entropy accumulation, a sequence of channels $\mathcal{M}_k : R_{k-1} \to A_k B_k R_k$

---

[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 $\left\| Q_{A_1^{2n} B_1^{2n}} - \prod_{i=1}^{2n} Q_{A_i} \cdot Q_{B_1^{2n}} \right\|_1 \to 2$.

for $1 \leq k \leq n$ sequentially act on a state $\rho_{R_0 E}^{(0)}$ to produce the state $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})$. 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 [8], 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 [11,12], which are implemented with imperfect devices [13,14]. 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

$$\rho_{A_1^k B_1^k E} = \text{tr}_{R_k} \circ \mathcal{M}_k \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)}) \right)$$
$$\approx_\epsilon \text{tr}_{R_k} \circ \mathcal{M}_k' \left(\mathcal{M}_{k-1} \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)}) \right) := \sigma_{A_1^k B_1^k E}^{(k)}.$$

Moreover, the assumptions on the channel $\mathcal{M}_k'$ guarantee that the state $\sigma_{A_1^k B_1^k E}^{(k)}$ 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 [15] to show that the divergence distance between the states $\rho_{A_1^n B_1^n E} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})$ and the virtual state $\sigma_{A_1^n B_1^n E} = \mathcal{M}_n' \circ \cdots \circ \mathcal{M}_1'(\rho_{R_0 E}^{(0)})$ 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 $\delta > 0$

$$H_{\min}^\delta(A_1^n | B_1^n E)_\rho \geq \sum_{k=1}^n \inf_\omega H(A_k | B_k \tilde{R}_{k-1})_{\mathcal{M}_k'(\omega)} - n O(\epsilon^{\frac{1}{24}}) - O\left(\frac{1}{\epsilon^{\frac{1}{24}}} \right) \quad (6)$$

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

In the companion paper [1], we use the techniques developed in this paper to provide a solution for the source correlation problem in quantum key distribution (QKD) [16]. Briefly speaking, the security proofs of QKD require that one of the honest parties produce randomly and independently sampled quantum states in each round of the protocol. However, the states produced by a realistic quantum source will be somewhat correlated across different rounds due to imperfections. These correlations are called source correlations. Proving security for QKD under such a correlated source is challenging and no general satisfying solution was known. In [1], we use the entropic triangle inequality to

---

[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.

reduce the security of a QKD protocol with a correlated source to that of the QKD protocol with a depolarised variant of the perfect source, for which security can be proven using existing techniques.

2. Background and Notation

For n quantum registers $(X_1, X_2, \ldots, X_n)$, the notation X_i^j refers to the set of registers $(X_i, X_{i+1}, \ldots, X_j)$. We use the notation [n] to denote the set $\{1, 2, \ldots, n\}$. 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 (or briefly just *state*) refers to a positive semidefinite operator with unit trace. At times, we will also need to consider positive semidefinite operators with trace less than equal to 1. We call these operators subnormalised states. We will denote the set of registers a quantum state describes (equivalently, its Hilbert space) using a subscript. For example, a quantum state on registers A and B, will be written as ρ_{AB} and its partial states on registers A and B, will be denoted as ρ_A and ρ_B. 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 normalised 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 supp(X) to denote the support of the Hermitian operator X and use $X \ll Y$ to denote that supp$(X) \subseteq$ supp(Y).

The trace norm is defined as $\|X\|_1 := \mathrm{tr}\left((X^\dagger X)^{\frac{1}{2}}\right)$. 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

$$F_*(\rho, \sigma) := \left(\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 + \sqrt{(1 - \mathrm{tr}\,\rho)(1 - \mathrm{tr}\,\sigma)}\right)^2. \tag{7}$$

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

$$P(\rho, \sigma) = \sqrt{1 - F_*(\rho, \sigma)}. \tag{8}$$

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

$$\|\mathcal{N}_{A \to B}\|_\diamond := \max_{X_{AR}:\|X_{AR}\|_1 \leq 1} \|\mathcal{N}_{A \to B}(X_{AR})\|_1 \tag{9}$$

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 [17] 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

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

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

$$\tilde{D}_\alpha(P||Q) = \begin{cases} \frac{1}{\alpha-1} \log \frac{\mathrm{tr}(Q^{-\frac{\alpha'}{2}} P Q^{-\frac{\alpha'}{2}})^\alpha}{\mathrm{tr}(P)} & \text{if } (\alpha < 1 \text{ and } P \not\perp Q) \text{ or } (P \ll Q) \\ \infty & \text{else.} \end{cases} \quad (11)$$

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

$$D_{\max}(P||Q) := \inf \left\{ \lambda \in \mathbb{R} : P \leq 2^\lambda Q \right\}. \quad (12)$$

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

$$D(P||Q) := \begin{cases} \frac{\mathrm{tr}(P \log P - P \log Q)}{\mathrm{tr}(P)} & \text{if } (P \ll Q) \\ \infty & \text{else.} \end{cases} \quad (13)$$

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 [18,19]

$$\mathbb{D}(\mathcal{N} || \mathcal{M}) := \sup_{\rho_{AR}} \mathbb{D}(\mathcal{N}_{A \to B}(\rho_{AR}) || \mathcal{M}_{A \to B}(\rho_{AR})) \quad (14)$$

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 subnormalised state ρ_{AB}:

$$\bar{H}_\alpha^\uparrow(A|B)_\rho := \sup_{\sigma_B} -\bar{D}_\alpha(\rho_{AB} || \mathbb{1}_A \otimes \sigma_B) \quad (15)$$

$$\tilde{H}_\alpha^\uparrow(A|B)_\rho := \sup_{\sigma_B} -\tilde{D}_\alpha(\rho_{AB} || \mathbb{1}_A \otimes \sigma_B) \quad (16)$$

$$\bar{H}_\alpha^\downarrow(A|B)_\rho := -\bar{D}_\alpha(\rho_{AB} || \mathbb{1}_A \otimes \rho_B) \quad (17)$$

$$\tilde{H}_\alpha^\downarrow(A|B)_\rho := -\tilde{D}_\alpha(\rho_{AB} || \mathbb{1}_A \otimes \rho_B) \quad (18)$$

for appropriate α in the domain of the divergences. The supremum in the definition for $\bar{H}_\alpha^\uparrow$ and $\tilde{H}_\alpha^\uparrow$ 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 $H(A|B)$. $\tilde{H}_\infty^\uparrow(A|B)_\rho$ 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

$$H_{\min}(A|B)_\rho := \sup \left\{ \lambda \in \mathbb{R} : \text{there exists state } \sigma_B \text{ such that } \rho_{AB} \leq 2^{-\lambda} \mathbb{1}_A \otimes \sigma_B \right\}. \quad (19)$$

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

$$B_\epsilon(\rho) := \{ \tilde{\rho} \geq 0 : P(\rho, \tilde{\rho}) \leq \epsilon \text{ and } \mathrm{tr}\, \tilde{\rho} \leq 1 \}. \quad (20)$$

We define the smooth max-relative entropy as

$$D_{\max}^{\epsilon}(\rho||\sigma) := \min_{\tilde{\rho} \in B_{\epsilon}(\rho)} D_{\max}(\tilde{\rho}||\sigma) \tag{21}$$

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

$$H_{\min}^{\epsilon}(A|B)_{\rho} := \max_{\tilde{\rho} \in B_{\epsilon}(\rho)} H_{\min}(A|B)_{\tilde{\rho}}. \tag{22}$$

3. Entropic Triangle Inequality for the Smooth Min-entropy

In this section, we derive a simple entropic triangle inequality (Lemma 3.5) for the smooth min-entropy of the form in Eq. 4. This Lemma is a direct consequence of the following triangle inequality for $\tilde{D}_{\alpha}$ (see [20, Theorem 3.1] for a triangle inequality, which changes the second argument of $\tilde{D}_{\alpha}$).

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

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

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

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

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,

$$2^{(\alpha-1)\tilde{D}_{\alpha}(\rho||Q)} = \frac{\text{tr}\left(Q^{-\frac{\alpha-1}{2\alpha}} \rho Q^{-\frac{\alpha-1}{2\alpha}}\right)^{\alpha}}{\text{tr}(\rho)}$$

$$\leq \frac{\text{tr}\left(Q^{-\frac{\alpha-1}{2\alpha}} 2^{D_{\max}(\rho||\eta)} \eta Q^{-\frac{\alpha-1}{2\alpha}}\right)^{\alpha}}{\text{tr}(\rho)}$$

$$= \frac{\text{tr}(\eta)}{\text{tr}(\rho)} 2^{\alpha D_{\max}(\rho||\eta)} 2^{(\alpha-1)\tilde{D}_{\alpha}(\eta||Q)}$$

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

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

$$\tilde{H}^{\uparrow}_{\alpha,\epsilon}(A|B)_\rho := \max_{\tilde{\rho}_{AB} \in B_\epsilon(\rho_{AB})} \tilde{H}^{\uparrow}_\alpha(A|B)_{\tilde{\rho}}. \tag{23}$$

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

$$\tilde{D}_\alpha(\tilde{\rho}_{AB}|| \mathbb{1}_A \otimes \sigma_B) \leq \tilde{D}_\alpha(\eta_{AB}|| \mathbb{1}_A \otimes \sigma_B) + \frac{\alpha}{\alpha - 1} D^\epsilon_{\max}(\rho_{AB}||\eta_{AB}) + \frac{1}{\alpha - 1} \log \frac{1}{1 - \epsilon^2} \tag{24}$$

where we used the fact that $\tilde{\rho}_{AB} \in B_\epsilon(\rho_{AB})$ which implies that $\text{tr}(\tilde{\rho}_{AB}) \geq 1 - \epsilon^2$. 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 $\tilde{H}^{\uparrow}_{\alpha,\epsilon}(A|B)_\rho \geq \tilde{H}^{\uparrow}_\alpha(A|B)_{\tilde{\rho}}$. $\square$

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

$$H^{\epsilon+\delta}_{\min}(A|B)_\rho \geq \sup_{\tilde{\rho} \in B_\epsilon(\rho_{AB})} H^\delta_{\min}(A|B)_{\tilde{\rho}}. \tag{25}$$

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

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

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. 25 is true.

Using this, we have

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

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

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

where we have used [8, Lemma B.10][5] (originally proven in [21]) in the second step. ☐

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

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

$$H^{\epsilon+\delta}_{\min}(A|B)_\rho \geq \tilde{H}^{\uparrow}_\alpha(A|B)_\eta - \frac{\alpha}{\alpha - 1} D^\epsilon_{\max}(\rho_{AB}||\eta_{AB}) - \frac{g_1(\delta, \epsilon)}{\alpha - 1} \tag{26}$$

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:

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

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

☐

We can use the asymptotic equipartition theorem for smooth min-entropy and max-relative entropy [21–23] 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*

$$H(A|B)_\rho \geq \tilde{H}^{\uparrow}_\alpha(A|B)_\eta - \frac{\alpha}{\alpha - 1} D(\rho_{AB}||\eta_{AB}). \tag{27}$$

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

$$H^{\epsilon+\delta}_{\min}(A^n_1|B^n_1)_{\rho^{\otimes n}} \geq \tilde{H}^{\uparrow}_\alpha(A^n_1|B^n_1)_{\eta^{\otimes n}} - \frac{\alpha}{\alpha - 1} D^\epsilon_{\max}(\rho^{\otimes n}_{AB}||\eta^{\otimes n}_{AB}) - \frac{g_1(\delta, \epsilon)}{\alpha - 1}$$

$$\Rightarrow \frac{1}{n} H^{\epsilon+\delta}_{\min}(A^n_1|B^n_1)_{\rho^{\otimes n}} \geq \tilde{H}^{\uparrow}_\alpha(A|B)_\eta - \frac{\alpha}{\alpha - 1}\frac{1}{n} D^\epsilon_{\max}(\rho^{\otimes n}_{AB}||\eta^{\otimes n}_{AB}) - \frac{1}{n}\frac{g_1(\delta, \epsilon)}{\alpha - 1}.$$

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

$$\lim_{n\to\infty} \frac{1}{n} H^{\epsilon+\delta}_{\min}(A^n_1|B^n_1)_{\rho^{\otimes n}} \geq \tilde{H}^{\uparrow}_\alpha(A|B)_\eta - \lim_{n\to\infty} \frac{\alpha}{\alpha - 1}\frac{1}{n} D^\epsilon_{\max}(\rho^{\otimes n}_{AB}||\eta^{\otimes n}_{AB})$$

$$- \lim_{n\to\infty} \frac{1}{n}\frac{g_1(\delta, \epsilon)}{\alpha - 1}$$

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

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 [8, 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 $k \in [n]$

$$\left\| \rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1} B} \right\|_1 \leq \epsilon. \tag{28}$$

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

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

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

$$\epsilon \geq \frac{1}{2}\|p - q\|_1 = \max_{H \subseteq \mathcal{X}} |p(H) - q(H)|$$

$$\geq q(S^c) \left| \frac{p(S^c)}{q(S^c)} - 1 \right|$$

$$\geq q(S^c) \left(\frac{p(S^c)}{q(S^c)} - 1 \right)$$

$$= q(S^c) \left(\frac{\sum_{x \in S^c} p(x)}{\sum_{x \in S^c} q(x)} - 1 \right)$$

$$\geq q(S^c) \left(\frac{\sum_{x \in S^c}(1 + \epsilon^{\frac{1}{2}})q(x)}{\sum_{x \in S^c} q(x)} - 1 \right)$$

$$= q(S^c)\epsilon^{\frac{1}{2}}$$

which implies that $q(S^c) \leq \epsilon^{\frac{1}{2}}$. Now, the statement of the Lemma follows. $\qquad\square$

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

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

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

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

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

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

$$\rho(a_1^n | b) = \prod_{k=1}^n \rho(a_k | a_1^{k-1}, b)$$

$$= \prod_{k:(a_1^n, b) \notin B_k} \rho(a_k | a_1^{k-1}, b) \prod_{k:(a_1^n, b) \in B_k} \rho(a_k | a_1^{k-1}, b)$$

$$\leq (1 + \sqrt{\epsilon})^n \prod_{k:(a_1^n, b) \notin B_k} \rho_{A_k}(a_k)$$

$$\leq (1 + \sqrt{\epsilon})^n 2^{-n(1-\epsilon^{\frac{1}{4}}) H_{\min}(A_1)}$$

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

$$H_{\min}^{\sqrt{2}\epsilon^{\frac{1}{8}}}(A_1^n | B) \geq n(1 - \epsilon^{\frac{1}{4}}) H_{\min}(A_1) - n \log(1 + \sqrt{\epsilon}). \tag{29}$$

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. 28 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. Furthermore, 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 [10,24] provides an upper bound on the smooth max relative entropy $D_{\max}^\epsilon(\rho||\sigma)$ between two states in terms of their relative entropy $D(\rho||\sigma)$.

Theorem 4.2 (Quantum substate theorem [24]). *Let ρ and σ be two states on the same Hilbert space. Then for any $\epsilon \in (0, 1)$, we have*

$$D_{\max}^{\sqrt{\epsilon}}(\rho||\sigma) \leq \frac{D(\rho||\sigma) + 1}{\epsilon} + \log \frac{1}{1-\epsilon}. \tag{30}$$

In this section, we will also frequently use the multipartite mutual information [25–27]. For a state $\rho_{X_1^n}$, the multipartite mutual information between the registers $(X_1, X_2, \ldots, X_n)$ is defined as

$$I(X_1 : X_2 : \cdots : X_n)_\rho := D(\rho_{X_1^n}||\rho_{X_1} \otimes \rho_{X_2} \otimes \cdots \otimes \rho_{X_n}). \tag{31}$$

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:

$$I(X_1 : X_2 : \cdots : X_n)_\rho = \sum_{k=1}^{n} H(X_k)_\rho - H(X_1 \cdots X_n)_\rho \tag{32}$$

$$= \sum_{k=2}^{n} I(X_k : X_1^{k-1}). \tag{33}$$

Going back to proving a bound for the quantum approximately independent registers problem, note that using the Alicki–Fannes–Winter (AFW) bound [28,29] for mutual information [30, Theorem 11.10.4], Eq. 28 implies that for every $k \in [n]$

$$I(A_k : A_1^{k-1}B)_\rho \leq \epsilon \log |A| + g_2\left(\frac{\epsilon}{2}\right) \tag{34}$$

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*

$$I(A_k : A_1^{k-1}B)_\rho \leq \epsilon \tag{35}$$

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

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

(36)

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

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

(37)

Proof. First note that we have,

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

Using the substate theorem, we now have

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

(38)

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:

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

$$\geq \sum_{k=1}^{n} H(A_k)_\rho - n(\alpha - 1)\log^2(1 + 2|A|) - \frac{\alpha}{\alpha - 1}n\sqrt{\epsilon}$$

$$- \frac{\alpha}{\alpha - 1}\frac{1}{\sqrt{\epsilon}} - \frac{\alpha}{\alpha - 1}\log(1 - \sqrt{\epsilon}) - \frac{g_1(\epsilon, \epsilon^{\frac{1}{4}})}{\alpha - 1}.$$

In the third line above, we have used [8, Lemma B.9] (which is an improvement of [21, 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

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

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

$\square$

We can now plug the bound in Eq. 34 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*

$$\left\|\rho_{A_1^k B} - \rho_{A_k} \otimes \rho_{A_1^{k-1}B}\right\|_1 \leq \epsilon. \tag{39}$$

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

$$H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n|B)_\rho \geq \sum_{k=1}^{n} H(A_k)_\rho - 3n\delta^{\frac{1}{4}}\log(1 + 2|A|)$$

$$- \frac{2\log(1 + 2|A|)}{\delta^{3/4}} - \frac{2\log(1 + 2|A|)}{\delta^{1/4}}\left(\log(1 - \sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}})\right) \tag{40}$$

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

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

$$- \frac{2\log(1 + 2|A|)}{\delta^{3/4}} - \frac{2\log(1 + 2|A|)}{\delta^{1/4}}\left(\log(1 - \sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}})\right). \tag{41}$$

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*

$$\left\| \rho_{A_1^k B_1^k E} - \rho_{A_k B_k} \otimes \rho_{A_1^{k-1} B_1^{k-1} E} \right\|_1 \le \epsilon. \tag{42}$$

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

$$H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n | B_1^n E)_\rho \ge \sum_{k=1}^n H(A_k|B_k)_\rho - 3n\delta^{\frac{1}{4}} \log(1 + 2|A|)$$
$$- \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}}\left(\log(1 - \sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}})\right) \tag{43}$$

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

$$H_{\min}^{\delta^{\frac{1}{4}}+\delta}(A_1^n | B_1^n E)_\rho \ge n\left(H(A_1|B_1)_\rho - 3\delta^{\frac{1}{4}}\log(1 + 2|A|)\right)$$
$$- \frac{2\log(1+2|A|)}{\delta^{3/4}} - \frac{2\log(1+2|A|)}{\delta^{1/4}}\left(\log(1 - \sqrt{\delta}) + g_1(\delta, \delta^{\frac{1}{4}})\right). \tag{44}$$

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

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

where we used the AFW bound for mutual information in the last line [30, Theorem 11.10.4]. The rest of the proof follows the proof of Theorem 4.3, only difference being that now we have $\tilde{H}_\alpha^\uparrow(A_1^n | B_1^n E)_\eta = \sum_{k=1}^n \tilde{H}_\alpha^\uparrow(A_k|B_k)_\rho$. □

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

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 $\omega_{\mathrm{avg}} < \omega_{\mathrm{exp}}$
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).

<div align="center">

Protocol 1.

</div>

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 [31] 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 $r_{k-1} := x_1^{k-1}, y_1^{k-1}, t_1^{k-1}, a_1^{k-1}, b_1^{k-1}$ 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 is used for key generation. We can further take an average over all the previous round's random variables to derive the probability of winning the kth game

$$\Pr[W_k] = \mathbb{E}_{r_{k-1}}\left[\Pr[W_k|r_{k-1}]\right]. \tag{45}$$

Alice and Bob randomly sample a subset of the rounds (using the random variable T_k) and play the CHSH game on this subset. If the average winning probability of CHSH game on this subset is small, they abort the protocol. For simplicity and brevity, we will assume here that the state $\rho^{(0)}_{E_A E_B E}$ distributed between Alice and Bob at the start of the
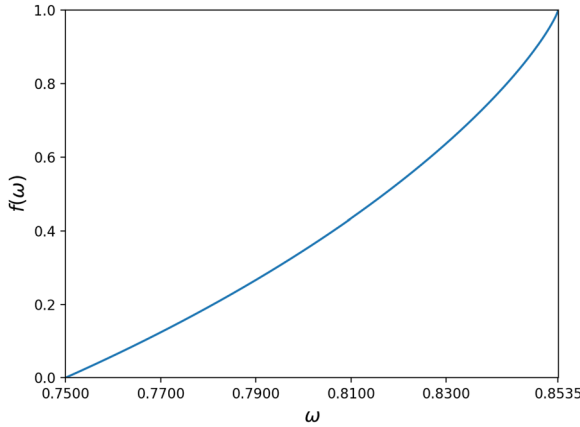
Fig. 1. The lower bound in Eq. 47 for the interval $[\frac{3}{4}, \frac{2+\sqrt{2}}{4}]$

protocol by Eve has an average winning probability at least $\omega_{\exp}$, that is,

$$\frac{1}{n}\sum_{k=1}^{n}\Pr[W_k] \geq \omega_{\exp} - \delta \tag{46}$$

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 $\sigma_{E_A E_B E}$ (where E_A is held by Alice, E_B is held by Bob and E is held by the adversary) and local measurement devices, if Alice and Bob win the CHSH game with a probability $\omega \in (\frac{3}{4}, \frac{2+\sqrt{2}}{4}]$, then Alice's answer A to the game is random given the questions X, Y and the register E held by adversary. This is quantified by the following entropic bound [32] (see [31, Lemma 5.3] for the following form)

$$H(A|XYE) \geq f(\omega) = \begin{cases} 1 - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16\omega(\omega-1)+3}\right) & \text{if } \omega \in [\frac{3}{4}, \frac{2+\sqrt{2}}{4}] \\ 0 & \text{if } \omega \in [0, \frac{3}{4}) \end{cases} \tag{47}$$

where $h(\cdot)$ is the binary entropy. The function f is convex over the interval $\left[0, \frac{2+\sqrt{2}}{4}\right]$. We plot it in the interval $[\frac{3}{4}, \frac{2+\sqrt{2}}{4}]$ in Fig. 1.

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

$$1 - f(\omega_{\exp} - \delta) = h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16(\omega_{\exp}-\delta)(\omega_{\exp}-\delta-1)+3}\right) \leq \epsilon^4. \tag{48}$$

We will now use Eq. 47 to bound the von Neumann entropy of the answers given Eve's information for the sequential DIQKD protocol. We have

$$H(A_1^n|X_1^n Y_1^n T_1^n E) = \sum_{k=1}^{n} H(A_k|A_1^{k-1} X_1^n Y_1^n T_1^n E)$$

$$\stackrel{(1)}{=} \sum_{k=1}^{n} H(A_k|A_1^{k-1} X_1^k Y_1^k T_1^k E)$$

$$\overset{(2)}{=} \sum_{k=1}^{n} H(A_k | X_k Y_k R_{k-1} E)$$

$$= \sum_{k=1}^{n} \mathbb{E}_{r_{k-1} \sim R_{k-1}} \left[H(A_k | X_k Y_k E)_{\rho^{(k)}_{|r_{k-1}}} \right]$$

$$\overset{(3)}{\geq} \sum_{k=1}^{n} \mathbb{E}_{r_{k-1} \sim R_{k-1}} \left[f\left(\Pr[W_k | r_{k-1}]\right) \right]$$

$$\geq \sum_{k=1}^{n} f\left(\Pr[W_k]\right)$$

$$\geq nf\left(\frac{1}{n} \sum_{k=1}^{n} \Pr[W_k] \right)$$

$$\geq nf(\omega_{\exp} - \delta) \geq n(1 - \epsilon^4)$$

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. 47, 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 [8,9] 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

$$I(A_1 : \cdots : A_n : X_1^n Y_1^n T_1^n E) = \sum_{k=1}^{n} H(A_k) + H(X_1^n Y_1^n T_1^n E) - H(A_1^n X_1^n Y_1^n T_1^n E)$$

$$= \sum_{k=1}^{n} H(A_k) - H(A_1^n | X_1^n Y_1^n T_1^n E)$$

$$\leq n - n(1 - \epsilon^4) = n\epsilon^4$$

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

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

$$= n(1 - 3\epsilon \log 5) - O\left(\frac{1}{\epsilon^3}\right) \tag{49}$$
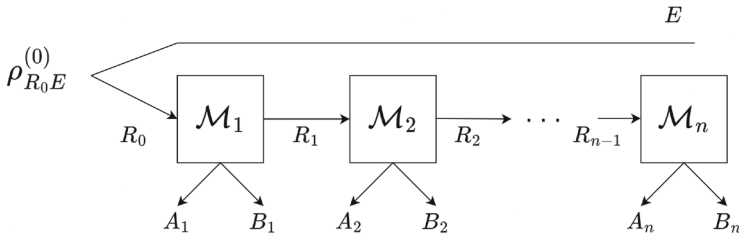
Fig. 2. The setting for entropy accumulation and Theorem 5.1. For $k \in [n]$, the channels $\mathcal{M}_k$ are repeatedly applied to the registers R_{k-1} to produce the "secret" information A_k and the side information B_k

where we have used the fact that the answers A_k can always be assumed to be uniformly distributed [31,32]. 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.

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) [8] provides a tight and simple lower bound for the smooth min-entropy $H^\epsilon_{\min}(A_1^n | B_1^n E)_\rho$ of sequential processes, under certain Markov chain conditions. The state $\rho_{A_1^n B_1^n E}$ in the setting for EAT is produced by a sequential process of the form shown in Fig. 2. The process begins with the registers R_0 and E. In the context of a cryptographic protocol, the register R_0 is usually held by the honest parties, whereas the register E is held by the adversary. Then, in each round $k \in [n]$ 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

$$H^\epsilon_{\min}(A_1^n | B_1^n E)_\rho \geq \sum_{k=1}^n \inf_{\omega_{R_{k-1}\tilde{R}}} H(A_k | B_k \tilde{R})_{\mathcal{M}_k(\omega)} - c\sqrt{n} \tag{50}$$

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

an approximate version of EAT. Consider the sequential process in Fig. 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

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

According to the Markov chain assumption for the channels $\mathcal{M}'_k$, the state $\sigma_{A_1^k B_1^k E}^{(k)}$, 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. The entropy accumulation theorem itself is used to prove the security of sequential device independent quantum key distribution (DIQKD) protocols [12]. 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 [13,14]). The approximate entropy accumulation theorem below can be used to prove security of DIQKD under such imperfections. We do not, however, pursue this example here and leave applications of this theorem for future work. In Sect. 5.4, we modify this Theorem to incorporate testing for EAT.

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*

$$\rho_{A_1^n B_1^n E} = \mathrm{tr}_{R_n} \circ \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)}) \tag{51}$$

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:*

$$\frac{1}{2} \left\| \mathcal{M}_k - \mathcal{M}'_k \right\|_\diamond \leq \epsilon \tag{52}$$

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 $\mathcal{M}'_k \circ \mathcal{N}_{k-1} \circ \cdots \circ \mathcal{N}_1(\rho_{R_0 E}^{(0)})$ satisfies the Markov chain*

$$A_1^{k-1} \leftrightarrow B_1^{k-1}E \leftrightarrow B_k. \tag{53}$$

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

$$
\begin{aligned}
H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n|B_1^n E)_\rho \geq{} & \sum_{k=1}^n \inf_{\omega_{R_{k-1}\tilde{R}}} H(A_k|B_k\tilde{R})_{\mathcal{M}_k'(\omega)} - n(\alpha-1)\log^2(1+2|A|) \\
& - \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(4^{\frac{\alpha-1}{\alpha}\log(|A||B|)} - 1\right)\right) \\
& - \frac{\alpha}{\alpha-1} n z_\beta(\epsilon,\delta) - \frac{1}{\alpha-1}\left(g_1(\epsilon_2,\epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta-1}\right).
\end{aligned}
\tag{54}
$$

where

$$
z_\beta(\epsilon,\delta) := \frac{\beta+1}{\beta-1}\log\left(\left(1 + \sqrt{(1-\delta)\epsilon}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta\beta}\right)^{\frac{1}{\beta+1}}\right)
\tag{55}
$$

and $g_1(x,y) = -\log(1-\sqrt{1-x^2}) - \log(1-y^2)$ and the infimum in Eq. 54 is taken over all input states $\omega_{R_{k-1}\tilde{R}}$ to the channels $\mathcal{M}_k'$ where $\tilde{R}$ is a reference register ($\tilde{R}$ can be chosen isomorphic to R_{k-1}).

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

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

We also have that

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

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

$$
\begin{aligned}
H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n|B_1^n E)_\rho \geq{} & \sum_{k=1}^n \inf_{\omega_{R_k\tilde{R}_k}} H(A_k|B_k\tilde{R}_k)_{\mathcal{M}_k'(\omega_{R_k\tilde{R}_k})} \\
& - n\sqrt{\epsilon_r}(\log^2(1+2|A|)+2) - \frac{1}{\sqrt{\epsilon_r}}(g_1(\epsilon_2,\epsilon_1)+2g_0(\epsilon_1))
\end{aligned}
\tag{56}
$$

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. 56, 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 Sect. 5.5.

In order to prove this theorem, we will use a channel divergence based chain rule. Recently proven chain rules for α-Rényi relative entropy [15, Corollary 5.2] state that for $\alpha > 1$ and states ρ_A and σ_A, and channels $\mathcal{E}_{A \to B}$ and $\mathcal{F}_{A \to B}$, we have

$$\tilde{D}_\alpha(\mathcal{E}_{A \to B}(\rho_A)||\mathcal{F}_{A \to B}(\sigma_A)) \leq \tilde{D}_\alpha(\rho_A||\sigma_A) + \tilde{D}_\alpha^{\text{reg}}(\mathcal{E}_{A \to B}||\mathcal{F}_{A \to B}) \qquad (57)$$

where $\tilde{D}_\alpha^{\text{reg}}(\mathcal{E}_{A \to B}||\mathcal{F}_{A \to B}) := \lim_{n \to \infty} \frac{1}{n} \tilde{D}_\alpha(\mathcal{E}_{A \to B}^{\otimes n}||\mathcal{F}_{A \to B}^{\otimes n})$ and $\tilde{D}_\alpha(\cdot||\cdot)$ is the channel divergence (see Eq. 14).

Now observe that if we were guaranteed that for the maps in Theorem 5.1 above, $\tilde{D}_\alpha^{\text{reg}}(\mathcal{M}_k || \mathcal{M}_k') \leq \epsilon$ for every k for some $\alpha > 1$. Then, we could use the chain rule in Eq. 57 as follows

$$\tilde{D}_\alpha(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})|| \mathcal{M}_n' \circ \cdots \circ \mathcal{M}_1'(\rho_{R_0 E}^{(0)}))$$

$$\leq \tilde{D}_\alpha(\mathcal{M}_{n-1} \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})|| \mathcal{M}_{n-1}' \circ \cdots \circ \mathcal{M}_1'(\rho_{R_0 E}^{(0)})) + \tilde{D}_\alpha^{\text{reg}}(\mathcal{M}_n || \mathcal{M}_n')$$

$$\leq \cdots$$

$$\leq \tilde{D}_\alpha(\rho_{R_0 E}^{(0)}||\rho_{R_0 E}^{(0)}) + \sum_{k=1}^{n} \tilde{D}_\alpha^{\text{reg}}(\mathcal{M}_k || \mathcal{M}_k')$$

$$\leq n\epsilon.$$

Once we have the above result we can simply use the well known relation between smooth max-relative entropy and α-Rényi relative entropy [17, Proposition 6.5] to get the bound

$$D_{\max}^{\epsilon'}(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})|| \mathcal{M}_n' \circ \cdots \circ \mathcal{M}_1'(\rho_{R_0 E}^{(0)}))$$

$$\leq \tilde{D}_\alpha(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})|| \mathcal{M}_n' \circ \cdots \circ \mathcal{M}_1'(\rho_{R_0 E}^{(0)})) + \frac{g_0(\epsilon')}{\alpha - 1}$$

$$\leq n\epsilon + O(1).$$

This bound can subsequently be used in Lemma 3.5 to relate the smooth min-entropy of the real state $\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})$ with the α−Rényi conditional entropy of the auxiliary state $\mathcal{M}_n' \circ \cdots \circ \mathcal{M}_1'(\rho_{R_0 E}^{(0)})$, 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 $\tilde{D}_\alpha^{\text{reg}}$ or any other channel divergence. We will get around this issue by instead using mixed channels $\mathcal{M}_k^\delta := (1 - \delta) \mathcal{M}_k' + \delta \mathcal{M}_k$. Also, instead of trying to bound channel divergence in terms of $\tilde{D}_\alpha^{\text{reg}}$, 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. [15] (see [33] for the following equivalent definition) in this section. For $\alpha > 1$ and two positive operators P and Q, it is defined

$$D_\alpha^\#(P||Q) := \min_{A \geq P} \hat{D}_\alpha(A||Q) \qquad (58)$$

where $\hat{D}_\alpha(A||Q)$ is the α-Rényi geometric divergence [34]. For $\alpha > 1$, it is defined as

$$\hat{D}_\alpha(A||Q) = \begin{cases} \frac{1}{\alpha-1} \log \operatorname{tr} \left(Q \left(Q^{-\frac{1}{2}} A Q^{-\frac{1}{2}} \right)^\alpha \right) & \text{if } A \ll Q \\ \infty & \text{otherwise.} \end{cases} \tag{59}$$

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 [17, Sect. 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*

$$X \leq (1 + t)\Pi X \Pi + \left(1 + \frac{1}{t}\right) \Pi_\perp X \Pi_\perp. \tag{60}$$

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

$$0 \leq \begin{pmatrix} t X_1 & -X_2 \\ -X_2^* & \frac{1}{t} X_3 \end{pmatrix}.$$

This is true because

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

since $X \geq 0$. $\qquad\square$

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

$$D_\alpha^\#(\rho||\sigma) \leq \frac{\alpha + 1}{\alpha - 1} \log \left((1 + \sqrt{\epsilon})^{\frac{\alpha}{\alpha+1}} + \left(2^{\alpha d} \sqrt{\epsilon}\right)^{\frac{1}{\alpha+1}} \right). \tag{61}$$

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

Further, let

$$\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}} = \sum_{i=1}^{n} \lambda_i |x_i\rangle\langle x_i| \tag{62}$$

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

$$\mathbb{E}_{I \sim q}[\lambda_I] = \sum_{i=1}^{n} \lambda_i \langle x_i|\sigma|x_i\rangle$$

$$= \mathrm{tr}\left(\sigma \sum_{i=1}^{n} \lambda_i |x_i\rangle\langle x_i|\right)$$

$$= \mathrm{tr}\left(\sigma\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}}\right)$$

$$= \mathrm{tr}(P)$$

$$\leq \epsilon.$$

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

$$S := \{i \in [n] : \lambda_i \leq \sqrt{\epsilon}\}. \tag{63}$$

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

$$\Pr_q(I \in S^c) = \Pr_q(\lambda_I > \sqrt{\epsilon})$$

$$\leq \frac{\mathbb{E}_{I \sim q}[\lambda_I]}{\sqrt{\epsilon}}$$

$$\leq \sqrt{\epsilon}.$$

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

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

$$= \Pr_q(I \in S^c)$$

$$\leq \sqrt{\epsilon}. \tag{64}$$

Moreover, by the definition of set S (Eq. 63) we have

$$\Pi\sigma^{-\frac{1}{2}} P \sigma^{-\frac{1}{2}}\Pi = \sum_{i \in S} \lambda_i |x_i\rangle\langle x_i| \leq \sqrt{\epsilon}\Pi \tag{65}$$

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

$$\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} \leq 2^d \, \mathbb{1} \,. \tag{66}$$

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

$$\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} \leq (1+t)\Pi\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\Pi + \left(1 + \frac{1}{t}\right)\Pi_\perp\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\Pi_\perp$$

$$\leq (1+t)\Pi\left(\mathbb{1} + \sigma^{-\frac{1}{2}}P\sigma^{-\frac{1}{2}}\right)\Pi + \left(1 + \frac{1}{t}\right)2^d\Pi_\perp$$

$$\leq (1+t)(1+\sqrt{\epsilon})\Pi + \left(1 + \frac{1}{t}\right)2^d\Pi_\perp$$

where we have used $\rho \leq \sigma + P$ to bound the first term and Eq. 66 to bound the second term in the second line, and Eq. 65 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:

$$\hat{D}_\alpha(A_t||\sigma) = \frac{1}{\alpha-1}\log\operatorname{tr}\left(\sigma\left(\sigma^{-\frac{1}{2}}A_t\sigma^{-\frac{1}{2}}\right)^\alpha\right)$$

$$= \frac{1}{\alpha-1}\log\operatorname{tr}\left(\sigma\left((1+t)(1+\sqrt{\epsilon})\Pi + \left(1 + \frac{1}{t}\right)2^d\Pi_\perp\right)^\alpha\right)$$

$$= \frac{1}{\alpha-1}\log\operatorname{tr}\left(\sigma\left((1+t)^\alpha(1+\sqrt{\epsilon})^\alpha\Pi + \left(1 + \frac{1}{t}\right)^\alpha 2^{d\alpha}\Pi_\perp\right)\right)$$

$$= \frac{1}{\alpha-1}\log\left((1+t)^\alpha(1+\sqrt{\epsilon})^\alpha\operatorname{tr}(\sigma\Pi) + \left(1 + \frac{1}{t}\right)^\alpha 2^{d\alpha}\operatorname{tr}(\sigma\Pi_\perp)\right)$$

$$\leq \frac{1}{\alpha-1}\log\left((1+t)^\alpha(1+\sqrt{\epsilon})^\alpha + \left(1 + \frac{1}{t}\right)^\alpha 2^{d\alpha}\sqrt{\epsilon}\right)$$

where in the last line we use $\operatorname{tr}(\sigma\Pi) \leq 1$ and $\operatorname{tr}(\sigma\Pi_\perp) \leq \sqrt{\epsilon}$ (Eq. 64). 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:

$$
\begin{aligned}
\frac{1}{2} \,\|\mathcal{M}_\delta - \mathcal{N}\|_\diamond &= \frac{1}{2} \,\|(1 - \delta)\,\mathcal{M} + \delta\,\mathcal{N} - \mathcal{N}\|_\diamond \\
&= (1 - \delta)\frac{1}{2} \,\|\mathcal{M} - \mathcal{N}\|_\diamond \\
&\leq (1 - \delta)\epsilon.
\end{aligned}
\tag{67}
$$

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. 67 have that

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

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 $\frac{1}{2} \|\mathcal{N} - \mathcal{M}\|_\diamond \leq \epsilon$. For $\delta \in (0, 1)$, we can define the mixed channel $\mathcal{M}_\delta := (1 - \delta)\mathcal{M} + \delta\mathcal{N}$. Then, for every $\alpha \in (1, \infty)$, we have the following bound on the channel divergence*

$$D_\alpha^\#(\mathcal{N} \,||\, \mathcal{M}_\delta) \leq \frac{\alpha + 1}{\alpha - 1} \log \left(\left(1 + \sqrt{(1-\delta)\epsilon} \right)^{\frac{\alpha}{\alpha+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\alpha} \right)^{\frac{1}{\alpha+1}} \right). \quad (68)$$

5.3. Proof of the approximate entropy accumulation theorem. We use the mixed channels defined in the previous section to define the auxiliary state $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0 E}^{(0)})$ for our proof. It is easy to show using the divergence bounds in Sect. 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 $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0 E}^{(0)})$ does not necessarily satisfy the Markov chain conditions required for entropy accumulation. Thus, we also need to reprove the entropy lower bound on this state by modifying the approach used in the proof of the original entropy accumulation theorem.

Proof of Theorem 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

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

$$:= z_\beta(\epsilon, \delta) \quad (69)$$

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

$$\sigma_{A_1^n B_1^n E} := \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0 E}^{(0)}). \quad (70)$$

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

$$D_{\max}^{\epsilon_1}(\rho_{A_1^n B_1^n E} \,||\, \sigma_{A_1^n B_1^n E})$$

$$\leq \tilde{D}_\beta(\rho_{A_1^n B_1^n E} \,||\, \sigma_{A_1^n B_1^n E}) + \frac{g_0(\epsilon_1)}{\beta - 1}$$

$$\leq D_\beta^\#(\rho_{A_1^n B_1^n E} \,||\, \sigma_{A_1^n B_1^n E}) + \frac{g_0(\epsilon_1)}{\beta - 1}$$

$$= D_\beta^\#(\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)}) \,||\, \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0 E}^{(0)})) + \frac{g_0(\epsilon_1)}{\beta - 1}$$

$$\leq D_\beta^\#(\mathcal{M}_{n-1} \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)}) \,||\, \mathcal{M}_{n-1}^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0 E}^{(0)})) + D_\beta^\#(\mathcal{M}_n \,||\, \mathcal{M}_n^\delta) + \frac{g_0(\epsilon_1)}{\beta - 1}$$

$$\leq \cdots$$

$$\leq \sum_{k=1}^{n} D_{\beta}^{\#}(\mathcal{M}_k \| \mathcal{M}_k^{\delta}) + \frac{g_0(\epsilon_1)}{\beta - 1}$$

$$\leq n z_{\beta}(\epsilon, \delta) + \frac{g_0(\epsilon_1)}{\beta - 1} \tag{71}$$

where the first line follows from [17, Proposition 6.5], the second line follows from [15, Proposition 3.4], fourth line follows from the chain rule for $D_{\beta}^{\#}$ [15, Proposition 4.5], and the last line follows from Eq. 69.

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

$$H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n | B_1^n E)_{\rho} \geq \tilde{H}_{\alpha}^{\uparrow}(A_1^n | B_1^n E)_{\sigma} - \frac{\alpha}{\alpha - 1} n z_{\beta}(\epsilon, \delta)$$

$$- \frac{1}{\alpha - 1} \left(g_1(\epsilon_2, \epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta - 1} \right). \tag{72}$$

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

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

1. Sample the classical random variable $C_k \in \{0, 1\}$ independently. $C_k = 1$ with probability $1 - \delta$ and 0 otherwise.
2. If $C_k = 1$ apply the map $\mathcal{M}_k'$ on R_{k-1}, else apply $\mathcal{M}_k$ on R_{k-1}.

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

$$\tilde{H}_{\alpha}^{\uparrow}(A_1^n | B_1^n E)_{\sigma} = \tilde{H}_{\alpha}^{\uparrow}(A_1^n | B_1^n E)_{\theta}$$

$$\geq \tilde{H}_{\alpha}^{\uparrow}(A_1^n | B_1^n C_1^n E)_{\theta}. \tag{73}$$

We will now focus on lower bounding $\tilde{H}_{\alpha}^{\uparrow}(A_1^n | B_1^n C_1^n E)_{\theta}$. Using [17, Proposition 5.1], we have that

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

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

Claim 5.5. *Define* $h_k := \inf_\omega \tilde{H}_\alpha^\downarrow (A_k|B_k\tilde{R}_{k-1})_{\mathcal{M}_k'(\omega)}$ *where the infimum is over all states* $\omega_{R_{k-1}\tilde{R}_{k-1}}$ *for a register* $\tilde{R}_{k-1}$, *which is isomorphic to* R_{k-1}, *and* $s := \log(|A||B|^2)$. *Then, we have*

$$\tilde{H}_\alpha^\uparrow (A_1^n|B_1^n E)_{\theta_{|c_1^n}} \geq \sum_{k=1}^n (\delta(c_k, 1)h_k - \delta(c_k, 0)s) \tag{74}$$

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

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

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

$$\tilde{H}_\alpha^\uparrow (A_1^k|B_1^k E)_{\theta_{|c_1^k}} \geq \tilde{H}_\alpha^\uparrow (A_1^{k-1}|B_1^k E)_{\theta_{|c_1^k}} - \log|A|$$

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

$$= \tilde{H}_\alpha^\uparrow (A_1^{k-1}|B_1^{k-1} E)_{\theta_{|c_1^{k-1}}} - s$$

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 $\theta_{R_{k-1} A_1^{k-1} B_1^{k-1} E|c_1^k} = \Phi_{k-1} \circ \Phi_{k-2} \cdots \circ \Phi_1(\rho_{R_0 E}^{(0)})$ 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 $\tilde{H}_\alpha^\uparrow$ counterpart for [8, Corollary 3.5], which is the main chain rule used for proving entropy accumulation), we have

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

$$= \tilde{H}_\alpha^\uparrow (A_1^{k-1}|B_1^{k-1} E)_{\theta_{|c_1^{k-1}}} + h_k$$

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

$$\tilde{H}_\alpha^\uparrow (A_1^k|B_1^k E)_{\theta_{|c_1^k}} \geq \tilde{H}_\alpha^\uparrow (A_1^{k-1}|B_1^{k-1} E)_{\theta_{|c_1^{k-1}}} + (\delta(c_k, 1)h_k - \delta(c_k, 0)s). \tag{75}$$

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

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

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

$$= \sum_{c_1^n} \prod_{k=1}^n \theta(c_k) 2^{\frac{1-\alpha}{\alpha} l_k(c_k)}$$

$$= \prod_{k=1}^n \sum_{c_k} \theta(c_k) 2^{\frac{1-\alpha}{\alpha} l_k(c_k)}. \tag{76}$$

Then, we have

$$\tilde{H}_\alpha^\uparrow(A_1^n | B_1^n C_1^n E)_\theta = \frac{\alpha}{1-\alpha} \log \sum_{c_1^n} \theta(c_1^n) \exp_2\left(\frac{1-\alpha}{\alpha} \tilde{H}_\alpha^\uparrow(A_1^n | B_1^n E)_{\theta_{|c_1^n}}\right).$$

$$\geq \frac{\alpha}{1-\alpha} \sum_{k=1}^n \log \sum_{c_k} \theta(c_k) 2^{\frac{1-\alpha}{\alpha} l_k(c_k)}$$

$$= \frac{\alpha}{1-\alpha} \sum_{k=1}^n \log\left((1-\delta)2^{\frac{1-\alpha}{\alpha} h_k} + \delta 2^{-\frac{1-\alpha}{\alpha} s}\right)$$

$$= \sum_{k=1}^n h_k - \frac{\alpha}{\alpha-1} \sum_{k=1}^n \log\left(1 - \delta + \delta 2^{\frac{\alpha-1}{\alpha}(s+h_k)}\right)$$

$$\geq \sum_{k=1}^n h_k - \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(2^{\frac{\alpha-1}{\alpha}(s+\log|A|)} - 1\right)\right) \tag{77}$$

where in the second line we have used Eq. 76 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 [8, Lemma B.9] to transform the above to

$$\tilde{H}_\alpha^\uparrow(A_1^n | B_1^n C_1^n E)_\theta \geq \sum_{k=1}^n \inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} H(A_k | B_k \tilde{R}_{k-1})_{\mathcal{M}_k'(\omega)} - n(\alpha-1)\log^2(1+2|A|)$$

$$- \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(2^{\frac{\alpha-1}{\alpha} 2\log(|A||B|)} - 1\right)\right). \tag{78}$$

Putting Eqs. 72, 73, and 78 together, we have

$$H_{\min}^{\epsilon_1+\epsilon_2}(A_1^n | B_1^n E)_\rho \geq \sum_{k=1}^n \inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} H(A_k | B_k \tilde{R}_{k-1})_{\mathcal{M}_k'(\omega)} - n(\alpha-1)\log^2(1+2|A|)$$

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

$$- \frac{\alpha}{\alpha-1} n z_\beta(\epsilon, \delta) \frac{1}{\alpha-1}\left(g_1(\epsilon_2, \epsilon_1) + \frac{\alpha g_0(\epsilon_1)}{\beta-1}\right).$$

$$\square$$

5.4. Testing. We will now incorporate testing into the approximate entropy accumulation theorem proven in Theorem 5.1. We follow [9], which is itself based on [35] for this purpose. Testing enables one to prove a lower bound on the smooth min-entropy of a state produced by the process in Fig. 2 conditioned on the output of a classical event. This is particularly useful for proving tight and practical bounds in cryptographic protocols.

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

$$\mathcal{T}_k(\omega_{A_k B_k}) = \sum_{y,z} \Pi^y_{A_k} \otimes \Pi^z_{B_k} \omega_{A_k B_k} \Pi^y_{A_k} \otimes \Pi^z_{B_k} \otimes |f(y,z)\rangle\langle f(y,z)|_{X_k} \tag{79}$$

where $\{\Pi^y_{A_k}\}_y$ and $\{\Pi^z_{B_k}\}_z$ are orthogonal projectors and f is some deterministic function which uses the measurements y and z to create the output register X_k.

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

$$\Sigma_k(q|\mathcal{N}_k) := \{\nu_{A_k B_k X_k R_k R} = \mathcal{N}_k(\omega_{R_{k-1}R}) : \text{ for a state } \omega_{R_{k-1}R} \text{ such that } \nu_{X_k} = q\}. \tag{80}$$

Definition 5.6. A function $f : \mathbb{P} \to \mathbb{R}$ is called a min-tradeoff function for the channels $\{\mathcal{N}_k\}_{k=1}^n$ if for every $k \in [n]$, it satisfies

$$f(q) \leq \inf_{\nu \in \Sigma_k(q|\mathcal{N}_k)} H(A_k|B_k R)_\nu. \tag{81}$$

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

$$\mathrm{Max}(f) := \max_{q \in \mathbb{P}} f(q) \tag{82}$$

$$\mathrm{Min}(f) := \min_{q \in \mathbb{P}} f(q) \tag{83}$$

$$\mathrm{Min}_\Sigma(f) := \min_{q : \Sigma(q) \neq \varnothing} f(q) \tag{84}$$

$$\mathrm{Var}(f) := \max_{q : \Sigma(q) \neq \varnothing} \sum_x q(x) f(\delta_x)^2 - \left(\sum_x q(x) f(\delta_x)\right)^2 \tag{85}$$

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

Theorem 5.7. *For $k \in [n]$, let the registers A_k and B_k be such that $|A_k| = |A|$ and $|B_k| = |B|$. For $k \in [n]$, let $\mathcal{M}_k$ be channels from $R_{k-1} \to R_k A_k B_k X_k$ and*

$$\rho_{A_1^n B_1^n X_1^n E} = \mathrm{tr}_{R_n} \circ \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho^{(0)}_{R_0 E}) \tag{86}$$

be the state produced by applying these maps sequentially. Further, let $\mathcal{M}_k$ be such that $\mathcal{M}_k = \mathcal{T}_k \circ \mathcal{M}^{(0)}_k$ for $\mathcal{T}_k$ defined in Eq. 79 and some channel $\mathcal{M}^{(0)}_k : R_{k-1} \to R_k A_k B_k$. Suppose the channels $\mathcal{M}_k$ are such that for every $k \in [n]$, there exists a channel $\mathcal{M}'_k$ from $R_{k-1} \to R_k A_k B_k X_k$ such that

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

$$\frac{1}{2} \left\| \mathcal{M}_k - \mathcal{M}'_k \right\|_\diamond \le \epsilon \tag{87}$$

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

$$A_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k. \tag{88}$$

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

$$
\begin{aligned}
H_{\min}^{\epsilon_2+\epsilon_3} &(A_1^n | B_1^n E)_{\rho|\Omega} \\
\ge\ & nh - \frac{(\alpha-1)\ln(2)}{2} \left(\log(2|A|^2 + 1) + \sqrt{2 + \text{Var}(f)} \right)^2 - n(\alpha-1)^2 K_\alpha \\
& - \frac{\alpha}{\alpha-1} n \log \left(1 + \delta \left(4^{\frac{\alpha-1}{\alpha}(\log(|A||B|)+\text{Max}(f)-\text{Min}(f)+1)} - 1 \right) \right) \\
& - \frac{\alpha}{\alpha-1} n z_\beta(\epsilon,\delta) - \frac{1}{\alpha-1} \left(\alpha \log \frac{1}{P_\rho(\Omega)-\epsilon_1} + g_1(\epsilon_3,\epsilon_2) + \frac{\alpha g_0(\epsilon_1)}{\beta-1} \right). \tag{89}
\end{aligned}
$$

where

$$z_\beta(\epsilon,\delta) := \frac{\beta+1}{\beta-1} \log \left(\left(1 + \sqrt{(1-\delta)\epsilon}\right)^{\frac{\beta}{\beta+1}} + \left(\frac{\sqrt{(1-\delta)\epsilon}}{\delta^\beta} \right)^{\frac{1}{\beta+1}} \right) \tag{90}$$

$$
\begin{aligned}
K_\alpha := & \frac{1}{6(2-\alpha)^3 \ln(2)} 2^{(\alpha-1)(2\log|A|+(\text{Max}(f)-\text{Min}_\Sigma(f)))} \\
& \ln^3 \left(2^{(2\log|A|+(\text{Max}(f)-\text{Min}_\Sigma(f)))} + e^2 \right) \tag{91}
\end{aligned}
$$

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

Proof. Just as in the proof of Theorem 5.1, we define

$$\mathcal{M}_k^\delta := (1-\delta)\mathcal{M}'_k + \delta \mathcal{M}_k \tag{92}$$

for every k and the state

$$\sigma_{A_1^n B_1^n X_1^n E} := \mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho^{(0)}_{R_0 E}). \tag{93}$$

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

$$D_{\max}^{\epsilon_1}(\rho_{A_1^n B_1^n X_1^n E} \| \sigma_{A_1^n B_1^n X_1^n E}) \le n z_\beta(\epsilon,\delta) + \frac{g_0(\epsilon_1)}{\beta-1}. \tag{94}$$

Define $d_\beta := n z_\beta(\epsilon, \delta) + \frac{g_0(\epsilon_1)}{\beta - 1}$. The bound above implies that there exists a state $\tilde{\rho}_{A_1^n B_1^n X_1^n E}$, which is also classical on X_1^n such that

$$P\left(\rho_{A_1^n B_1^n X_1^n E}, \tilde{\rho}_{A_1^n B_1^n X_1^n E}\right) \leq \epsilon_1 \tag{95}$$

and

$$\tilde{\rho}_{A_1^n B_1^n X_1^n E} \leq 2^{d_\beta} \sigma_{A_1^n B_1^n X_1^n E}. \tag{96}$$

The registers X_1^n for $\tilde{\rho}$ can be chosen to be classical, since the channel measuring X_1^n only decreases the distance between $\tilde{\rho}$ and ρ, and the new state produced would also satisfy Eq. 96. As the registers X_1^n are classical for both σ and $\tilde{\rho}$, we can condition these states on the event Ω. We will call the probability of the event Ω for the state σ and $\tilde{\rho}$ $P_\sigma(\Omega)$ and $P_{\tilde{\rho}}(\Omega)$ respectively. Using Lemma G.1 and the Fuchs-van de Graaf inequality, we have

$$P\left(\rho_{A_1^n B_1^n X_1^n E|\Omega}, \tilde{\rho}_{A_1^n B_1^n X_1^n E|\Omega}\right) \leq 2\sqrt{\frac{\epsilon_1}{P_\rho(\Omega)}}. \tag{97}$$

Conditioning Eq. 96 on Ω, we get

$$P_{\tilde{\rho}}(\Omega) \tilde{\rho}_{A_1^n B_1^n X_1^n E|\Omega} \leq 2^{d_\beta} P_\sigma(\Omega) \sigma_{A_1^n B_1^n X_1^n E|\Omega}. \tag{98}$$

Together, the above two equations imply that

$$D_{\max}^{\epsilon_2}(\rho_{A_1^n B_1^n X_1^n E|\Omega} \| \sigma_{A_1^n B_1^n X_1^n E|\Omega}) \leq n z_\beta(\epsilon, \delta) + \frac{g_0(\epsilon_1)}{\beta - 1} + \log \frac{P_\sigma(\Omega)}{P_{\tilde{\rho}}(\Omega)} \tag{99}$$

for $\epsilon_2 := 2\sqrt{\frac{\epsilon_1}{P_\rho(\Omega)}}$.

For $\epsilon_3 > 0$ and $\alpha \in (1, 2)$, we can plug the above in the bound provided by Lemma 3.5 to get

$$H_{\min}^{\epsilon_2 + \epsilon_3}(A_1^n | B_1^n E)_{\rho|\Omega} \geq \tilde{H}_\alpha^\uparrow(A_1^n | B_1^n E)_{\sigma|\Omega} - \frac{\alpha}{\alpha - 1} n z_\beta(\epsilon, \delta)$$
$$- \frac{1}{\alpha - 1}\left(\alpha \log \frac{P_\sigma(\Omega)}{P_{\tilde{\rho}}(\Omega)} + g_1(\epsilon_3, \epsilon_2) + \frac{\alpha g_0(\epsilon_1)}{\beta - 1}\right). \tag{100}$$

Now, note that using Eq. 79 and [8, Lemma B.7] we have

$$\tilde{H}_\alpha^\uparrow(A_1^n | B_1^n E)_{\sigma|\Omega} = \tilde{H}_\alpha^\uparrow(A_1^n X_1^n | B_1^n E)_{\sigma|\Omega}. \tag{101}$$

For every k, we introduce a register D_k of dimension $|D_k| = \lceil 2^{\max(f) - \min(f)} \rceil$ and a channel $\mathcal{D}_k : X_k \to X_k D_k$ as

$$\mathcal{D}_k(\omega) := \sum_x \langle x|\omega|x\rangle |x\rangle\langle x| \otimes \tau_x \tag{102}$$

where for every x, the state τ_x is a mixture between a uniform distribution on $\{1, 2, \ldots \lfloor 2^{\max(f) - f(\delta_x)} \rfloor\}$ and a uniform distribution on $\{1, 2, \ldots \lceil 2^{\max(f) - f(\delta_x)} \rceil\}$, so that

$$H(D_k)_{\tau_x} = \max(f) - f(\delta_x) \tag{103}$$

where δ_x is the distribution with unit weight at element x.

Define the channels $\bar{\mathcal{M}}_k := \mathcal{D}_k \circ \mathcal{M}_k$, $\bar{\mathcal{M}}'_k := \mathcal{D}_k \circ \mathcal{M}'_k$ and $\bar{\mathcal{M}}^\delta_k := \mathcal{D}_k \circ \mathcal{M}^\delta_k = (1-\delta)\bar{\mathcal{M}}'_k + \delta\bar{\mathcal{M}}_k$ and the state

$$\bar{\sigma}_{A_1^n B_1^n X_1^n D_1^n E} := \bar{\mathcal{M}}^\delta_n \circ \cdots \bar{\mathcal{M}}^\delta_1(\rho^{(0)}_{R_0 E}) \tag{104}$$

Note that $\bar{\sigma}_{A_1^n B_1^n X_1^n E} = \sigma_{A_1^n B_1^n X_1^n E}$. [9, Lemma 4.5] implies that this satisfies

$$\tilde{H}^\uparrow_\alpha(A_1^n X_1^n | B_1^n E)_{\sigma|\Omega} = \tilde{H}^\uparrow_\alpha(A_1^n X_1^n | B_1^n E)_{\bar{\sigma}|\Omega}$$

$$\geq \tilde{H}^\uparrow_\alpha(A_1^n X_1^n D_1^n | B_1^n E)_{\bar{\sigma}|\Omega} - \max_{x_1^n \in \Omega} H_\alpha(D_1^n)_{\bar{\sigma}_{|x_1^n}} \tag{105}$$

For $x_1^n \in \Omega$, we have

$$H_\alpha(D_1^n)_{\bar{\sigma}_{|x_1^n}} \leq H(D_1^n)_{\bar{\sigma}_{|x_1^n}}$$

$$\leq \sum_{k=1}^n H(D_k)_{\tau_{x_k}}$$

$$= \sum_{k=1}^n \max(f) - f(\delta_{x_k})$$

$$= n\max(f) - nf(\mathrm{freq}(x_1^n))$$

$$\leq n\max(f) - nh. \tag{106}$$

We can get rid of the conditioning on the right-hand side of Eq. 105 by using [8, Lemma B.5]

$$\tilde{H}^\uparrow_\alpha(A_1^n X_1^n D_1^n | B_1^n E)_{\bar{\sigma}|\Omega} \geq \tilde{H}^\uparrow_\alpha(A_1^n X_1^n D_1^n | B_1^n E)_{\bar{\sigma}} - \frac{\alpha}{\alpha-1}\log\frac{1}{P_\sigma(\Omega)}. \tag{107}$$

We now show that the channels $\bar{\mathcal{M}}'_k$ satisfy the second condition in Theorem 5.1. For an arbitrary $k \in [n]$ and a sequence of channels $\mathcal{N}_i \in \{\bar{\mathcal{M}}_i, \bar{\mathcal{M}}'_i\}$ for every $1 \leq i < k$, let

$$\eta_{A_1^k B_1^k X_1^k D_1^k E} = \bar{\mathcal{M}}'_k \circ \mathcal{N}_{k-1} \cdots \circ \mathcal{N}_1(\rho^{(0)}_{R_0 E}).$$

For this state, we have

$$I(A_1^{k-1} D_1^{k-1} : B_k | B_1^{k-1} E)_\eta = I(A_1^{k-1} : B_k | B_1^{k-1} E)_\eta + I(D_1^{k-1} : B_k | A_1^{k-1} B_1^{k-1} E)_\eta$$

$$= 0$$

where $I(A_1^{k-1} : B_k | B_1^{k-1} E)_\eta = 0$ because of the condition in Eq. 88, and $I(D_1^{k-1} : B_k | A_1^{k-1} B_1^{k-1} E)_\eta = 0$ since X_1^{k-1} and hence D_1^{k-1} are determined by $A_1^{k-1} B_1^{k-1}$. This implies that for this state $A_1^{k-1} D_1^{k-1} \leftrightarrow B_1^{k-1} E \leftrightarrow B_k$. Thus, the maps $\bar{\mathcal{M}}_k$ and $\bar{\mathcal{M}}'_k$ satisfy the conditions required for applying Theorem 5.1. Specifically, we can use the bounds in Eqs. 73 and 77 for bounding α-conditional Rényi entropy in Eq. 107

$$\tilde{H}^\uparrow_\alpha(A_1^n X_1^n D_1^n | B_1^n E)_{\bar{\sigma}}$$

$$\geq \tilde{H}^\uparrow_\alpha(A_1^n D_1^n | B_1^n E)_{\bar{\sigma}}$$

$$\geq \sum_{k=1}^{n} \inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} \tilde{H}_\alpha^\downarrow(A_k D_k | B_k \tilde{R}_{k-1})_{\bar{\mathcal{M}}'_k(\omega)}$$
$$- \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(2^{\frac{\alpha-1}{\alpha} 2\log(|A||D||B|)} - 1\right)\right). \tag{108}$$

The analysis in the proof of [35, Proposition V.3] shows that the first term above can be bounded as

$$\inf_{\omega_{R_{k-1}\tilde{R}_{k-1}}} \tilde{H}_\alpha^\downarrow(A_k D_k | B_k \tilde{R}_{k-1})_{\bar{\mathcal{M}}'_k(\omega)}$$
$$\geq \text{Max}(f) - \frac{(\alpha-1)\ln(2)}{2}\left(\log(2|A|^2+1) + \sqrt{2+\text{Var}(f)}\right)^2 - (\alpha-1)^2 K_\alpha \tag{109}$$

Combining Eqs. 105, 106, 107, 108 and 109, we have

$$\tilde{H}_\alpha^\uparrow(A_1^n X_1^n | B_1^n E)_{\bar{\sigma}_{|\Omega}}$$
$$\geq nh - \frac{(\alpha-1)\ln(2)}{2}\left(\log(2|A|^2+1) + \sqrt{2+\text{Var}(f)}\right)^2 - n(\alpha-1)^2 K_\alpha$$
$$- \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(2^{\frac{\alpha-1}{\alpha} 2\log(|A||D||B|)} - 1\right)\right) - \frac{\alpha}{\alpha-1}\log\frac{1}{P_\sigma(\Omega)}. \tag{110}$$

Plugging this into Eq. 100, we get

$$H_{\min}^{\epsilon_2+\epsilon_3}(A_1^n | B_1^n E)_{\rho_{|\Omega}}$$
$$\geq nh - \frac{(\alpha-1)\ln(2)}{2}\left(\log(2|A|^2+1) + \sqrt{2+\text{Var}(f)}\right)^2 - n(\alpha-1)^2 K_\alpha$$
$$- \frac{\alpha}{\alpha-1} n \log\left(1 + \delta\left(4^{\frac{\alpha-1}{\alpha}(\log(|A||B|)+\max(f)-\min(f)+1)} - 1\right)\right)$$
$$- \frac{\alpha}{\alpha-1} n z_\beta(\epsilon,\delta) - \frac{1}{\alpha-1}\left(\alpha\log\frac{1}{P_\rho(\Omega)-\epsilon_1} + g_1(\epsilon_3,\epsilon_2) + \frac{\alpha g_0(\epsilon_1)}{\beta-1}\right). \tag{111}$$

where we have used $P_{\tilde{\rho}}(\Omega) \geq P_\rho(\Omega) - \epsilon_1$ since $\frac{1}{2}\|\rho - \tilde{\rho}\|_1 \leq P(\rho,\tilde{\rho}) \leq \epsilon_1$. Note that the probability of Ω under the auxiliary state σ cancels out. □

5.5. Limitations and further improvements.

As we pointed out previously, the dependence of the entropy loss per round on ϵ is very poor (behaves as $\sim \epsilon^{1/24}$) in Theorem 5.1. 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 $(1-\delta)\mathcal{M}'_k + \delta\tau_{A_k B_k} \otimes \text{tr}_{A_k B_k} \circ \mathcal{M}_k$ (which is what is done classically), where $\tau_{A_k B_k}$ is the completely mixed state on registers $A_k B_k$. Here, instead of mixing the channel

$\mathcal{M}'_k$ with $\mathcal{M}_k$, we mix it with $\tau_{A_k B_k} \otimes \operatorname{tr}_{A_k B_k} \circ \mathcal{M}_k$, which also keeps $D_{\max}(\mathcal{M}_k \,\|\, \mathcal{M}_k^\delta)$ small enough. Moreover, this definition ensures that the registers B_k produced by the map $\mathcal{M}_k^\delta$ always satisfy the Markov chain conditions. If it were possible to show that the divergence between the real state $\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 E}^{(0)})$ and the auxiliary state $\mathcal{M}_n^\delta \circ \cdots \circ \mathcal{M}_1^\delta(\rho_{R_0 E}^{(0)})$ is small for this definition of $\mathcal{M}_k^\delta$, then one could directly use the entropy accumulation theorem for lower bounding the entropy for the auxiliary state. We cannot do this in our proof as this definition of the mixed channel $\mathcal{M}_k^\delta$ also increases the distance from the original channel $\mathcal{M}_k$ to $\epsilon + 2\delta$ and this makes the upper bound in Lemma 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. The necessity of such a bound also rules out a similar approximate extension of the generalised entropy accumulation theorem (GEAT).

Appendices

Appendix 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

$$H_{\min}^{\epsilon'}(A|B)_\rho \geq H_{\min}^{\epsilon}(A|B)_\eta - O(D_{\max}^{\epsilon''}(\rho\|\eta)) \tag{112}$$

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. 112, were true then we would have

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

$$\leq H_{\min}^{\epsilon'}(A|B)_{p|E} = O(1)$$

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.

Appendix 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*

$$D_\alpha(p||q) \leq \frac{1}{\alpha - 1} \log\left((1 + \sqrt{\epsilon})^{\alpha-1}(1 - 2\sqrt{\epsilon}) + 2^{d(\alpha-1)+1}\sqrt{\epsilon}\right). \tag{113}$$

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

$$D(p||q) \leq (1 - 2\sqrt{\epsilon}) \log(1 + \sqrt{\epsilon}) + 2\sqrt{\epsilon}d. \tag{114}$$

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

$$\sum_{x \in \mathcal{X}} p(x) \left(\frac{p(x)}{q(x)}\right)^{\alpha-1} = \sum_{x \in S} p(x) \left(\frac{p(x)}{q(x)}\right)^{\alpha-1} + \sum_{x \notin S} p(x) \left(\frac{p(x)}{q(x)}\right)^{\alpha-1}$$

$$\leq \sum_{x \in S} (1 + \sqrt{\epsilon})^{\alpha-1} p(x) + \sum_{x \notin S} 2^{d(\alpha-1)} p(x)$$

$$= (1 + \sqrt{\epsilon})^{\alpha-1} p(S) + 2^{d(\alpha-1)} p(S^c)$$

$$\leq (1 + \sqrt{\epsilon})^{\alpha-1}(1 - 2\sqrt{\epsilon}) + 2^{d(\alpha-1)+1}\sqrt{\epsilon}$$

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. $\qquad\square$

We observed in Sect. 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. 114 exists for $\lim_{\alpha \to 1} D_\alpha^\#(\rho||\sigma) = \hat{D}(\rho||\sigma)$ [33]. 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

$$\hat{D}(\rho||\sigma) \leq f(\epsilon, d) \tag{115}$$

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. 114 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. [6] *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

$$\hat{D}(\rho||\sigma) = \mathrm{tr}\left(\rho \log\left(\rho^{\frac{1}{2}}\sigma^{-1}\rho^{\frac{1}{2}}\right)\right)$$
$$= \mathrm{tr}\left(|\rho\rangle\langle\rho| \log\left(|\rho\rangle\langle\rho|\sigma^{-1}|\rho\rangle\langle\rho|\right)\right)$$
$$= \mathrm{tr}\left(|\rho\rangle\langle\rho| \log(\langle\rho|\sigma^{-1}|\rho\rangle)|\rho\rangle\langle\rho|\right)$$
$$= \log\langle\rho|\sigma^{-1}|\rho\rangle.$$

Next, we have that

$$D_{\max}(\rho||\sigma) = \log\left\|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right\|_\infty$$
$$= \log\left\|\sigma^{-\frac{1}{2}}|\rho\rangle\langle\rho|\sigma^{-\frac{1}{2}}\right\|_\infty$$
$$= \log\mathrm{tr}\left(\sigma^{-\frac{1}{2}}|\rho\rangle\langle\rho|\sigma^{-\frac{1}{2}}\right)$$
$$= \log\langle\rho|\sigma^{-1}|\rho\rangle.$$

$\qquad\square$

---

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

To obtain a contradiction, let $\epsilon \in [0, a^2)$. Define the states

$$\rho := |0\rangle\langle0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\sigma'_\epsilon := (\sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|1\rangle)(\sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|1\rangle)^\dagger$$

$$= \begin{pmatrix} 1-\epsilon & \sqrt{\epsilon(1-\epsilon)} \\ \sqrt{\epsilon(1-\epsilon)} & \epsilon \end{pmatrix}$$

$$\sigma_\epsilon := (1-\delta)\sigma'_\epsilon + \delta\rho$$

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

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 $\frac{1}{2}\|\rho - \sigma_\epsilon\|_1 \le \sqrt{\epsilon} \in [0, a)$. For these definitions, we have

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

which implies that $\hat{D}(\rho||\sigma_\epsilon) = \log\frac{1}{\delta}$ using Lemma 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. 115 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. 115 (also see [36]). Consequently, any kind of bound on $\hat{D}_\alpha$ or $D_\alpha^\#$ which results in a bound of the form in Eq. 115 as $\alpha \to 1$, for example, the bound in Eq. 113, 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 [15, Remark 3.3](otherwise the above counterexample would also give a no-go argument for $D_\alpha^\#$).

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

We have to redo the Lemmas used in [8] using $\tilde{H}_\alpha^\uparrow$ because we were only able to prove the dimension bound we need ($\tilde{H}_\alpha^\uparrow(A|BC) \ge \tilde{H}_\alpha^\uparrow(A|B) - 2\log|C|$) in terms of $\tilde{H}_\alpha^\uparrow$

Lemma C.1 [8, Lemma 3.1]. *For $\rho_{A_1A_2B}$ and σ_B be states and $\alpha \in (0, \infty)$, we have the chain rule*

$$\tilde{D}_\alpha(\rho_{A_1B}||\mathbb{1}_{A_1}\otimes\sigma_B) - \tilde{D}_\alpha(\rho_{A_1A_2B}||\mathbb{1}_{A_1A_2}\otimes\sigma_B) = \tilde{H}_\alpha^\downarrow(A_2|A_1B)_\nu \qquad (116)$$

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

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

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

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

Corollary C.2. (Chain rule for $\tilde{H}_\alpha^\downarrow$ [8, Theorem 3.2]) *For $\alpha \in (0, \infty)$, a state $\rho_{A_1 A_2 B}$, we have the chain rule*

$$\tilde{H}_\alpha^\downarrow(A_1 A_2 | B)_\rho = \tilde{H}_\alpha^\downarrow(A_1 | B)_\rho + \tilde{H}_\alpha^\downarrow(A_2 | A_1 B)_\nu \tag{117}$$

where the state $\nu_{A_1 A_2 B}$ is defined as

$$\nu_{A_1 B} := \frac{\left(\rho_{A_1 B}^{\frac{1}{2}} \rho_B^{-\alpha'} \rho_{A_1 B}^{\frac{1}{2}}\right)^\alpha}{\mathrm{tr}\left(\rho_{A_1 B}^{\frac{1}{2}} \rho_B^{-\alpha'} \rho_{A_1 B}^{\frac{1}{2}}\right)^\alpha}$$

$$\nu_{A_1 A_2 B} := \nu_{A_1 B}^{\frac{1}{2}} \rho_{A_2 | A_1 B} \nu_{A_1 B}^{\frac{1}{2}}$$

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

We can modify [8, Theorem 3.2], which is in terms of $\tilde{H}_\alpha^\downarrow$, to the following, which is a chain rule in terms of $\tilde{H}_\alpha^\uparrow$. The chain rule in this Corollary was also observed in [8].

Corollary C.3. (Chain rule for $\tilde{H}_\alpha^\uparrow$) *For $\alpha \in (0, \infty)$, a state $\rho_{A_1 A_2 B}$ and for any state σ_B such that $\tilde{H}_\alpha^\uparrow(A_1 | B)_\rho = -\tilde{D}_\alpha(\rho_{A_1 B} || \mathbb{1}_{A_1} \otimes \sigma_B)$, we have*

$$\tilde{H}_\alpha^\uparrow(A_1 A_2 | B)_\rho \geq \tilde{H}_\alpha^\uparrow(A_1 | B)_\rho + \tilde{H}_\alpha^\downarrow(A_2 | A_1 B)_\nu \tag{118}$$

where the state $\nu_{A_1 A_2 B}$ is defined as

$$\nu_{A_1 B} := \frac{\left(\rho_{A_1 B}^{\frac{1}{2}} \sigma_B^{-\alpha'} \rho_{A_1 B}^{\frac{1}{2}}\right)^\alpha}{\mathrm{tr}\left(\rho_{A_1 B}^{\frac{1}{2}} \sigma_B^{-\alpha'} \rho_{A_1 B}^{\frac{1}{2}}\right)^\alpha}$$

$$\nu_{A_1 A_2 B} := \nu_{A_1 B}^{\frac{1}{2}} \rho_{A_2 | A_1 B} \nu_{A_1 B}^{\frac{1}{2}}$$

and $\alpha' := \frac{\alpha - 1}{\alpha}$. For $\alpha \in (0, \infty)$, state $\rho_{A_1 A_2 B}$ and any state σ_B such that $\tilde{H}_\alpha^\uparrow(A_1 A_2 | B)_\rho = -\tilde{D}_\alpha(\rho_{A_1 A_2 B} || \mathbb{1}_{A_1 A_2} \otimes \sigma_B)$, we have

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

where the state $\nu_{A_1 A_2 B}$ is defined the same as above.

Proof. Let σ_B be a state such that $\tilde{H}_\alpha^\uparrow(A_1 | B)_\rho = -\tilde{D}_\alpha(\rho_{A_1 B} || \mathbb{1} \otimes \sigma_B)$. Then, using Lemma C.1, we have

$$\begin{aligned}
\tilde{H}_\alpha^\uparrow(A_1 A_2 | B)_\rho &\geq -\tilde{D}_\alpha(\rho_{A_1 A_2 B} || \mathbb{1}_{A_1 A_2} \otimes \sigma_B) \\
&= -\tilde{D}_\alpha(\rho_{A_1 B} || \mathbb{1}_{A_1} \otimes \sigma_B) + \tilde{H}_\alpha^\downarrow(A_2 | A_1 B)_\nu \\
&= \tilde{H}_\alpha^\uparrow(A_1 | B)_\rho + \tilde{H}_\alpha^\downarrow(A_2 | A_1 B)_\nu
\end{aligned}$$

for $\nu_{A_1 A_2 B}$ defined as in the Lemma. Similarly, if $\tilde{H}_\alpha^\uparrow(A_1 A_2 | B)_\rho = -\tilde{D}_\alpha(\rho_{A_1 A_2 B} || \mathbb{1}_{A_1 A_2} \otimes \sigma_B)$, then

$$\tilde{H}_\alpha^\uparrow(A_1 A_2 | B)_\rho = -\tilde{D}_\alpha(\rho_{A_1 A_2 B} || \mathbb{1}_{A_1 A_2} \otimes \sigma_B)$$

$$= -\tilde{D}_\alpha(\rho_{A_1 B} || \mathbb{1}_{A_1} \otimes \sigma_B) + \tilde{H}_\alpha^\downarrow(A_2|A_1 B)_\nu$$
$$\leq \tilde{H}_\alpha^\uparrow(A_1|B)_\rho + \tilde{H}_\alpha^\downarrow(A_2|A_1 B)_\nu$$

for $\nu_{A_1 A_2 B}$ defined as in the Lemma. $\qquad\square$

We transform [8, Theorem 3.3] to a statement about $\tilde{H}_\alpha^\uparrow$ in the following.

Lemma C.4. *Let $\alpha \in \left[\frac{1}{2}, \infty\right)$ and $\rho_{A_1 A_2 B_1 B_2}$ be a state which satisfies the Markov chain $A_1 \leftrightarrow B_1 \leftrightarrow B_2$. Then, we have*

$$\tilde{H}_\alpha^\uparrow(A_1 A_2|B_1 B_2)_\rho \geq \tilde{H}_\alpha^\uparrow(A_1|B_1)_\rho + \inf_\nu \tilde{H}_\alpha^\downarrow(A_2|A_1 B_1 B_2)_\nu \qquad (120)$$

where the infimum is taken over all states $\nu_{A_1 A_2 B_1 B_2}$ such that $\nu_{A_2 B_2|A_1 B_1} = \rho_{A_2 B_2|A_1 B_1}$.

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 [37, Theorem 5.4]

$$B_1 = \bigoplus_{j \in J} a_j \otimes c_j$$

such that

$$\rho_{A_1 B_1 B_2} = \bigoplus_{j \in J} p(j)\rho_{A_1 a_j} \otimes \rho_{c_j B_2}. \qquad (121)$$

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 $\mathcal{R}_{B_1 \to B_1 B_2}$ for $\rho_{A_1 B_1 B_2}$ as

$$\mathcal{R}_{B_1 \to B_1 B_2}(X) := \bigoplus_{j \in J} \text{tr}_{c_j}\left(\Pi_{a_j} \otimes \Pi_{c_j} X \Pi_{a_j} \otimes \Pi_{c_j}\right) \otimes \rho_{c_j B_2} \qquad (122)$$

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

$$\mathcal{R}_{B_1 \to B_1 B_2}(\rho_{A_1 B_1}) = \rho_{A_1 B_1 B_2}. \qquad (123)$$

We can now show that the optimisation for the conditional entropy $\tilde{H}_\alpha^\uparrow(A_1|B_1 B_2)_\rho$ can be restricted to states of the form $\mathcal{R}_{B_1 \to B_1 B_2}(\sigma_{B_1})$. This follows as

$$\tilde{H}_\alpha^\uparrow(A_1|B_1 B_2)_\rho = \sup_{\sigma_{B_1 B_2}} -\tilde{D}_\alpha(\rho_{A_1 B_1 B_2} || \mathbb{1}_{A_1} \otimes \sigma_{B_1 B_2})$$

$$\leq \sup_{\sigma_{B_1 B_2}} -\tilde{D}_\alpha(\mathcal{R}_{B_1 \to B_1 B_2} \circ \text{tr}_{B_2}(\rho_{A_1 B_1 B_2}) || \mathcal{R}_{B_1 \to B_1 B_2} \circ \text{tr}_{B_2}(\mathbb{1}_{A_1} \otimes \sigma_{B_1 B_2}))$$

$$= \sup_{\sigma_{B_1}} -\tilde{D}_\alpha(\rho_{A_1 B_1 B_2} || \mathbb{1}_{A_1} \otimes \mathcal{R}_{B_1 \to B_1 B_2}(\sigma_{B_1}))$$

$$\leq \sup_{\sigma_{B_1 B_2}} -\tilde{D}_\alpha(\rho_{A_1 B_1 B_2} || \mathbb{1}_{A_1} \otimes \sigma_{B_1 B_2})$$

$$= \tilde{H}_\alpha^\uparrow(A_1|B_1 B_2)_\rho$$

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 $B_1 B_2$, and the last line simply follows from the definition of $\tilde{H}_\alpha^\uparrow (A_1|B_1 B_2)_\rho$. As a result, it follows that

$$\tilde{H}_\alpha^\uparrow (A_1|B_1 B_2)_\rho = \sup_{\sigma_{B_1}} -\tilde{D}_\alpha(\rho_{A_1 B_1 B_2} \| \mathbb{1}_{A_1} \otimes \mathcal{R}_{B_1 \to B_1 B_2}(\sigma_{B_1})) \tag{124}$$

Let $\sigma_{B_1 B_2} = \mathcal{R}_{B_1 \to B_1 B_2}(\eta_{B_1})$ be such that $\tilde{H}_\alpha^\uparrow (A_1|B_1 B_2)_\rho = -\tilde{D}_\alpha(\rho_{A_1 B_1 B_2} \| \mathbb{1}_{A_1} \otimes \sigma_{B_1 B_2})$. Using Corollary C.3, for this choice of $\sigma_{B_1 B_2}$, we have that

$$\tilde{H}_\alpha^\uparrow (A_1 A_2|B_1 B_2)_\rho \geq \tilde{H}_\alpha^\uparrow (A_1|B_1 B_2)_\rho + \tilde{H}_\alpha^\downarrow (A_2|A_1 B_1 B_2)_\nu \tag{125}$$

where the state $\nu_{A_1 A_2 B_1 B_2}$ is defined as

$$\nu_{A_1 B_1 B_2} := \frac{\left(\rho_{A_1 B_1 B_2}^{\frac{1}{2}} \sigma_{B_1 B_2}^{-\alpha'} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} \right)^\alpha}{\operatorname{tr}\left(\rho_{A_1 B_1 B_2}^{\frac{1}{2}} \sigma_{B_1 B_2}^{-\alpha'} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} \right)^\alpha}$$

$$\nu_{A_1 A_2 B_1 B_2} := \nu_{A_1 B_1 B_2}^{\frac{1}{2}} \rho_{A_2|A_1 B_1 B_2} \nu_{A_1 B_1 B_2}^{\frac{1}{2}}.$$

We will now show that $\nu_{A_2 B_2|A_1 B_1} = \rho_{A_2 B_2|A_1 B_1}$. For this it is sufficient to show that

$$\nu_{A_1 B_1}^{-\frac{1}{2}} \nu_{A_1 B_1 B_2}^{\frac{1}{2}} = \rho_{A_1 B_1}^{-\frac{1}{2}} \rho_{A_1 B_1 B_2}^{\frac{1}{2}}.$$

We have that

$$\sigma_{B_1 B_2} = \mathcal{R}_{B_1 \to B_1 B_2}(\eta_{B_1})$$
$$= \bigoplus_{j \in J} \operatorname{tr}_{c_j}\left(\Pi_{a_j} \otimes \Pi_{c_j} \eta_{B_1} \Pi_{a_j} \otimes \Pi_{c_j} \right) \otimes \rho_{c_j B_2}$$
$$= \bigoplus_{j \in J} q(j) \omega_{a_j} \otimes \rho_{c_j B_2}$$

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

Since $\tilde{D}_\alpha(\rho_{A_1 B_1 B_2} \| \mathbb{1}_{A_1} \otimes \sigma_{B_1 B_2}) = -\tilde{H}_\alpha^\uparrow (A_1|B_1 B_2)_\rho \leq \log|A_1| < \infty$, we have that

$$\rho_{A_1 B_1 B_2} \ll \mathbb{1}_{A_1} \otimes \sigma_{B_1 B_2}$$
$$\Rightarrow \bigoplus_{j \in J'} p(j) \rho_{A_1 a_j} \otimes \rho_{c_j B_2} \ll \mathbb{1}_{A_1} \otimes \bigoplus_{j \in J} q(j) \omega_{a_j} \otimes \rho_{c_j B_2}$$
$$\Rightarrow \text{ for every } j \in J' : \rho_{A_1 a_j} \ll \mathbb{1}_{A_1} \otimes \omega_{a_j} \text{ and } q(j) > 0. \tag{126}$$

This decomposition can be used to evaluate $v_{A_1 B_1 B_2}$ as follows

$$
v_{A_1 B_1 B_2} = \frac{1}{N} \left(\rho_{A_1 B_1 B_2}^{\frac{1}{2}} \sigma_{B_1 B_2}^{-\alpha'} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} \right)^{\alpha}
$$

$$
= \frac{1}{N} \left(\bigoplus_{j \in J'} p(j)^{\frac{1}{2}} \rho_{A_1 a_j}^{\frac{1}{2}} \otimes \rho_{c_j B_2}^{\frac{1}{2}} \bigoplus_{j \in J} q(j)^{-\alpha'} \omega_{a_j}^{-\alpha'} \|! \otimes \rho_{c_j B_2}^{-\alpha'} \bigoplus_{j \in J'} p(j)^{\frac{1}{2}} \rho_{A_1 a_j}^{\frac{1}{2}} \otimes \rho_{c_j B_2}^{\frac{1}{2}} \right)^{\alpha}
$$

$$
= \frac{1}{N} \left(\bigoplus_{j \in J'} p(j) q(j)^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \otimes \rho_{c_j B_2}^{1-\alpha'} \right)^{\alpha}
$$

$$
= \frac{1}{N} \bigoplus_{j \in J'} p(j)^{\alpha} q(j)^{1-\alpha} \left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right)^{\alpha} \otimes \rho_{c_j B_2}
$$

for $N := \mathrm{tr}\left(\rho_{A_1 B_1 B_2}^{\frac{1}{2}} \sigma_{B_1 B_2}^{-\alpha'} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} \right)^{\alpha}$. Further, we have

$$
v_{A_1 B_1}^{-\frac{1}{2}} v_{A_1 B_1 B_2}^{\frac{1}{2}}
$$

$$
= \frac{1}{N^{-\frac{1}{2}}} \bigoplus_{j \in J'} p(j)^{-\frac{\alpha}{2}} q(j)^{-\frac{1-\alpha}{2}} \left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right)^{-\frac{\alpha}{2}} \otimes \rho_{c_j}^{-\frac{1}{2}}
$$

$$
\cdot \frac{1}{N^{\frac{1}{2}}} \bigoplus_{j \in J'} p(j)^{\frac{\alpha}{2}} q(j)^{\frac{1-\alpha}{2}} \left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right)^{\frac{\alpha}{2}} \otimes \rho_{c_j B_2}^{\frac{1}{2}}
$$

$$
= \bigoplus_{j \in J'} \left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right)^{0} \otimes \rho_{c_j}^{-\frac{1}{2}} \rho_{c_j B_2}^{\frac{1}{2}}
$$

$$
= \bigoplus_{j \in J'} \rho_{A_1 a_j}^{0} \otimes \rho_{c_j}^{-\frac{1}{2}} \rho_{c_j B_2}^{\frac{1}{2}}
$$

where in the last line we have used that the projector $\left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right)^{0}$ is equal to the projector $\rho_{A_1 a_j}^{0}$ 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

$$
\mathrm{im}\left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right) \subseteq \mathrm{im}\left(\rho_{A_1 a_j} \right). \tag{127}
$$

Second, we have that Eq. 126 above implies that $\omega_{a_j}^{0} \rho_{A a_j}^{0} = \rho_{A a_j}^{0}$ for every $j \in J'$. Now, for $j \in J'$ we have the following inequality

$$
\left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right) \geq m \left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{0} \rho_{A_1 a_j}^{\frac{1}{2}} \right)
$$

$$
= m \rho_{A_1 a_j}
$$

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)

$$\left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right)^0 \geq \rho_{A_1 a_j}^0. \tag{128}$$

Equations 127 and 128 together imply that for $j \in J'$

$$\left(\rho_{A_1 a_j}^{\frac{1}{2}} \omega_{a_j}^{-\alpha'} \rho_{A_1 a_j}^{\frac{1}{2}} \right)^0 = \rho_{A_1 a_j}^0.$$

Finally, we have that

$$
\rho_{A_1 B_1}^{-\frac{1}{2}} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} = \bigoplus_{j \in J'} p(j)^{-\frac{1}{2}} \rho_{A_1 a_j}^{-\frac{1}{2}} \otimes \rho_{c_j}^{-\frac{1}{2}} \bigoplus_{j \in J'} p(j)^{\frac{1}{2}} \rho_{A_1 a_j}^{\frac{1}{2}} \otimes \rho_{c_j B_2}^{\frac{1}{2}}
$$
$$
= \bigoplus_{j \in J'} \rho_{A_1 a_j}^0 \otimes \rho_{c_j}^{-\frac{1}{2}} \rho_{c_j B_2}^{\frac{1}{2}}.
$$

This proves that

$$\nu_{A_1 B_1}^{-\frac{1}{2}} \nu_{A_1 B_1 B_2}^{\frac{1}{2}} = \rho_{A_1 B_1}^{-\frac{1}{2}} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} \tag{129}$$

and hence

$$
\nu_{A_2 B_2 | A_1 B_1} = \nu_{A_1 B_1}^{-\frac{1}{2}} \nu_{A_1 B_1 B_2}^{\frac{1}{2}} \nu_{A_2 | A_1 B_1 B_2} \nu_{A_1 B_1 B_2}^{\frac{1}{2}} \nu_{A_1 B_1}^{-\frac{1}{2}}
$$
$$
= \rho_{A_1 B_1}^{-\frac{1}{2}} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} \rho_{A_2 | A_1 B_1 B_2} \rho_{A_1 B_1 B_2}^{\frac{1}{2}} \rho_{A_1 B_1}^{-\frac{1}{2}}
$$
$$
= \rho_{A_2 B_2 | A_1 B_1}
$$

where we have used the fact that $\nu_{A_2 | A_1 B_1 B_2} = \rho_{A_2 | A_1 B_1 B_2}$ and Eq. 129. We can now modify Eq. 125 to get

$$\tilde{H}_\alpha^\uparrow (A_1 A_2 | B_1 B_2)_\rho \geq \tilde{H}_\alpha^\uparrow (A_1 | B_1 B_2)_\rho + \inf_\nu \tilde{H}_\alpha^\downarrow (A_2 | A_1 B_1 B_2)_\nu$$

where the infimum is over states ν such that $\nu_{A_2 B_2 | A_1 B_1} = \rho_{A_2 B_2 | A_1 B_1}$. We can use the data processing inequality to get

$$\tilde{H}_\alpha^\uparrow (A_1 | B_1 B_2)_\rho = \tilde{H}_\alpha^\uparrow (A_1 | B_1 B_2)_{\mathcal{R}_{B_1 \to B_1 B_2}(\rho_{AB_1})}$$
$$\geq \tilde{H}_\alpha^\uparrow (A_1 | B_1)_\rho.$$

Together with the above inequality this proves the Lemma.   □

We will use the following modification of [8, Corollary 3.5].

Corollary C.5. *Let $\mathcal{M}_{R \to A_2 B_2}$ be a channel and $\rho_{A_1 A_2 B_1 B_2} = \mathcal{M}(\rho'_{A_1 B_1 R})$ such that the Markov chain $A_1 \leftrightarrow B_1 \leftrightarrow B_2$ holds. Then, we have*

$$\tilde{H}_\alpha^\uparrow (A_1 A_2 | B_1 B_2)_\rho \geq \tilde{H}_\alpha^\uparrow (A_1 | B_1)_\rho + \inf_\omega \tilde{H}_\alpha^\downarrow (A_2 | A_1 B_1 B_2)_{\mathcal{M}(\omega)} \tag{130}$$

where the infimum is taken over all states $\omega_{A_1 B_1 R}$. Moreover, if $\rho'_{A_1 B_1 R}$ is pure then we can restrict the optimisation to pure states.

Proof. The proof is the same as [8, Corollary 3.5]. We include it here for the sake of completeness.

It is sufficient to show that for every state ν such that $\nu_{A_2 B_2 | A_1 B_1} = \rho_{A_2 B_2 | A_1 B_1}$, there exists an $\omega_{A_1 B_1 R}$ such that $\nu_{A_1 A_2 B_1 B_2} = \mathcal{M}(\omega)$. For such a ν, we can define

$$\omega_{R A_1 B_1} = \nu_{A_1 B_1}^{\frac{1}{2}} \rho_{A_1 B_1}^{-\frac{1}{2}} \rho'_{A_1 B_1 R} \rho_{A_1 B_1}^{-\frac{1}{2}} \nu_{A_1 B_1}^{\frac{1}{2}}$$

which can be seen to be a valid state and also satisfy $\nu_{A_1 A_2 B_1 B_2} = \mathcal{M}(\omega)$. □

Appendix D: Dimension Bounds for Conditional Rényi Entropies

Lemma D.1 (Dimension bound). *For $\alpha \in [\frac{1}{2}, \infty]$, a state $\rho_{A_1 A_2 B}$, the following bounds hold for the sandwiched conditional entropies*

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

For $\alpha \in [0, 2]$ and a state $\rho_{A_1 A_2 B}$, the following bounds hold for the Petz conditional entropies

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

Proof. For the sandwiched conditional entropies, we simply use the corresponding chain rules (Corollary C.2 or Corollary C.3) along with the fact that for all states ν, $\tilde{H}_\alpha^\downarrow(A_2|A_1 B)_\nu \in [-\log|A_2|, \log|A_2|]$ [17, Lemma 5.2].

For the Petz conditional entropies, we will make use of the Jensen's inequality for operators [38, 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 $\alpha \in [0, 1]$

$$\operatorname{tr}_X P_{XY}^\alpha = \sum_{i=1}^{|X|} \mathbb{1}_Y \otimes \langle e_i|_X P_{XY}^\alpha \mathbb{1}_Y \otimes |e_i\rangle_X$$
$$\leq |X| \left(\sum_{i=1}^{|X|} \frac{1}{|X|} \mathbb{1}_Y \otimes \langle e_i|_X P_{XY} \mathbb{1}_Y \otimes |e_i\rangle_X \right)^\alpha$$
$$= |X|^{1-\alpha} P_Y^\alpha \tag{131}$$

where in the second step we have used the operator Jensen's inequality with the operators $\left\{ \frac{1}{\sqrt{|X|}} \mathbb{1}_Y \otimes |e_i\rangle_X \right\}_{i=1}^{|X|}$ along with the fact that the map $X \mapsto X^\alpha$ is operator concave. For $\alpha \in [1, 2]$ and positive operator P_{XY}, we can use the same argument as above and the fact that $X \mapsto X^\alpha$ is operator convex in this regime and derive

$$\operatorname{tr}_X P_{XY}^\alpha \geq |X|^{1-\alpha} P_Y^\alpha. \tag{132}$$

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

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

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

The following Lemma was originally proven in [39, Proposition 8]. We reproduce the proof argument here.

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

$$
\tilde{H}_\alpha^\uparrow(A|BC)_\rho \geq \tilde{H}_\alpha^\uparrow(AC|B)_\rho - \log |C| \tag{133}
$$

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

$$
\bar{H}_\alpha^\uparrow(A|BC)_\rho \geq \bar{H}_\alpha^\uparrow(AC|B)_\rho - \log |C| \tag{134}
$$

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

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

where we simply restrict the supremum in the second line to states of the form $\eta_{BC} = \eta_B \otimes \frac{\mathbb{1}_C}{|C|}$ to derive the inequality. The same proof also works with $\bar{H}_\alpha^\uparrow$ entropy. □

The following lemma was originally proven in [40, 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*

$$
\tilde{H}_\alpha^\uparrow(A|BC)_\rho \geq \tilde{H}_\alpha^\uparrow(A|B)_\rho - 2\log |C|. \tag{135}
$$

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

$$
\tilde{H}_\alpha^\uparrow(A|BC)_\rho \geq \tilde{H}_\alpha^\uparrow(A|B)_\rho - \log |C|. \tag{136}
$$

Proof. This bound can be proven by combining Lemmas D.1 and D.2. In the case that C is classical, we have the inequality $\tilde{H}_\alpha^\uparrow(AC|B)_\rho \geq \tilde{H}_\alpha^\uparrow(A|B)_\rho$ [17, Lemma 5.3]. □

Appendix E: Necessity for Constraints on Side Information Size for Approximate AEP and EAT and its Implication for Approximate GEAT

It turns out that it is necessary to place some sort of bound on the size of the side information for an approximate entropy accumulation theorem of the form in Theorem 5.1. The following classical example demonstrates this. This example also demonstrates the necessity for a bound on the size of the side information in an approximate asymptotic equipartition of the form in Theorem 4.5.

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*

$$\left\| \Phi(\rho_{RR'}) - \Phi'(\rho_{RR'}) \right\|_1 \leq \left\| P_{AC}^\Phi - P_{AC}^{\Phi'} \right\|_1 \tag{137}$$

where P_{AC}^Φ and $P_{AC}^{\Phi'}$ 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

$$\rho_{CR'} = \sum_c |c\rangle\langle c|_C \otimes \mathrm{tr}_R \left(|c\rangle\langle c|_R \rho_{RR'} \right)$$

$$= \sum_c p(c)|c\rangle\langle c|_C \otimes \rho_{R'|c}$$

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 $p_{A|C}^\Phi$ and the action of channel Φ' on register C can be similarly represented using $p_{A|C}^{\Phi'}$. We can define the states

$$\rho_{ACR'}^\Phi := \sum_{ac} p_{A|C}^\Phi(a|c)p(c)|a, c\rangle\langle a, c| \otimes \rho_{R'|c}$$

$$\rho_{ACR'}^{\Phi'} := \sum_{ac} p_{A|C}^{\Phi'}(a|c)p(c)|a, c\rangle\langle a, c| \otimes \rho_{R'|c}.$$

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

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

$\square$

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

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

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 $\left\|\mathcal{M}_k - \mathcal{M}'_k\right\|_{\diamond} \le \epsilon$.

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 $H^{\epsilon'}_{\min}(A_1^n|B_1^n C_1^n)_R = O(1)$ as long as $\epsilon' \le \frac{1}{4}$. 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 $p(E) \ge 1 - \epsilon'$.

Lemma E.2. *Let P_{AB} be a subnormalised probability distribution such that $A = f(B)$ for some function f (that is, $P(a, b) > 0$ only if $a = f(b)$). Then, $H^{\epsilon}_{\min}(A|B)_P \le \log \frac{1}{\mathrm{tr}(P) - \sqrt{2\epsilon}}$.*

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

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

which implies that $H_{\min}(A|B)_{P'} \leq \log \frac{1}{\text{tr}(P) - \sqrt{2\epsilon}}$. Since, this is true for every distribution ϵ-close to P, it also holds for $H^{\epsilon}_{\min}(A|B)_P$. $\square$

We then have that

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

where in the first line we have used [41, 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 theorem (GEAT) [9] 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.

Appendix 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*

$$\left\| p_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} - q^{(k)}_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} \right\|_\infty \leq \epsilon \tag{138}$$

where $\|v\|_\infty := \max_i |v(i)|$ and the $q^{(k)}_{B_k | a_1^{k-1}, b_1^{k-1}, e} = q^{(k)}_{B_k | b_1^{k-1}, e}$ or equivalently $q^{(k)}$ satisfies the Markov chain $A_k \leftrightarrow B_1^{k-1} E \leftrightarrow B_k$. Also, let $|A_k| = |A|, |B_k| = |B|$ for every $k \in [n]$.

Then, for $\epsilon' \in (0, 1)$ and $\alpha \in \left(1, 1 + \frac{1}{\log(1+2|A|)} \right)$, we have that

$$H_{\min}^{\epsilon'}(A_1^n | B_1^n E)_p \geq \sum_{k=1}^n \inf_q H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}}$$

$$- n(\alpha - 1) \log^2(2|A| + 1) - \frac{\alpha}{\alpha - 1} n \log(1 + \epsilon|A||B|) - \frac{g_0(\epsilon')}{\alpha - 1}. \tag{139}$$

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

$$H_{\min}^{\epsilon'}(A_1^n | B_1^n E)_p \geq \sum_{k=1}^n \inf_q H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}}$$

$$- n\sqrt{\epsilon} \left(\log^2(2|A| + 1) - 2|A||B| \right) - \frac{g_0(\epsilon')}{\alpha - 1} \tag{140}$$

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

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

$$\Rightarrow p_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} \leq q^{(k)}_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} + \epsilon \, \mathbb{1}_{A_k B_k}$$

$$\Rightarrow p_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} \leq q^{(k)}_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} + \epsilon |A||B| u_{A_k B_k}$$

$$\Rightarrow p_{A_k B_k | a_1^{k-1}, b_1^{k-1}, e} \leq (1 + |A||B|\epsilon) r^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E}$$

Define the distribution

$$r_{A_1^n B_1^n E} = \prod_{k=1}^{n} r^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} \, p_E. \tag{141}$$

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

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

$$= \frac{1}{1+|A||B|\epsilon} q^{(k)}_{B_k | B_1^{k-1} E}(b_k | b_1^{k-1} e) + \frac{\epsilon}{1+|A||B|\epsilon},$$

which implies

$$r_{B_k | B_1^{k-1} E}(b_k | b_1^{k-1} e) = \sum_{\bar{a}_1^{k-1}} r_{A_1^{k-1} | B_1^{k-1} E}(\bar{a}_1^{k-1} | b_1^{k-1} e) r_{B_k | A_1^{k-1} B_1^{k-1} E}(b_k | \bar{a}_1^{k-1} b_1^{k-1} e)$$

$$= \sum_{\bar{a}_1^{k-1}} r_{A_1^{k-1} | B_1^{k-1} E}(\bar{a}_1^{k-1} | b_1^{k-1} e) \left(\frac{1}{1+|A||B|\epsilon} q^{(k)}_{B_k | B_1^{k-1} E}(b_k | b_1^{k-1} e) + \frac{\epsilon}{1+|A||B|\epsilon} \right)$$

$$= r_{B_k | A_1^{k-1} B_1^{k-1} E}(b_k | a_1^{k-1} b_1^{k-1} e).$$

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

$$p_{A_1^n B_1^n E}(a_1^n, b_1^n, e) = \prod_{k=1}^{n} p_{A_k B_k | A_1^{k-1}, B_1^{k-1}, E}(a_k, b_k | a_1^{k-1}, b_1^{k-1}, e) p_E(e)$$

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

$$= (1 + \epsilon |A||B|)^n r_{A_1^n B_1^n E}(a_1^n, b_1^n, e)$$

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 [8, Corollary 3.5] along with [8, Lemma B.9] n-times to bound the entropy of this auxiliary distribution. We get

$$\tilde{H}_\alpha^\uparrow(A_1^n | B_1^n E)_r \geq \sum_{k=1}^{n} \inf_{q_{A_1^{k-1} B_1^{k-1} E}} \tilde{H}_\alpha^\downarrow(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{r^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}}$$

$$\geq \sum_{k=1}^{n} \inf_{q_{A_1^{k-1} B_1^{k-1} E}} H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{r^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}}$$

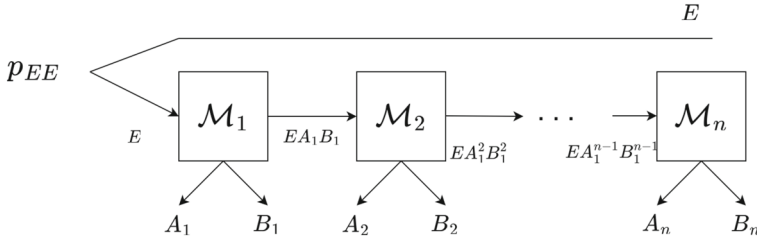**Fig. 3.** Setting for classical EAT

$$- n(\alpha - 1)\log^2(2|A| + 1)$$

$$\geq \sum_{k=1}^{n} \left(\inf_{q_{A_1^{k-1} B_1^{k-1} E}} \frac{1}{1 + |A||B|\epsilon} H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}} \right.$$

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

$$\geq \sum_{k=1}^{n} \inf_{q_{A_1^{k-1} B_1^{k-1} E}} H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}}$$

$$- n(\alpha - 1)\log^2(2|A| + 1)$$

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

$$H_{\min}^{\epsilon'}(A_1^n | B_1^n E)_p \geq \tilde{H}_{\alpha}^{\uparrow}(A_1^n | B_1^n E)_r - \frac{\alpha}{\alpha - 1} D_{\max}(p_{A_1^n B_1^n E} || r_{A_1^n B_1^n E}) - \frac{g_1(\epsilon', 0)}{\alpha - 1}$$

$$\geq \sum_{k=1}^{n} \inf_q H(A_k | B_k A_1^{k-1} B_1^{k-1} E)_{q^{(k)}_{A_k B_k | A_1^{k-1} B_1^{k-1} E} q_{A_1^{k-1} B_1^{k-1} E}}$$

$$- n(\alpha - 1)\log^2(2|A| + 1) - \frac{\alpha}{\alpha - 1} n \log(1 + \epsilon|A||B|) - \frac{g_0(\epsilon')}{\alpha - 1}.$$

$\square$

Appendix G: Lemma to Bound Distance after Conditioning

The following Lemma relates the distance of two states conditioned on an event to the distance between them without conditioning.

Lemma G.1. *Suppose* $\rho_{XA} = \sum_{x \in \mathcal{X}} p(x)|x\rangle\langle x| \otimes \rho_{A|x}$ *and* $\tilde{\rho}_{XA} = \sum_{x \in \mathcal{X}} \tilde{p}(x)|x\rangle\langle x| \otimes \tilde{\rho}_{A|x}$ *are classical-quantum states such that* $\frac{1}{2}\|\rho_{XA} - \tilde{\rho}_{XA}\|_1 \leq \epsilon$. *Then, for* $x \in \mathcal{X}$ *such that* $p(x) > 0$, *we have*

$$\frac{1}{2}\left\|\rho_{A|x} - \tilde{\rho}_{A|x}\right\|_1 \leq \frac{2\epsilon}{p(x)} \tag{142}$$

Proof.

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

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

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

and

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

Using these inequalities, we have

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

$\square$

References

1. Marwah, A., Dupuis, F.: Proving Security of bb84 Under Source Correlations (2024). arXiv:2402.12346 [quant-ph]
2. Tomamichel, M., Lim, C.C.W., Gisin, N., Renner, R.: Tight finite-key analysis for quantum cryptography. Nat. Commun. **3**(1), 634 (2012). https://doi.org/10.1038/ncomms1631
3. Konig, R., Wehner, S., Wullschleger, J.: Unconditional security from noisy quantum storage. IEEE Trans. Inf. Theory **58**(3), 1962–1984 (2012). https://doi.org/10.1109/TIT.2011.2177772
4. Renner, R.: Security of Quantum Key Distribution. Ph.D. Thesis (2006)
5. Renner, R., König, R.: Universally composable privacy amplification against quantum adversaries. In: Kilian, J. (ed.) Theory of Cryptography, pp. 407–425. Springer, Berlin (2005)
6. Tomamichel, M., Renner, R., Schaffner, C., Smith, A.: Leftover hashing against quantum side information. In: IEEE International Symposium on Information Theory, pp. 2703–2707 (2010). https://doi.org/10.1109/ISIT.2010.5513652
7. Vitanov, A., Dupuis, F., Tomamichel, M., Renner, R.: Chain rules for smooth min- and max-entropies. IEEE Trans. Inf. Theory **59**(5), 2603–2612 (2013). https://doi.org/10.1109/TIT.2013.2238656
8. Dupuis, F., Fawzi, O., Renner, R.: Entropy accumulation. Commun. Math. Phys. **379**(3), 867–913 (2020). https://doi.org/10.1007/s00220-020-03839-5
9. Metger, T., Fawzi, O., Sutter, D., Renner, R.: Generalised entropy accumulation (2022). https://doi.org/10.48550/ARXIV.2203.04989. arXiv:2203.04989
10. Jain, R., Radhakrishnan, J., Sen, P.: 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, pp. 429–438 (2002). https://doi.org/10.1109/SFCS.2002.1181967
11. Ekert, A.K.: Quantum cryptography based on Bell's theorem. Phys. Rev. Lett. **67**, 661–663 (1991). https://doi.org/10.1103/PhysRevLett.67.661
12. Arnon-Friedman, R., Dupuis, F., Fawzi, O., Renner, R., Vidick, T.: Practical device-independent quantum cryptography via entropy accumulation. Nat. Commun. **9**(1), 459 (2018). https://doi.org/10.1038/s41467-017-02307-4

13. Jain, R., Kundu, S.: A Direct Product Theorem for Quantum Communication Complexity with Applications to Device-Independent Cryptography (2023)
14. Tan, E.Y.-Z.: Robustness of Implemented Device-Independent Protocols Against Constrained Leakage (2023)
15. Fawzi, H., Fawzi, O.: Defining quantum divergences via convex optimization. Quantum **5**, 387 (2021). https://doi.org/10.22331/q-2021-01-26-387
16. Pereira, M., Currás-Lorenzo, G., Navarrete, A., Mizutani, A., Kato, G., Curty, M., Tamaki, K.: Modified BB84 Quantum Key Distribution Protocol Robust to Source Imperfections (2022). https://doi.org/10.48550/ARXIV.2210.11754. arXiv:2210.11754
17. Tomamichel, M.: Quantum Information Processing with Finite Resources. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-21891-5
18. Cooney, T., Mosonyi, M., Wilde, M.M.: Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication. Commun. Math. Phys. **344**(3), 797–829 (2016). https://doi.org/10.1007/s00220-016-2645-4
19. Leditzky, F., Kaur, E., Datta, N., Wilde, M.M.: Approaches for approximate additivity of the Holevo information of quantum channels. Phys. Rev. A (2018). https://doi.org/10.1103/physreva.97.012332
20. Christandl, M., Müller-Hermes, A.: Relative entropy bounds on quantum, private and repeater capacities. Commun. Math. Phys. **353**(2), 821–852 (2017)
21. Tomamichel, M., Colbeck, R., Renner, R.: A fully quantum asymptotic equipartition property. IEEE Trans. Inf. Theory **55**(12), 5840–5847 (2009). https://doi.org/10.1109/TIT.2009.2032797
22. Tomamichel, M.: A Framework for Non-asymptotic Quantum Information Theory. Ph.D. Thesis (2012). https://doi.org/10.48550/ARXIV.1203.2142. https://arxiv.org/abs/1203.2142
23. Tomamichel, M., Hayashi, M.: A hierarchy of information quantities for finite block length analysis of quantum tasks. IEEE Trans. Inf. Theory **59**(11), 7693–7710 (2013). https://doi.org/10.1109/TIT.2013.2276628
24. Jain, R., Nayak, A.: Short Proofs of the Quantum Substate Theorem (2011). https://doi.org/10.48550/ARXIV.1103.6067. arXiv:1103.6067
25. Watanabe, S.: Information theoretical analysis of multivariate correlation. IBM J. Res. Dev. **4**(1), 66–82 (1960). https://doi.org/10.1147/rd.41.0066
26. Horodecki, R.: Informationally coherent quantum systems. Phys. Lett. A **187**(2), 145–150 (1994). https://doi.org/10.1016/0375-9601(94)90052-3
27. Cerf, N.J., Massar, S., Schneider, S.: Multipartite classical and quantum secrecy monotones. Phys. Rev. A (2002). https://doi.org/10.1103/physreva.66.042309
28. Alicki, R., Fannes, M.: Continuity of quantum conditional information. J. Phys. A: Math. Gen. **37**(5), 55–57 (2004). https://doi.org/10.1088/0305-4470/37/5/l01
29. Winter, A.: Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. Commun. Math. Phys. **347**(1), 291–313 (2016). https://doi.org/10.1007/s00220-016-2609-8
30. Wilde, M.M.: Quantum Information Theory. Cambridge University Press, Cambridge (2013). https://doi.org/10.1017/CBO9781139525343
31. Arnon-Friedman, R.: Device-Independent Quantum Information Processing. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60231-4
32. Pironio, S., Acín, A., Brunner, N., Gisin, N., Massar, S., Scarani, V.: Device-independent quantum key distribution secure against collective attacks. New J. Phys. **11**(4), 045021 (2009). https://doi.org/10.1088/1367-2630/11/4/045021
33. Bergh, B., Salzmann, R., Datta, N.: The $\alpha \to 1$ limit of the sharp quantum rényi divergence. J. Math. Phys. **62**(9), 092205 (2021). https://doi.org/10.1063/5.0049791
34. Matsumoto, K.: A new quantum version of F-divergence. In: Ozawa, M., Butterfield, J., Halvorson, H., Rédei, M., Kitajima, Y., Buscemi, F. (eds.) Reality and measurement in algebraic quantum theory, pp. 229–273. Springer, Singapore (2018)
35. Dupuis, F., Fawzi, O.: Entropy accumulation with improved second-order term. IEEE Trans. Inf. Theory **65**(11), 7596–7612 (2019). https://doi.org/10.1109/TIT.2019.2929564
36. Bluhm, A., Capel, Gondolf, P., Pérez-Hernández, A.: Continuity of Quantum Entropic Quantities via Almost Convexity (2022)
37. Sutter, D.: Approximate quantum Markov chains. In: Approximate Quantum Markov Chains, pp. 75–100. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78732-9_5
38. Bhatia, R.: Matrix Analysis, vol. 169. Springer, Cham (1997)
39. Müller-Lennert, M., Dupuis, F., Szehr, O., Fehr, S., Tomamichel, M.: On quantum Rényi entropies: a new generalization and some properties. J. Math. Phys. **54**(12), 122203 (2013). https://doi.org/10.1063/1.4838856
40. Leditzky, F.: Relative Entropies and Their Use in Quantum Information Theory. Ph.D. Thesis (2016)

41. Tomamichel, M., Leverrier, A.: A largely self-contained and complete security proof for quantum key distribution. Quantum **1**, 14 (2017). https://doi.org/10.22331/q-2017-07-14-14