Entropy Accumulation

Frédéric Dupuis\textsuperscript{1,2,5}, Omar Fawzi\textsuperscript{3}, Renato Renner\textsuperscript{4}

\textsuperscript{1} Université de Lorraine, CNRS, Inria, LORIA, 54000 Nancy, France
\textsuperscript{2} Faculty of Informatics, Masaryk University, Brno, Czech Republic
\textsuperscript{3} Laboratoire de l’Informatique du Parallélisme, ENS de Lyon, Lyon, France.
E-mail: omar.fawzi@ens-lyon.fr
\textsuperscript{4} Institute for Theoretical Physics, ETH Zurich, Zurich, Switzerland
\textsuperscript{5} Département d’informatique et de recherche opérationnelle, Université de Montréal, Montreal, Quebec, Canada

Received: 18 December 2017 / Accepted: 23 June 2020
Published online: 23 September 2020 – © The Author(s) 2020

Abstract: We ask the question whether entropy accumulates, in the sense that the operationally relevant total uncertainty about an \(n\)-partite system \(A = (A_1, \ldots, A_n)\) corresponds to the sum of the entropies of its parts \(A_i\). The Asymptotic Equipartition Property implies that this is indeed the case to first order in \(n\)—under the assumption that the parts \(A_i\) are identical and independent of each other. Here we show that entropy accumulation occurs more generally, i.e., without an independence assumption, provided one quantifies the uncertainty about the individual systems \(A_i\) by the von Neumann entropy of suitably chosen conditional states. The analysis of a large system can hence be reduced to the study of its parts. This is relevant for applications. In device-independent cryptography, for instance, the approach yields essentially optimal security bounds valid for general attacks, as shown by Arnon-Friedman \textit{et al.} (SIAM J Comput 48(1):181–225, 2019).

1. Introduction

In classical information theory, the uncertainty one has about a variable \(A\) given access to side information \(B\) can be operationally quantified by the number of bits one would need to learn, in addition to \(B\), in order to reconstruct \(A\). While this number generally fluctuates, it is—except with probability of order \(\varepsilon > 0\)—not larger than the \(\varepsilon\)-smooth max-entropy, \(H_{\text{max}}^\varepsilon(A|B)_\rho\), evaluated for the joint distribution \(\rho\) of \(A\) and \(B\) \cite{45}.\textsuperscript{1} Conversely, it is in the same way not smaller than the \(\varepsilon\)-smooth min-entropy, \(H_{\text{min}}^\varepsilon(A|B)_\rho\). This may be summarised by saying that the number of bits needed to reconstruct \(A\) from \(B\) is with probability at least 1 —\(O(\varepsilon)\) contained in the interval

\[ I = \left[ H_{\text{min}}^\varepsilon(A|B)_\rho, H_{\text{max}}^\varepsilon(A|B)_\rho \right], \]  

\textsuperscript{1} There is some freedom in how to count the number of bits, but the statement always holds up to additive terms of the order \(\log(1/\varepsilon)\).
whose boundaries are defined by the smooth entropies. We refer to Definition 2.2 below for a precise definition of these quantities.

This approach to quantifying uncertainty can be extended to the case where $A$ and $B$ are quantum systems. The conclusion remains the same: the operationally relevant uncertainty interval is $I$ as defined by (1). The only difference is that $\rho$ is now a density operator, which describes the joint state of $A$ and $B$ [41, 44, 51].

Finding the boundaries of the interval $I$ is a central task of information theory. However, the smooth entropies of a large system $A$ are often difficult to calculate. It is therefore rather common to introduce certain assumptions to render this task more feasible. One extremely popular approach in standard information theory is to assume that the system consists of many mutually independent and identically distributed (IID) parts. More precisely, the IID Assumption demands that the system be of the form $A = A_1^n = A_1 \otimes \cdots \otimes A_n$, that the side information have an analogous structure $B = B_1^n = B_1 \otimes \cdots \otimes B_n$, and that the joint state of these systems be of the form $\rho_{A_1 B_1 \cdot \cdot \cdot A_n B_n} = \nu_{A_B} \otimes^n$, for some density operator $\nu_{A_B}$. A fundamental result from information theory, the Asymptotic Equipartition Property (AEP) [48] (see [54] for the quantum version), then asserts that the uncertainty interval satisfies

$$I \subset \left[ n \left( H(A|B)_v - \frac{c_v}{\sqrt{n}} \right), n \left( H(A|B)_v + \frac{c_v}{\sqrt{n}} \right) \right],$$

where $c_v$ is a constant (independent of $n$) and where $H(A|B)_v$ is the conditional von Neumann entropy evaluated for the state $\nu_{A_B}$. In other words, for large $n$, the operationally relevant total uncertainty one has about $A_1^n$ given $B_1^n$ is well approximated by $nH(A|B)_v = \sum_i H(A_i|B_i)_\rho$. In this sense, the entropy of the individual systems $A_i$ accumulates to the entropy of the total system $A_1^n$.

In this work, we generalise this statement to the case where the individual pairs $A_i, B_i$ are no longer independent of each other, i.e., where the IID assumption does not hold. Without loss of generality one may think of the pairs $A_i, B_i$ as being generated by a sequence of processes $\mathcal{M}_i$, as shown in Fig. 1. Each process $\mathcal{M}_i$ may pass information on to the next one via a “memory” register $R_i$. The state of the “future” pairs can thus depend on the “past” ones. The only assumption we make is that, given the side information $B_i^1$ generated until step $i$, the systems $A_i^1$ are independent of the next piece of side information $B_{i+1}^1$. This is captured by the requirement that $A_i^1 \leftrightarrow B_i^1 \leftrightarrow B_{i+1}$ forms a quantum Markov chain.

Entropy accumulation is then the claim that

$$I \subset \left[ \sum_{i=1}^n \left( \inf_{\omega_{R_{i-1}}^1} H(A_i|B_i R_i)_{\mathcal{M}_i(\omega)} - \frac{c_v}{\sqrt{n}} \right), \sum_{i=1}^n \left( \sup_{\omega_{R_{i-1}}^1} H(A_i|B_i R_i)_{\mathcal{M}_i(\omega)} + \frac{c_v}{\sqrt{n}} \right) \right],$$

where, in the $i$th term of each sum, the infimum or supremum ranges over joint states $\omega_{R_{i-1}}^1$ of the memory $R_{i-1}$ and a system $R$ isomorphic to it, and the conditional von Neumann entropy is evaluated for the state $(\mathcal{M}_i \otimes I_R)(\omega_{R_{i-1}}^1)$, abbreviated by $\mathcal{M}_i(\omega)$, which describes the output pair $A_i B_i$ generated by $\mathcal{M}_i$ jointly with $R$.

---

2 We note that the value of $c_v$ governing the second order term is well understood; see [57] and the book [52] for more details on this.

3 The IID assumption corresponds to the special case where the systems $R_i$ are trivial (ensuring the mutual independence of the pairs $A_i B_i$) and where the maps $\mathcal{M}_i$ are all the same (ensuring that the state of each pair $A_i B_i$ is identical to all others).

4 The necessity of this condition is discussed in “Appendix C”.
To illustrate (3) it is useful to think of a communication scenario with two parties, Alice and Bob, who are receiving information $A_1^n$ and $B_1^n$, respectively. Suppose that a source with memory $R_i$ generates this information sequentially in $n$ steps, described by maps $\mathcal{M}_i$ as depicted in Fig. 1. Suppose furthermore that Bob would like to infer all $n$ values $A_i$ (which, for the purpose of this example, we assume to be classical). As discussed above, for this he would require $N$ additional classical bits from Alice, where $N$ fluctuates (up to probability $\varepsilon$) within an interval $I$ with boundaries given by the entropies $H_{\min}^\varepsilon(A_1^n|B_1^n)$ and $H_{\max}^\varepsilon(A_1^n|B_1^n)$, which quantify Bob’s uncertainty about $A_1^n$. While these entropies depend on the joint state $\rho_{A_1^n B_1^n}$ of the entire information generated by the source over all $n$ steps, they can, according to (3), be lower (or upper) bounded by a sum of terms that merely depend on the individual steps $\mathcal{M}_i$. Specifically, the minimum (or maximum) number $N$ of bits that Alice needs to send to Bob so that he can infer her values $A_i$ grows for each such value by the von Neumann entropy $H(A_i|B_i R_i)$, minimised (or maximised) over all possible states the memory $R_{i-1}$ could have been in right before the pair $A_i B_i$ was produced, and conditioned on $B_i$ as well as any information $R$ about this memory.\footnote{In the special case where the source is IID, the memory register is trivial and no minimisation (or maximisation) is necessary. Expression (3) then reduces to (2), and one retrieves the (well known) result that the number of bits that Alice needs to communicate to Bob per value $A_i$ is, up to second order terms, given by $H(A_i|B_i)$.}

The main result we derive in this work is actually a bit more general than (3), allowing one to take into account global information about the statistics of $A_1^n$ and $B_1^n$. This is relevant for applications. In quantum key distribution, for instance, $\mathcal{M}_i$ models the generation of the $i$th bit of the raw key. However, in this cryptographic scenario, $\mathcal{M}_i$ can depend on the attack strategy of an adversary, and is thus partially unknown. Hence, in order to bound the entropy (which characterises an adversary’s uncertainty) of the raw key bits, one must as well take into account global statistical properties. These are inferred by tests carried out by the quantum key distribution protocol on a small sample of the generated bits. To incorporate such statistical information in the analysis, we consider for each $i$ an additional classical value $X_i$ derived from $A_i$ and $B_i$, as depicted by Fig. 2. Specifically, $X_i$ shall tell us whether position $i$ was included in the statistical test, and if so, the outcome of the test performed at step $i$. For this extended scenario, (3) still holds, but now the infimum and supremum are taken over a restricted set, containing only those states $\omega$ for which the resulting probability distribution on $X_i$ corresponds to the observed statistics.

Entropy accumulation has a number of theoretical and practical implications. For example, it serves as a technique to turn cryptographic security proofs that were restricted to collective attacks to security proofs against general attacks. This application is demonstrated in [5] for the case of a fully device-independent quantum key distribu-
Fig. 2. Circuit diagram illustrating the decomposition of states $\rho_{A_1^n B_1^n X_1^n}$ relevant for the full version of our main theorem, which can take into account statistical information $X_i$. The individual pieces $X_i$ of this statistical information are classical values that can be determined from $A_i$ and $B_i$ without disturbing them. When $A_i$ and $B_i$ are themselves classical, this means that $X_i$ is a deterministic function of $A_i$ and $B_i$. For a precise definition in the general case we refer to Sect. 4.

The resulting security bounds are essentially tight, implying that device-independent cryptography is possible with state-of-the-art technology. To illustrate the basic ideas behind such applications, we will present two concrete examples in more detail. The first is a proof of security of a variant of the E91 Quantum Key Distribution protocol. This new security proof has two advantages. First, its structure is modular and it may therefore be adapted to other cryptographic schemes (see also the discussion in Sect. 6). In addition, it achieves a strong level of security where no assumption is made on Bob’s devices. This is sometimes referred to as one-sided measurement device independence and this level of security was partially achieved in [58] (they used a memoryless devices assumption which we do not need) and later fully in [56] though with sub-optimal rates.

The second example is the derivation of an upper bound on the fidelity achievable by Fully Quantum Random Access Codes.

The proof of the main result, Eq. (3), has a similar structure as the proof of the Quantum Asymptotic Equipartition Property [54], which we can retrieve as a special case (see Corollary 4.10). The idea is to first bound the smooth entropy of the entire sequence $A_1^n$ conditioned on $B_1^n$ by a conditional Rényi entropy of order $\alpha$, then decompose this entropy into a sum of conditional Rényi entropies for the individual terms $A_i$, and finally bound these in terms of von Neumann entropies. However, in contrast to previous arguments, we use a recently introduced version of conditional Rényi entropies, termed “sandwiched Rényi entropies” [37,64]. For these entropies, we derive a novel chain rule, which forms the core technical part of our proof. In addition, some of the concepts used in this work generalise techniques proposed in the recent security proofs for device-independent cryptography presented in [34,35]. In particular, the dominant terms of the lower bound on the amount of randomness obtained in [35], called rate curves, are similar to the tradeoff functions considered here (cf. Definition 4.1).6

Paper organisation: We begin with preliminaries and notation in Sect. 2. Section 3 is devoted to the central technical ingredient of our argument, a chain rule for Rényi entropies. The main result, the theorem on entropy accumulation, is then stated and proved in Sect. 4. In Sect. 5 we present the two sample applications mentioned above, before concluding with remarks and suggestions for future work in Sect. 6.

6 While the tradeoff functions considered in this work are defined in terms of conditional von Neumann entropies, the rate curves of [35] are equal to a difference of $(1+\varepsilon)$-Rényi entropies (see [35, Section 6]). The latter cannot be larger than the tradeoff functions, which yield asymptotically optimal randomness extraction rates (as shown in [5]).
2. Preliminaries

2.1. Notation. In the table below, we summarise some of the notation used throughout the paper:

| Symbol | Definition |
|--------|------------|
| $A, B, C, \ldots$ | Quantum systems, and their associated Hilbert spaces |
| $\mathcal{L}(A, B)$ | Set of linear operators from $A$ to $B$ |
| $\mathcal{L}(A)$ | $\mathcal{L}(A, A)$ |
| $X_{AB}$ | Operator in $\mathcal{L}(A \otimes B)$ |
| $X_{B \rightarrow A}$ | Operator in $\mathcal{L}(A, B)$ |
| $D(A)$ | Set of normalised density operators on $A$ |
| $D_{\leq}(A)$ | Set of sub-normalised density operators on $A$ |
| $\text{Pos}(A)$ | Set of positive semidefinite operators on $A$ |
| $X^{-1}$ for $X \in \text{Pos}(A)$ | Generalised inverse, such that $XX^{-1}X = X$ holds |
| $X_A \geq Y_A$ | $X_A - Y_A \in \text{Pos}(A)$ |
| $A_i^j$ (with $j \geq i$) | Given $n$ systems $A_1, \ldots, A_n$, this is a shorthand for $A_i, \ldots, A_j$ |
| $A^B$ | Often used as shorthand for $A_1, \ldots, A_n$ |
| $\log(x)$ | Logarithm of $x$ in base 2 |

Throughout the paper, we restrict ourselves to finite-dimensional Hilbert spaces. Furthermore, we use the following notation for classical-quantum states $\rho_{XA} \in D(X \otimes A)$ with respect to the basis $\{|x\rangle\}_{x \in \mathcal{X}}$ of the system $X$. For any $x \in \mathcal{X}$, we let $\rho_{A,x} = \langle x | \rho_{AX} | x \rangle$ so that $\rho_{XA} = \sum_{x \in \mathcal{X}} |x\rangle \langle x| \otimes \rho_{A,x}$. To refer to the conditional state, we write $\rho_{A|x} = \frac{\rho_{A,x}}{\text{tr}(\rho_{A,x})}$. An event $\Omega \subseteq \mathcal{X}$ in this paper refers to a subset of $\mathcal{X}$ and we can similarly define $\rho_{XA|\Omega} = \frac{1}{\text{tr}(\rho_{AX|\Omega})} \sum_{x \in \Omega} |x\rangle \langle x| \otimes \rho_{A,x}$, where we introduced the notation $\rho[\Omega] = \sum_{x \in \Omega} \text{tr}(\rho_{A,x})$. We also use the usual notation for the partial trace for conditional states, e.g., $\rho_{XA|\Omega} = \text{tr}_B(\rho_{XAB|\Omega})$.

For a density operator $\rho_{AB} \in D(A \otimes B)$ on a bipartite Hilbert space $A \otimes B$ we define the operator\footnote{Note that this operator is well defined, for the support of $\rho_{AB}$ is contained in the support of $\text{id}_A \otimes \rho_B$.}

$$\rho_{A|B} = (\text{id}_A \otimes \rho_B)^{-\frac{1}{2}} \rho_{AB} (\text{id}_A \otimes \rho_B)^{-\frac{1}{2}},$$

which may be interpreted as the state of $A$ conditioned on $B$, analogous to a conditional probability distribution. This operator was previously defined and studied in [6,30]. In the following, we will usually drop identity operators from the notation when they are clear from the context. We would thus write, for instance,

$$\rho_{A|B} = \rho_B^{-\frac{1}{2}} \rho_{AB} \rho_B^{-\frac{1}{2}}.$$

Remark 2.1. Let $A$ and $\tilde{A}$ be two isomorphic Hilbert spaces with orthonormal bases $\{|i\rangle_A\}$ and $\{|i\rangle_{\tilde{A}}\}$ and define

$$|\Theta\rangle = \sum_i |i\rangle_A \otimes |i\rangle_{\tilde{A}}.$$

Then any trace-non-increasing map $\mathcal{M} = \mathcal{M}_{B \rightarrow \tilde{A}}$ from $\mathcal{L}(\tilde{A})$ to $\mathcal{L}(B)$ can be represented as a “conditional state” (also known as the Choi-Jamiołkowski state) $M_{B|A}$ on $A \otimes B$ with the property that

$$M_{B|A} \geq 0 \quad \text{and} \quad \text{tr}_B[M_{B|A}] \leq \text{id}_A \quad (4)$$
and such that
\[ \mathcal{M}(\rho_A^{\frac{1}{2}}|\Theta\rangle\langle\Theta|\rho_A^{\frac{1}{2}}) = \rho_A^{\frac{1}{2}} M_{B|A} \rho_A^{\frac{1}{2}} \quad \forall \rho_A \] (5)
holds. Specifically, for any map \( \mathcal{M} \) one may define
\[ M_{B|A} = \mathcal{M}(|\Theta\rangle\langle\Theta|) ; \] (6)
it is then straightforward to verify the properties above.

Conversely, for any \( M_{B|A} \) such that (4) holds the map defined by
\[ \mathcal{M}(|i\rangle_{\bar{A}} \langle j|_{\bar{A}}) = \langle i|_{A} M_{B|A} |j\rangle_{A} \]
satisfies (6) and hence (5). It is also easy to verify that it is completely positive and trace non-increasing.

We mention here a slight abuse of terminology: for a completely positive map \( M_{B\leftarrow A} \) from \( \mathcal{L}(A) \) to \( \mathcal{L}(B) \), we often use a shorthand to indicate the systems it acts on and simply say that it maps \( A \) to \( B \).

2.2. Background on quantum Markov chains. The concept of quantum Markov chains will be used throughout the paper, and here we give some relevant basic facts about them. Let \( \{a_j\}_{j \in J} \) and \( \{c_j\}_{j \in J} \) be families of Hilbert spaces and let \( B \) be a Hilbert space such that
\[ B \cong \bigoplus_{j \in J} a_j \otimes c_j , \] (7)
holds. Let us furthermore denote by \( V = \bigoplus_{j \in J} V_{a_jc_j \leftarrow B} \) the corresponding isomorphism. It is convenient to treat \( \bigoplus_{j \in J} a_j \otimes c_j \) as a subspace of the product \( a \otimes c \) of the spaces
\[ a = a \quad \text{and} \quad c = c . \]

The mapping \( V \) may then be viewed as an embedding of \( B \) into \( a \otimes c \). Given a density operator \( \rho_B \), we denote by \( \rho_{ac} \) the density operator \( V\rho_B V^\dagger \). More generally, for a multipartite density operator \( \rho_{AB} \), we write \( \rho_{Aac} \) for \( V\rho_{AB} V^\dagger \). Furthermore, for any \( j \in J \), we denote by \( \rho_{a_jc_j} \) the projection of \( \rho_{Aac} \) onto the subspace defined by \( a_j \otimes c_j \), i.e.,
\[ \rho_{a_jc_j} = V_{a_jc_j \leftarrow B} \rho_{AB} V_{a_jc_j \leftarrow B}^\dagger . \] (8)

A tri-partite density operator \( \rho_{ABC} \) is said to obey the Markov chain condition \( A \leftrightarrow B \leftrightarrow C \) if there exists a decomposition of \( B \) of the form (7) such that
\[ \rho_{ABC} \cong \rho_{Aac} = \bigoplus_{j \in J} q_j \hat{\rho}_{a_j} \otimes \hat{\rho}_{c_jC} \] (9)
where \( \{q_j\}_{j \in J} \) is a probability distribution and \( \{\hat{\rho}_{a_j}\}_{j \in J} \) and \( \{\hat{\rho}_{c_jC}\}_{j \in J} \) are families of density operators \([26,28,39]\). It follows from this decomposition that a state \( \rho_{ABC} \)

\[ \oplus \] denotes the orthogonal direct sum.
obeying the Markov chain condition can be reconstructed from \( \rho_{AB} \) with a map \( T_{BC} \leftarrow B \)
acting only on \( B \) \[39\]:

\[
\rho_{ABC} = \mathcal{I}_A \otimes T_{BC} \leftarrow B(\rho_{AB}).
\] (10)

Another useful characterization of the Markov chain condition for \( \rho_{ABC} \) is given by the entropic equality

\[
I(A : C|B)_\rho = H(AB)_\rho + H(BC)_\rho - H(B)_\rho - H(ABC)_\rho
\]
where \( H(A)_\rho = -\text{tr}(\rho_A \log \rho_A) \) is the von Neumann entropy.

2.3. Entropic quantities. The formulation of the main claim refers to smooth entropies, which can be defined as follows.

**Definition 2.2.** For any density operator \( \rho_{AB} \) and for \( \varepsilon \in [0, 1] \) the \( \varepsilon \)-smooth min- and max-entropies of \( A \) conditioned on \( B \) are

\[
H^\varepsilon_{\text{min}}(A|B)_\rho = -\log \inf \inf_{\tilde{\rho}_{AB}, \sigma_B} \| \frac{1}{\sqrt{\tilde{\rho}_{AB}}} \sigma_B^{-\frac{1}{2}} \|_\infty^2,
\]

\[
H^\varepsilon_{\text{max}}(A|B)_\rho = \log \inf \sup_{\tilde{\rho}_{AB}, \sigma_B} \| \frac{1}{\sqrt{\tilde{\rho}_{AB}}} \sigma_B^{-\frac{1}{2}} \|_1^2,
\]
respectively, where \( \tilde{\rho} \) is any non-negative operator with trace at most 1 that is \( \varepsilon \)-close to \( \rho \) in terms of the purified distance \( P(\rho, \tilde{\rho}) = \sqrt{1 - \left(\text{tr}\left(\sqrt{\sqrt{\rho} \sqrt{\tilde{\rho}}}\right)\right)^2} \) whenever either \( \rho \) or \( \tilde{\rho} \) is normalized.

The proof we present here heavily relies on the sandwiched relative Rényi entropies introduced in \[37, 64\]. These relative entropies can be used to define a conditional entropy.10

**Definition 2.3.** For any density operator \( \rho_{AB} \) and for \( \alpha \in (0, 1) \cup (1, \infty) \) the sandwiched \( \alpha \)-Rényi entropy of \( A \) conditioned on \( B \) is defined as

\[
H_\alpha(A|B)_\rho = -\frac{1}{\alpha'} \log \| \rho_{AB}^{1-\frac{1}{2\alpha'}} \rho_B^{\frac{1}{2\alpha'}} \|_{2\alpha}^2,
\]

where \( \alpha' = \frac{\alpha - 1}{\alpha} \) and where \( \| X \|_\alpha = \text{tr}\left( (X^\dagger X)^\frac{\alpha}{2} \right)^\frac{1}{\alpha} \). Note that \( \alpha' \) is the inverse of the Hölder conjugate of \( \alpha \).

We note that, while the function \( X \mapsto \| X \|_\alpha \) is a norm for \( \alpha \geq 1 \), this is not the case for \( \alpha < 1 \) since it does not satisfy the triangle inequality. Some key properties of this function are summarised in Appendix A. Using them, the sandwiched Rényi entropies may be rewritten as

\[
H_\alpha(A|B)_\rho = \frac{\alpha}{1 - \alpha} \log \left( \left\| \rho_B^{\frac{1-\alpha}{2\alpha}} \rho_{AB} \rho_B^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha \right)
= \frac{1}{1 - \alpha} \log \text{tr}\left( \left( \rho_B^{\frac{1-\alpha}{2\alpha}} \rho_{AB} \rho_B^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right).
\]

It turns out that there are multiple ways of defining conditional entropies from relative entropies. Another variant that will be needed in this work is the following:

\[9\] The purified distance is defined as \( P(\rho, \tilde{\rho}) = \sqrt{1 - (\text{tr}\left(\sqrt{\sqrt{\rho} \sqrt{\tilde{\rho}}}\right))^2} \) whenever either \( \rho \) or \( \tilde{\rho} \) is normalized.

\[10\] We note that there are at least two common variants for how to define a conditional entropy based on a relative entropy. We refer to “Appendix B” for more details.
Lemma 3.1. Let \( \rho_{AB} \) and \( \sigma_B \), be density operators and let \( \alpha \in (0, 1) \cup (1, \infty) \), we define

\[
H_\alpha^+(A|B)_\rho = -\inf_{\sigma_B} \frac{1}{\alpha'} \log \left\| \frac{1}{\alpha} \rho_{AB} \sigma_B^{-\frac{\alpha'}{\alpha}} \right\|_2^2 ,
\]

where the infimum is over all sub-normalised density operators on \( B \).

Other relevant facts about the sandwiched Rényi entropy and the corresponding notion of relative entropy can be found in Appendix B.

3. Chain Rule for Rényi Entropies

As explained in the introduction, our main result can be regarded as a generalisation of the Quantum Asymptotic Equipartition Property [54], corresponding to (2). The approach used for the proof of the latter is to bound both the smooth min-entropy and the von Neumann entropy by Rényi entropies with an appropriate parameter \( \alpha \). The IID assumption is then used to decompose the Rényi entropy into a sum of \( n \) terms. However, since our main claim, Eq. (3), is supposed to hold for general non-IID states, we do not have this luxury, and we must somehow decompose the Rényi entropy into \( n \) terms using other means. The tool we will use for this purpose is a chain rule for Rényi entropies, which we present as a separate theorem (Theorem 3.2). We start by stating a more general version that will be useful in the proof of the main theorem.

Lemma 3.1. Let \( \rho_{A_1A_2B} \) and \( \sigma_B \) be density operators and let \( \alpha \in (0, \infty) \). Then

\[
D_\alpha(\rho_{A_1B} \| \text{id}_{A_1} \otimes \sigma_B) - D_\alpha(\rho_{A_1A_2B} \| \text{id}_{A_1A_2} \otimes \sigma_B) = H_\alpha(A_2|A_1B)_\nu ,
\]

where

\[
v_{A_1A_2B} = v_{A_1B}^{\frac{1}{2}} \rho_{A_2A_1B} v_{A_1B}^{\frac{1}{2}} \quad \text{with} \quad v_{A_1B} = \frac{1}{\text{tr} \left( \rho_{A_1B}^{\frac{1}{2}} \right)} \left( \rho_{A_1B}^{\frac{1}{2}} \right)^{\alpha} . \tag{11}
\]

We note that \( v_{A_1B} = \text{tr}_{A_2}(v_{A_1A_2B}) \), which justifies the notation.

Proof. When \( \alpha = 1 \), this equality follows directly from the definition of the entropies. To prove the equality for \( \alpha \in (0, 1) \cup (1, \infty) \) we consider a purification \( |\psi\rangle_{A_1A_2BE} \) of \( \rho_{A_1A_2B} \). Using Lemma B.2 and setting \( \alpha' = \frac{\alpha-1}{\alpha} \), we have

\[
D_\alpha(\rho_{A_1A_2B} \| \text{id}_{A_1A_2} \otimes \sigma_B)
= \sup_{\tau_E} \frac{1}{\alpha'} \log \left\| \sigma_B^{-\frac{\alpha'}{\alpha}} \otimes \frac{\alpha'}{\alpha} \right\|_2^2
= \sup_{\tau_E} \frac{1}{\alpha'} \log \left\| \left( \sigma_B^{-\frac{\alpha'}{\alpha}} \rho_{A_1B}^{\frac{1}{2}} \otimes \frac{\alpha'}{\alpha} \right) \rho_{A_1B}^{\frac{1}{2}} |\psi\rangle \langle |\psi| \rho_{A_1B}^{\frac{1}{2}} \right\|_2^2
= \sup_{\tau_E} \frac{1}{\alpha'} \log \text{tr} \left( \left( \sigma_B^{-\frac{\alpha'}{\alpha}} \rho_{A_1B}^{\frac{1}{2}} \otimes \frac{\alpha'}{\alpha} \right) \rho_{A_1B}^{\frac{1}{2}} |\psi\rangle \langle |\psi| \rho_{A_1B}^{\frac{1}{2}} \right) \left( \rho_{A_1B}^{\frac{1}{2}} \sigma_B^{-\frac{\alpha'}{\alpha}} \otimes \frac{\alpha'}{\alpha} \right) \rho_{A_1B}^{\frac{1}{2}} |\psi\rangle \langle |\psi| \rho_{A_1B}^{\frac{1}{2}} \right) \left( \text{id}_{A_1B} \otimes \frac{\alpha'}{\alpha} \right)
= \sup_{\tau_E} \frac{1}{\alpha'} \log \text{tr} \left( \left( \rho_{A_1B}^{\frac{1}{2}} \sigma_B^{-\frac{\alpha'}{\alpha}} \rho_{A_1B}^{\frac{1}{2}} \otimes \frac{\alpha'}{\alpha} \right) \rho_{A_1B}^{\frac{1}{2}} |\psi\rangle \langle |\psi| \rho_{A_1B}^{\frac{1}{2}} \right) \left( \rho_{A_1B}^{\frac{1}{2}} \sigma_B^{-\frac{\alpha'}{\alpha}} \otimes \frac{\alpha'}{\alpha} \right) \rho_{A_1B}^{\frac{1}{2}} |\psi\rangle \langle |\psi| \rho_{A_1B}^{\frac{1}{2}} \right) \left( \text{id}_{A_1B} \otimes \frac{\alpha'}{\alpha} \right) .
\]
By the definition of \( \nu_{A_1B} \), we get

\[
D_{\alpha}(\rho_{A_1A_2B} \| \text{id}_{A_1A_2} \otimes \sigma_B)
= \frac{1}{\alpha'} \log \left( \text{tr} \left( \frac{1}{\alpha} \rho_{A_1B}^{\alpha} \rho_{A_1B}^{-\alpha'} \frac{1}{\alpha} \right) \right)^{\frac{1}{\alpha'}}
+ \sup_{\tau_E} \frac{1}{\alpha'} \log \text{tr} \left( \left( \frac{1}{\alpha} \rho_{A_1B}^{\alpha} \right)^{\frac{1}{\alpha'}} \rho_{A_1B}^{-\frac{1}{\alpha'}} \right)

= D_{\alpha}(\rho_{A_1B} \| \text{id}_{A_1} \otimes \sigma_B)
+ \sup_{\tau_E} \frac{1}{\alpha'} \log \left( \left( \frac{1}{\alpha} \rho_{A_1B}^{\alpha} \right)^{\frac{1}{\alpha'}} \rho_{A_1B}^{-\frac{1}{\alpha'}} \right)

= D_{\alpha}(\rho_{A_1B} \| \text{id}_{A_1} \otimes \sigma_B) + \sup_{\tau_E} \frac{1}{\alpha'} \log \left\| \frac{1}{\rho_{A_1B}} \otimes \frac{1}{\sigma_B} \right\|^2,
\]

where we defined the pure state \( |v\rangle_{A_1A_2B} = \frac{1}{\rho_{A_1B}} \rho_{A_1B}^{-\frac{1}{\alpha'}} |\psi\rangle \), which is a purification of \( \nu_{A_1A_2B} \). To conclude we use the fact that \( \nu_{A_1B} = \text{tr}_{A_2}(\nu_{A_1A_2B}) \) and Lemma B.2. \( \square \)

By choosing \( \sigma_B = \rho_B \) in Lemma 3.1, we directly obtain a chain rule for the Rényi entropies:

**Theorem 3.2.** Let \( \rho_{A_1A_2B} \) be a density operator and let \( \alpha \in (0, \infty) \). Then

\[
H_{\alpha}(A_1A_2|B)_\rho = H_{\alpha}(A_1|B)_\rho + H_{\alpha}(A_2|A_1B)_\rho, \quad (12)
\]

where

\[
\nu_{A_1A_2B} = \frac{1}{\rho_{A_1B}} \rho_{A_2A_1B}^{\frac{1}{\alpha'}} \nu_{A_1B} \quad \text{with} \quad \nu_{A_1B} = \left( \frac{1}{\rho_{A_1B}} \rho_{A_1B}^{-\frac{1}{\alpha'}} \right)^{\frac{1}{\alpha'}} \left( \frac{1}{\rho_{A_1B}} \rho_{A_1B}^{-\frac{1}{\alpha'}} \right)^{\frac{1}{\alpha'}}. \quad (13)
\]

One drawback of the above result is that we are seldom interested in the particular state \( \nu \) defined in the theorem statement. It is therefore generally more useful to present the result in a slightly weaker form, where the state \( \nu \) is chosen to be the worst case over an appropriate class of density operators. When \( \rho \) obeys the Markov chain condition \( A_1 \leftrightarrow B_1 \leftrightarrow B_2 \), we obtain the following result.

**Theorem 3.3.** Let \( \rho_{A_1B_1A_2B_2} \) be a density operator such that the Markov chain condition \( A_1 \leftrightarrow B_1 \leftrightarrow B_2 \) holds and let \( \alpha \in (0, \infty) \). Then

\[
\inf_{\nu} H_{\alpha}(A_2|B_2A_1B_1)_\nu \leq H_{\alpha}(A_1A_2|B_1B_2)_\rho - H_{\alpha}(A_1|B_1)_\rho \leq \sup_{\nu} H_{\alpha}(A_2|B_2A_1B_1)_\nu
\]

(14)

where the supremum and infimum range over density operators \( \nu \) such that \( \nu_{A_2B_2|A_1B_1} = \rho_{A_2B_2|A_1B_1} \) holds.
Proof. We apply Theorem 3.2 with $B = B_1B_2$. The Markov chain condition implies that $H_\alpha(A_1|B_1B_2)_\rho = H_\alpha(A_1|B_1)_\rho$. To see this for $\alpha \in \left(\frac{1}{2}, \infty\right)$, we could use the recoverability condition (10) for Markov chains together with the monotonicity of $D_\alpha$ under quantum channels [8,24,37,64]. We can also see it for all $\alpha \in (0, \infty)$ using the structure of a Markov chain stated in (9). Namely, there exists a decomposition $\bigoplus_j a_j \otimes b_j$ of the system $B_1$ such that

$$
\rho_{A_1B_1B_2} = \bigoplus_j q_j \hat{\rho}_{A_1a_j} \otimes \hat{\rho}_{b_jB_2}
$$

(15)

holds, where $\{q_j\}$ is a probability distribution and where $\{\hat{\rho}_{Aa_j}\}$ and $\{\hat{\rho}_{b_jB}\}$ are families of density operators. Then,

$$
H_\alpha(A_1|B_1B_2)_\rho = \frac{1}{1 - \alpha} \log \text{tr} \left( \left( \rho_{B_1B_2} \rho_{A_1B_1B_2} \rho_{A_1B_1B_2}^{-\alpha} \right)^\alpha \right)
$$

$$
= \frac{1}{1 - \alpha} \log \text{tr} \left( \bigoplus_j q_j \left( \hat{\rho}_{A_1a_j}^{-\alpha} \otimes \hat{\rho}_{b_jB_2} \left( \hat{\rho}_{A_1a_j} \otimes \hat{\rho}_{b_jB_2} \right) \right)^\alpha \right)
$$

$$
= \frac{1}{1 - \alpha} \log \text{tr} \left( \bigoplus_j q_j \left( \hat{\rho}_{A_1a_j}^{-\alpha} \otimes \hat{\rho}_{b_jB_2} \right)^\alpha \right) = H_\alpha(A_1|B_1)_\rho .
$$

To prove (14), it only remains to show that the state $v_{A_1A_2B_1B_2}$ defined in (13) satisfies $v_{A_2B_2|A_1B_1} = \rho_{A_2B_2|A_1B_1}$. For that, we again use the fact that $\rho_{A_1B_1B_2}$ forms a Markov chain. As we will be using this statement later in other contexts, we state it as a claim.

\[\square\]

Claim 3.4. Let $\rho_{A_1A_2B_2}$ be a density operator such that the Markov chain condition $A_1 \leftrightarrow B_1 \leftrightarrow B_2$ holds, let $\alpha \in (0, \infty)$ and let $v_{A_1A_2B_2}$ be as in (13) with $B \rightarrow B_1B_2$. Then $v_{A_2B_2|A_1B_1} = \rho_{A_2B_2|A_1B_1}$.

Letting $Z = \text{tr} \left( \frac{1}{\alpha} \rho_{A_1B_1B_2}^{-\alpha} \rho_{B_1B_2} \rho_{A_1B_1B_2}^{-\alpha} \right)^\alpha$, the decomposition (15) allows us to write

$$
v_{A_1B_1B_2} = \frac{1}{Z} \left( \bigoplus_j q_j \left( \hat{\rho}_{A_1a_j}^{-\alpha} \otimes \hat{\rho}_{b_jB_2} \right)^\alpha \right) = \frac{1}{Z} \left( \bigoplus_j q_j \left( \hat{\rho}_{A_1a_j}^{-\alpha} \otimes \hat{\rho}_{b_jB_2} \right)^\alpha \right)
$$

(16)

It follows that

$$
v_{-\frac{1}{2}}^{-\frac{1}{2}} = \bigoplus_j \hat{\rho}_{A_1a_j}^{\frac{1}{2}} \otimes \hat{\rho}_{b_jB_2}^{\frac{1}{2}}
$$

(19)

$$
= \rho_{A_1B_1}^{\alpha} \rho_{A_1B_1B_2}^{\alpha} .
$$

(20)
where \( \hat{\rho}^{0}_{A_{1}A_{j}} \) is the projector onto the support of \( \hat{\rho}_{A_{1}A_{j}} \). We used for the first equality the fact that the support of the operator \( \left( \hat{\rho}^{\frac{1}{2}}_{A_{1}A_{j}} \hat{\rho}^{-\alpha} \hat{\rho}^{\frac{1}{2}}_{A_{1}A_{j}} \right)^{\alpha} \) is the same as the support of \( \hat{\rho}_{A_{1}A_{j}} \). As a result, we find

\[
\begin{align*}
\nu_{A_{2}B_{1}|A_{1}B_{1}} & = -\frac{1}{2} \nu_{A_{1}B_{1}} v_{A_{1}B_{1}}^{1} \nu_{A_{2}|A_{1}} v_{A_{2}|A_{1}B_{1}} v_{A_{1}B_{1}} \nu_{A_{1}B_{1}} \nu_{A_{1}B_{1}}^{1} \\
 & = \rho_{A_{1}B_{1}}^{1} \rho_{A_{1}B_{1}}^{1} \nu_{A_{2}|A_{1}} v_{A_{2}|A_{1}B_{1}} v_{A_{1}B_{1}} \nu_{A_{1}B_{1}} \nu_{A_{1}B_{1}}^{1} \\
 & = \rho_{A_{2}B_{1}|A_{1}B_{1}}. \tag{21}
\end{align*}
\]

This concludes the proof of Claim 3.4 and gives the desired statement. □

The following simple corollary expresses the above chain rules in terms of quantum channels, i.e., trace preserving completely positive (TPCP) maps, rather than conditional states.

**Corollary 3.5.** Let \( \rho_{RA_{1}B_{1}}^{0} \) be a density operator on \( R \otimes A_{1} \otimes B_{1} \), \( \mathcal{M} = \mathcal{M}_{A_{2}B_{2} \leftarrow R} \) be a TPCP map and \( \alpha \in (0, \infty) \). Assuming that \( \rho_{A_{1}B_{1}A_{2}B_{2}} = \mathcal{M}(\rho_{RA_{1}B_{1}}^{0}) \) satisfies the Markov condition \( A_{1} \leftrightarrow B_{1} \leftrightarrow B_{2} \), we have

\[
\inf_{\omega} H_{\alpha}(A_{2}|B_{2}A_{1}B_{1})_{\mathcal{M}(\omega)} \leq H_{\alpha}(A_{1}A_{2}|B_{1}B_{2})_{\mathcal{M}(\rho^{0})} - H_{\alpha}(A_{1}|B_{1})_{\rho^{0}} \leq \sup_{\omega} H_{\alpha}(A_{2}|B_{2}A_{1}B_{1})_{\mathcal{M}(\omega)}
\]

where the supremum and infimum range over density operators \( \omega_{RA_{1}B_{1}} \) on \( R \otimes A_{1} \otimes B_{1} \). Moreover, if \( \rho_{RA_{1}B_{1}}^{0} \) is pure then we can optimise over pure states \( \omega_{RA_{1}B_{1}} \).

**Proof.** We apply Theorem 3.3 to \( \rho_{A_{1}B_{1}A_{2}B_{2}} \). It suffices to show that the optimisation over \( \nu_{A_{1}B_{1}A_{2}B_{2}} \) satisfying \( \nu_{A_{2}B_{1}|A_{1}B_{1}} = \rho_{A_{2}B_{2}|A_{1}B_{1}} \) is contained in the optimisation over \( \omega_{RA_{1}B_{1}} \). For this, let \( \nu_{A_{1}B_{1}A_{2}B_{2}} \) be any density operator satisfying \( \nu_{A_{2}B_{2}|A_{1}B_{1}} = \rho_{A_{2}B_{2}|A_{1}B_{1}} \), i.e.,

\[
\nu_{A_{1}B_{1}A_{2}B_{2}} = \nu_{A_{1}B_{1}}^{\frac{1}{2}} \rho_{A_{1}B_{1}}^{\frac{1}{2}} \nu_{A_{2}B_{2}}^{\frac{1}{2}} \nu_{A_{1}B_{1}}^{\frac{1}{2}}. \tag{24}
\]

Now we choose

\[
\omega_{RA_{1}B_{1}} = \nu_{A_{1}B_{1}}^{\frac{1}{2}} \rho_{A_{1}B_{1}}^{\frac{1}{2}} \nu_{A_{2}B_{2}}^{\frac{1}{2}} \nu_{A_{1}B_{1}}^{\frac{1}{2}}. \tag{25}
\]

We then see that

\[
\mathcal{M}(\omega_{RA_{1}B_{1}}) = \nu_{A_{1}B_{1}}^{\frac{1}{2}} \rho_{A_{1}B_{1}}^{\frac{1}{2}} \mathcal{M}(\rho_{RA_{1}B_{1}}^{0}) \rho_{A_{1}B_{1}}^{\frac{1}{2}} \nu_{A_{1}B_{1}}^{\frac{1}{2}}
\]

\[
= \nu_{A_{1}B_{1}}^{\frac{1}{2}} \rho_{A_{2}B_{2}|A_{1}B_{1}} \nu_{A_{1}B_{1}}^{\frac{1}{2}}.
\]

□
4. Entropy Accumulation

This section is devoted to the main result on entropy accumulation. The statement is formulated in its fully general form as Theorem 4.4 and presented in a slightly simplified version as Corollary 4.8. We also give a formulation that corresponds to statement (3) of the introduction (Corollary 4.9). Finally, we show how the Quantum Asymptotic Equipartition Property follows as a special case (cf. Corollary 4.10).

For \( i \in \{1, \ldots, n\} \), let \( \mathcal{M}_i \) be a TPCP map from \( R_{i-1} \) to \( X_i A_i B_i R_i \), where \( A_i \) is finite-dimensional and where \( X_i \) represents a classical value from an alphabet \( \mathcal{X} \) that is determined by \( A_i \) and \( B_i \) together. More precisely, we require that, \( \mathcal{M}_i = T_i \circ \mathcal{M}'_i \) where \( \mathcal{M}'_i \) is an arbitrary TPCP map from \( R_{i-1} \) to \( A_i B_i R_i \) and \( T_i \) is a TPCP map from \( A_i B_i \) to \( X_i A_i B_i \) of the form

\[
T_i(W_{A_i B_i}) = \sum_{y \in \mathcal{Y}, z \in \mathcal{Z}} (\Pi_{A_i, y} \otimes \Pi_{B_i, z}) W_{A_{i}, B_{i}} (\Pi_{A_{i}, y} \otimes \Pi_{B_{i}, z}) \otimes |t(y, z)\rangle\langle t(y, z)|_{X_i},
\]

(25)

where \( \{\Pi_{A_i, y}\} \) and \( \{\Pi_{B_i, z}\} \) are families of mutually orthogonal projectors on \( A_i \) and \( B_i \), and where \( t : \mathcal{Y} \times \mathcal{Z} \rightarrow \mathcal{X} \) is a deterministic function (cf. Figs. 1 and 2). Special cases of interest are when \( X_i \) is trivial and \( T_i \) is the identity map, and when \( X_i = t(Y_i, Z_i) \) where \( Y_i \) and \( Z_i \) are classical parts of \( A_i \) and \( B_i \), respectively. Note that the maps \( T_i \) have the property that, for any operator \( W_{X_i A_i B_i} \), if \( W_{X_i A_i B_i} = T_i(W_{A_i B_i}) \) then \( W_{X_i A_i B_i} = T_i(W_{A_i B_i}) \).

The entropy accumulation theorem stated below will hold for states of the form

\[
\rho_{A_i^0 B_i^0 X_i^0 E} = (\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1 \otimes \mathcal{I}_E)(\rho_{R_0 E}^0)
\]

(26)

where \( \rho_{R_0 E}^0 \in \text{D}(R_0 \otimes E) \) is a density operator on \( R_0 \) and an arbitrary system \( E \). In addition, we require that the Markov conditions

\[
A_i^{i-1} \leftrightarrow B_i^{i-1} E \leftrightarrow B_i
\]

(27)

be satisfied for all \( i \in \{1, \ldots, n\} \).

Let \( \mathbb{P} \) be the set of probability distributions on the alphabet \( \mathcal{X} \) of \( X_i \), and let \( R \) be a system isomorphic to \( R_{i-1} \). For any \( q \in \mathbb{P} \) we define the set of states

\[
\Sigma_i(q) = \{v_{X_i A_i B_i R_i} = (\mathcal{M}_i \otimes \mathcal{I}_R)(\omega_{R_{i-1} R}) : \omega \in \text{D}(R_{i-1} \otimes R) \text{ and } v_{X_i} = q\},
\]

(28)

where \( v_{X_i} \) denotes the probability distribution over \( \mathcal{X} \) with the probabilities given by \( \langle x | v_{X_i} | x \rangle \).

**Definition 4.1.** A real function \( f \) on \( \mathbb{P} \) is called a min- or max-tradeoff function for \( \mathcal{M}_i \) if it satisfies

\[
f(q) \leq \inf_{v \in \Sigma_i(q)} H(A_i | B_i R)_v \quad \text{or} \quad f(q) \geq \sup_{v \in \Sigma_i(q)} H(A_i | B_i R)_v,
\]

respectively.\(^{11}\)

\(^{11}\) If the set \( \Sigma_i(q) \) is empty then the infimum and supremum are by definition equal to \( \infty \) and \(-\infty\), respectively, so that the conditions are trivial.
Remark 4.2. To determine the infimum $\inf_{\nu \in \Sigma_i(q)} H(A_i | B_i R)_{\nu}$, we may assume that $\omega_{R_{i-1} R}$ in the definition of $\Sigma_i(q)$ is pure. In fact, including a purifying system in $R$ cannot increase $H(A_i | B_i R)$ because of strong subadditivity. Similarly, to calculate the supremum $\sup_{\nu \in \Sigma_i(q)} H(A_i | B_i R)_{\nu}$, we may assume that $\omega_{R_{i-1} R}$ is a product state or that $R$ is trivial. This justifies the fact that we assumed $R$ is isomorphic to $R_{i-1}$ in the definition of $\Sigma_i(q)$.

Remark 4.3. As we will see in the proof below, one can also impose the constraint on the set $\Sigma_i(q)$ that the system $R$ be isomorphic to $A_i^{-1} B_i^{-1} E$. Furthermore, if a part of the latter is classical in $\rho$, one can restrict $\Sigma_i(q)$ to states satisfying this property.

In the following, we denote by $\nabla f$ the gradient of a function $f$. (Note that in Theorem 4.4 and Proposition 4.5 $f$ is an affine function, so that $\nabla f$ is a constant.) We write $\text{freq}(X^n_1)$ for the distribution on $X'$ defined by $\text{freq}(X^n_1)(x) = |\{i \in [1, \ldots, n] : X_i = x\}|$. We also recall that in this context, an event $\Omega$ is defined by a subset of $\mathcal{X}^n$ and we write $\rho[\Omega] = \sum_{x_1^n \in \Omega} \text{tr}(\rho_{A^n_1 B^n_1 E, x^n_1})$ for the probability of the event $\Omega$ and

$$
\rho_{X^n_1 A_i^n B_i^n E | \Omega} = \frac{1}{\rho[\Omega]} \sum_{x_1^n \in \Omega} |x_1^n| |x_1^n| \otimes \rho_{A_i^n B_i^n E, x^n_1}
$$

for the state conditioned on $\Omega$ (cf. Section 2.1).

Theorem 4.4. Let $\mathcal{M}_1, \ldots, \mathcal{M}_n$ and $\rho_{A^n_1 B^n_1 X^n_1 E}$ be such that (26) and the Markov conditions (27) hold, let $h \in \mathbb{R}$, let $f$ be an affine min-tradeoff function for $\mathcal{M}_1, \ldots, \mathcal{M}_n$, and let $\epsilon \in (0, 1)$. Then, for any event $\Omega \subseteq \mathcal{X}^n$ that implies $f(\text{freq}(X^n_1)) \geq h$,\footnote{We say that the event $\Omega$ implies $f(\text{freq}(X^n_1)) \geq h$ if for every $x^n_1 \in \Omega$, $f(\text{freq}(x^n_1)) \geq h$.}

$$
H_{\min}^\epsilon(A^n_1 | B^n_1 E)_{\rho | \Omega} > nh - c \sqrt{n}
$$

(29)

holds for $c = 2(\log(1 + 2d_A) + \|\nabla f\|_\infty) \sqrt{1 - 2 \log(\epsilon \rho[\Omega])}$, where $d_A$ is the maximum dimension of the systems $A_i$. Similarly,

$$
H_{\max}^\epsilon(A^n_1 | B^n_1 E)_{\rho | \Omega} < nh + c \sqrt{n}
$$

(30)

holds if $f$ is replaced by an affine max-tradeoff function and if $\Omega$ implies $f(\text{freq}(X^n_1)) \leq h$.

Before proceeding to the proof, some remarks are in order. The first is that the Markov chain assumption on the state is important as argued in Appendix C. Secondly, the system $E$ could have been included in $B_1$, but for the applications we consider, it is clearer to keep a separate system $E$ that is not affected by the processes $\mathcal{M}_1, \ldots, \mathcal{M}_n$. Thirdly, concerning the second order term, it is possible to replace $d_A$ with appropriate entropic quantities, as in the Quantum Asymptotic Equipartition Property [54], which could be useful when the systems $A_i$ are infinite-dimensional. The dependence of the second order term in the state and in the tradeoff function $f$ is studied in more detail in the subsequent work [19]. Finally, we note that the constraint that the tradeoff function be affine is not a severe restriction: given a convex min-tradeoff function, one can always choose a tangent hyperplane at a point of interest as an affine lower bound. This is illustrated in Corollary 4.7.

To prove the theorem, we will first show the following proposition, which is essentially a Rényi version of entropy accumulation. We then show how Theorem 4.4 follows from this proposition.
Proposition 4.5. Let $\mathcal{M}_1, \ldots, \mathcal{M}_n$ and $\rho_{A_1^n B_1^n X_1^n E}$ be such that (26) and the Markov conditions (27) hold, let $h \in \mathbb{R}$, and let $f$ be an affine min-tradeoff function $f$ for $\mathcal{M}_1, \ldots, \mathcal{M}_n$. Then, for any event $\Omega$ which implies $f(\text{freq}(X_1^n)) \geq h$,

$$H_{\alpha}^\uparrow(A_1^n | B_1^n E)_{\rho|\Omega} > nh - n \left(\frac{\alpha - 1}{4}\right) V^2 - \frac{\alpha}{\alpha - 1} \log \frac{1}{\rho[\Omega]} \quad (31)$$

holds for $\alpha$ satisfying $1 < \alpha < 1 + \frac{2}{V}$, and $V = 2 \left[\|\nabla f\|_{\infty}\right] + 2 \log(1 + 2d_A)$, where $d_A$ is the maximum dimension of the systems $A_i$. Similarly,

$$H_{\bar{\alpha}}^\downarrow(A_1^n | B_1^n E)_{\rho|\Omega} < nh + n \left(\frac{\alpha - 1}{4}\right) V^2 + \frac{\alpha}{\alpha - 1} \log \frac{1}{\rho[\Omega]} \quad (32)$$

holds if $f$ is replaced by an affine max-tradeoff function and if $\Omega$ implies $f(\text{freq}(X_1^n)) \leq h$.

Proof. We focus on proving the first inequality (31). The proof of the second inequality (32) is similar, we only point out the main differences in the course of the proof.

The first step of the proof is to construct a state that will allow us to lower-bound $H_{\alpha}^\uparrow(A_1^n | B_1^n E)_{\rho|\Omega}$ using the chain rule of Theorem 3.3, while ensuring that the tradeoff function is taken into account. Let $[g_{\min}, g_{\max}]$ be the smallest real interval that contains the range $f(\mathbb{P})$ of $f$, and set $\bar{g} = \frac{1}{2}(g_{\min} + g_{\max})$. Furthermore, for every $i$, let $D_i : X_i \rightarrow X_i D_i \bar{D}_i$, with dim $D_i = \text{dim} \bar{D}_i$, be a TPCP map defined as

$$D_i(W_{X_i}) = \sum_{x \in X} \langle x | W_{X_i} | x \rangle \cdot | x \rangle \langle x | X_i \otimes \tau(x) D_i \bar{D}_i,$$

where $\tau(x)$ is a mixture between a maximally entangled state on $D_i \otimes \bar{D}_i$ and a fully mixed state such that the marginal on $\bar{D}_i$ is uniform and such that $H_{\alpha}(D_i | \bar{D}_i)_{\tau(x)} = \bar{g} - f(\delta_x)$ (here $\delta_x$ stands for the distribution with all the weight on element $x$). To ensure that this is possible, we need to choose dim $D_i$ large enough, so we need to bound how large $\bar{g} - f(\delta_x)$ can be, positive or negative. By the definition of $\bar{g}$, $|\bar{g} - f(\delta_x)|$ cannot be larger than $\frac{1}{2}|g_{\max} - g_{\min}| \leq \|\nabla f\|_{\infty}$. We therefore take the dimension of the spaces $D_i$ to be equal to

$$d_D := \left\lceil 2\|\nabla f\|_{\infty} \right\rceil \leq 2\|\nabla f\|_{\infty}.$$

For later use, we note that we have

$$\log(1 + 2d_A d_D) \leq \lceil\|\nabla f\|_{\infty}\rceil + \log(1 + 2d_A) = \frac{V}{2}. \quad (33)$$

Now, let

$$\bar{\rho} := (D_n \circ \cdots \circ D_1)(\rho). \quad (34)$$

Note that $\bar{\rho}_{X_1^n A_1^n B_1^n E} = \rho_{X_1^n A_1^n B_1^n E}$.

One can think of the $D_i$ systems as an “entropy price” that encodes the tradeoff function. With these systems in place, the output entropy includes an extra term that allows the tradeoff function to be taken into account in the optimisation arising in Theorem 3.3. This is formalised by the following facts, which are proven in Claim 4.6:

$$H_{\alpha}^\uparrow(A_1^n | B_1^n E)_{\rho|\Omega} \geq H_{\alpha}^\uparrow(A_1^n D_1^n | B_1^n E \bar{D}_1^n)_{\bar{\rho}|\Omega} - n \bar{g} + nh, \quad (35)$$
\[ H_{\frac{1}{\alpha}}(A_1^n|B_1^nE)|_{\rho_{\Omega}} \leq H_{\frac{1}{\alpha}}(A_1^nD_1^n|B_1^nE\tilde{D}_1^n)|_{\tilde{\rho}_{\Omega}} - n\tilde{g} + nh. \] (36)

The next step is to relate the entropies on the conditional state \(\rho_{\Omega}\) to those on the unconditional state. To do this, we use Lemmas B.5 and B.6 applied to \(\tilde{\rho} = \rho[\Omega]\tilde{\rho}_{\Omega} + (\tilde{\rho} - \rho[\Omega]\tilde{\rho}_{\Omega})\), together with the fact that \(H_{\frac{1}{\alpha}} \geq H_{\alpha}\), and obtain\(^{13}\)

\[
H_{\frac{1}{\alpha}}(A_1^n|B_1^nE)|_{\rho_{\Omega}} \geq H_{\alpha}(A_1^nD_1^n|B_1^nE\tilde{D}_1^n)|_{\tilde{\rho}} - \frac{\alpha}{\alpha - 1} \log \frac{1}{\rho[\Omega]} - n\tilde{g} + nh,
\]

(37)

\[
H_{\frac{1}{\alpha}}(A_1^n|B_1^nE)|_{\rho_{\Omega}} \leq H_{\frac{1}{\alpha}}(A_1^nD_1^n|B_1^nE\tilde{D}_1^n)|_{\tilde{\rho}} + \frac{\alpha}{\alpha - 1} \log \frac{1}{\rho[\Omega]} - n\tilde{g} + nh.
\]

(38)

To show the desired inequality (31), it now suffices to prove that \(H_{\alpha}(A_1^nD_1^n|B_1^nE\tilde{D}_1^n)|_{\tilde{\rho}}\) is lower bounded by (roughly) \(n\tilde{g}\). To do that, we are now going to use the chain rule for Rényi entropies in the form of Corollary 3.5 \(n\) times on the state \(\tilde{\rho}\), with the following substitutions at step \(i\):

- \(A_1 \to A_1^{i-1}D_1^{i-1}\)
- \(B_1 \to B_1^{i-1}E\tilde{D}_1^{i-1}\)
- \(A_2 \to A_iD_i\)
- \(B_2 \to B_i\tilde{D}_i\)
- \(R \to R_{i-1}\)
- \(\mathcal{M} \to \text{tr}_{X_i} \circ \mathcal{D}_i \circ \mathcal{M}_i\).

To establish the Markov chain condition, we compute the conditional mutual information. Using the chain rule, we obtain

\[
I(A_1^{i-1}D_1^{i-1} : B_i\tilde{D}_i|B_1^{i-1}E\tilde{D}_1^{i-1}) = I(A_1^{i-1} : B_i\tilde{D}_i|B_1^{i-1}E\tilde{D}_1^{i-1}) + I(D_1^{i-1} : B_i\tilde{D}_i|A_1^{i-1}B_1^{i-1}E\tilde{D}_1^{i-1}).
\]

(39)

We first show that the second term is zero. By construction, \(D_1^{i-1}\tilde{D}_1^{i-1}\) conditioned on \(X_1^{i-1}\) is independent of all the other systems. This implies that \(I(D_1^{i-1}\tilde{D}_1^{i-1} : B_i\tilde{D}_i|X_1^{i-1}A_1^{i-1}B_1^{i-1}E) = 0\). In addition, using the fact that \(X_1^{i-1}\) is determined by \(A_1^{i-1}B_1^{i-1}\), the systems \(X_1^{i-1}\) can be removed from the conditioned without changing the value. Then, using the chain rule and together with the non-negativity of the conditional mutual information, this shows that \(I(D_1^{i-1} : B_i\tilde{D}_i|A_1^{i-1}B_1^{i-1}E\tilde{D}_1^{i-1}) = 0\).

To compute the first term in (39), we use the fact that \(\tilde{D}_1^n\) is uniform independently of \(A_1^nB_1^nE\) so that \(I(A_1^{i-1} : B_i\tilde{D}_i|B_1^{i-1}E\tilde{D}_1^{i-1}) = I(A_1^{i-1} : B_iB_1^{i-1}E)\). But then the assumed Markov condition on \(\rho_{A_1^nB_1^nE}\) implies that this quantity is zero and establishes the required condition to apply Corollary 3.5.

We thus obtain

\[
H_{\alpha}(A_1^nD_1^n|B_1^nE\tilde{D}_1^n)|_{\tilde{\rho}} \\
\geq \sum_{i} \inf_{\omega_{R_{i-1}}} H_{\alpha}(A_iD_i|B_i\tilde{D}_iR)(\mathcal{D}_i\circ\mathcal{M}_i)(\omega) \\
\geq \sum_{i} \inf_{\omega_{R_{i-1}}} H(A_iD_i|B_i\tilde{D}_iR)(\mathcal{D}_i\circ\mathcal{M}_i)(\omega) - n(\alpha - 1) \log^2(1 + 2d_A d_D)
\]

\(^{13}\) Note that the reason why we use \(H_{\alpha}^\dagger\) in one case but not the other is due to the difference between Lemmas B.5 and B.6.
\[ \sum_{i} \inf_{\omega_{R_{i+1}}^n} H(A_i D_i | B_i \bar{D}_i R)_{(D_i \circ \mathcal{M}_i)(\omega)} - \frac{n(\alpha - 1)}{4} V^2, \]  

where we have invoked Lemma B.9 in the second inequality and (33) in the last. Note that the restriction of this lemma that \( \alpha \) satisfy \( 1 < \alpha < 1 + 1/ \log(1 + 2d_A d_D) \) is implied by our assumption that \( \alpha < 1 + 2/V \).

The inﬁmum is taken over all states \( \omega_{R_{i+1}}^n \), where the system \( R \) is isomorphic to \( A_1^{i-1} D_1^{i-1} B_1^{i-1} \bar{D}_1^{i-1} E \). This condition can be further strengthened by redoing the above argument with Theorem 3.2 instead of Corollary 3.5. It turns out that the system \( R \) can be taken to be isomorphic to \( A_1^{i-1} B_1^{i-1} E \), as noted in Remark 4.3.

To prove that we can restrict in our optimisation the system \( R \) to be isomorphic to \( A_1^{i-1} B_1^{i-1} E \) and drop the systems \( D_1^{i-1} \bar{D}_1^{i-1} \), we use Theorem 3.2 directly instead of Corollary 3.5. In particular, using Lemma B.7 as for (43), we can write

\[
H_\alpha(A^n_1 D^n_1 | B^n_1 E \bar{D}^n_1)_{\bar{\rho}} = H_\alpha(A^n_1 X^n_1 D^n_1 | B^n_1 E \bar{D}^n_1)_{\bar{\rho}} = H_\alpha(A^n_1 X^n_1 D^n_1 | B^n_1 E \bar{D}^n_1)_{\bar{\rho}} + H_\alpha(A_n X_n D_n | B^n_1 E \bar{D}^n_1 A_1^{n-1} X_1^{n-1} D_1^{n-1})_{\nu^n},
\]

where we used the Markov chain condition \( A_1^{n-1} X_1^{n-1} D_1^{n-1} \leftrightarrow B_1^{n-1} E \bar{D}_1^{n-1} \leftrightarrow B_n D_n \)

and we deﬁned for all \( i \in \{1, \ldots, n\} \)

\[
v_{A_i X_i D_i}^{i} = \left( \frac{\tilde{p}_{A_i}^{1} X_i^{i-1} D_i^{i-1} B_i E \bar{D}_i \rho_{B_i E \bar{D}_i}^{1} \rho_{A_i}^{1} X_i^{i-1} D_i^{i-1} B_i E \bar{D}_i}{Z_i} \right)^{\alpha} \]

\[
v_{A_i X_i D_i}^{i} = \left( v_{A_i X_i D_i}^{i} \right)^{\alpha} \left( v_{A_i X_i D_i}^{i} \right)^{1/2} \tilde{p}_{A_i}^{1} X_i^{i-1} D_i^{i-1} B_i E \bar{D}_i \left( v_{A_i X_i D_i}^{i} \right)^{1/2},
\]

with \( Z_i = \text{tr} \left( \tilde{p}_{A_1}^{1} X_1^{1} D_1^{1} B_1 E \bar{D}_1 \rho_{B_1 E \bar{D}_1}^{1} \rho_{A_1}^{1} X_1^{1} D_1^{1} B_1 E \bar{D}_1 \right)^{\alpha} \).

Then we use the chain rule \( n - 2 \) more times to get

\[
H_\alpha(A^n_1 D^n_1 | B^n_1 E \bar{D}^n_1)_{\bar{\rho}} = \sum_{i} H_\alpha(A_i X_i D_i | B_i E \bar{D}_i A_1^{i-1} X_1^{i-1} D_1^{i-1})_{\nu^i}.
\]

We now use the properties of \( \bar{\rho} \) to simplify the entropies in the right hand side.

\[
v_{A_i X_i D_i}^{i} = \frac{1}{Z_i} \sum_{x \in \mathcal{X}_i^{i-1}} |x \rangle \langle x |_{X_i^{i-1}} \otimes \left( \tilde{p}_{A_i X_i D_i}^{1} B_i E \bar{D}_i \rho_{B_i E \bar{D}_i}^{1} \right)^{\alpha} \right)\]

Using the properties of the systems \( D_i \bar{D}_i \), we get for any \( x \in \mathcal{X}_i^{i-1}, \)

\[
\tilde{p}_{A_i}^{1} X_i^{i-1} D_i^{i-1} B_i E \bar{D}_i \rho_{B_i E \bar{D}_i}^{1} \rho_{A_i}^{1} X_i^{i-1} D_i^{i-1} B_i E \bar{D}_i \]

\[
= \left( \rho_{A_i}^{1} B_i E, x \otimes \tau(x) D_i^{i-1} \rho_{B_i E \bar{D}_i}^{1} \right)^{\alpha} \left( \rho_{B_i E} \otimes \rho_{\bar{D}_i} \right)^{1/2} \left( \rho_{B_i E} \otimes \rho_{\bar{D}_i} \right)^{1/2}.
\]
\[
\left( \rho_{A_1^{-1}B_1' E,x} \otimes \tau(x) \right)_{D_1^{-1} \tilde{D}_1^{-1}} \otimes \tilde{\rho}_{D_1} \right)^{\frac{1}{2}} \\
= \left( \rho_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \left( \rho_{B_1'E} \right)^{\frac{1}{2}} \left( \rho_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \otimes \tau(x) \left( \rho_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \otimes \tau(x) \left( \rho_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \otimes \tilde{\rho}_{D_1} \otimes \tilde{\rho}_{D_1},
\]
where we used the fact that \( \tilde{\rho}_{D_1} = \otimes_{j=1}^i \tilde{\rho}_{D_j} \). Letting
\[
\tau'(x)_{D_1^{-1} \tilde{D}_1^{-1}} = \left( \left( \rho_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \left( \rho_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \otimes \left( \rho_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \right)^{\frac{1}{2}} \alpha \frac{1}{\alpha} \frac{1}{\alpha} \frac{1}{\alpha},
\]
we can write
\[
\nu^j_{A_1^{-1}B_1' E,x} = \sum_{x \in \mathcal{X}_1^{-1}} |x\rangle \langle x|_{X_1^{-1}} \otimes \nu^j_{A_1^{-1}B_1' E,x} \otimes \tau'(x)_{D_1^{-1} \tilde{D}_1^{-1}} \otimes \tilde{\rho}_{D_1}.
\]
In addition
\[
\tilde{\rho}_{A_1^{-1}X_1^{-1}D_1^{-1}B_1' E}\tilde{D}_1 = \sum_{x \in \mathcal{X}_1^{-1}} |x\rangle \langle x|_{X_1^{-1}} \otimes \rho_{A_1^{-1}B_1' E,x} \otimes \tau(x)_{D_1^{-1} \tilde{D}_1^{-1}} \otimes \tilde{\rho}_{D_1}.
\]
As a result,
\[
\nu^j_{A_1^{-1}X_1^{-1}D_1^{-1}B_1' E\tilde{D}_1,x} = \sum_{x \in \mathcal{X}_1^{-1}} |x\rangle \langle x|_{X_1^{-1}} \otimes \nu^j_{A_1^{-1}X_1^{-1}D_1^{-1}B_1' E\tilde{D}_1,x}
\]
with
\[
\nu^j_{A_1^{-1}X_1^{-1}D_1^{-1}B_1' E\tilde{D}_1,x} = \left( \frac{1}{\nu^j_{A_1^{-1}B_1' E,x}} \right)^{\frac{1}{2}} \rho_{A_1^{-1}B_1' E,x} \otimes \tau'(x)_{D_1^{-1} \tilde{D}_1^{-1}} \otimes \tilde{\rho}_{A_1^{-1}B_1' E,x} \\
\times \left( \rho_{A_1^{-1}B_1' E,x} \nu^j_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \otimes \tau(x)_{D_1^{-1} \tilde{D}_1^{-1}} \otimes \tilde{\rho}_{A_1^{-1}B_1' E,x} \\
= \left( \nu^j_{A_1^{-1}B_1' E,x} \right)^{\frac{1}{2}} \rho_{A_1^{-1}B_1' E,x} \nu^j_{A_1^{-1}B_1' E,x} \otimes \tau(x)_{D_1^{-1} \tilde{D}_1^{-1}} \otimes \tilde{\rho}_{A_1^{-1}B_1' E,x} \\
\otimes \tau'(x)_{D_1^{-1} \tilde{D}_1^{-1}} \\
\otimes \tilde{\rho}_{A_1^{-1}B_1' E,x} \nu^j_{A_1^{-1}B_1' E,x} \otimes \tau(x)_{D_1^{-1} \tilde{D}_1^{-1}} \otimes \tilde{\rho}_{A_1^{-1}B_1' E,x} \\
\otimes \tau'(x)_{D_1^{-1} \tilde{D}_1^{-1}} \cdot
\]
As the system \( D_1^{-1} \tilde{D}_1^{-1} \) can be generated by only acting on \( X_1^{-1} \), we have by data processing
\[
H_\alpha(A_1X_1D_1|B_1'E\tilde{D}_1A_1^{-1}X_1^{-1}D_1^{-1})_{ij} = H_\alpha(A_1X_1D_1|B_1'E\tilde{D}_1A_1^{-1}X_1^{-1})_{ij}.
\]
We can then write
\[ v^j_{A^i} X^i_t | B^j_t E D_t \tilde{D}_t = (v^j_{A^i} X^i_t | B^j_t E)^2 \rho^{\frac{1}{2}} A^i_{x^i} X^i_t | B^j_t E \tilde{P}_t A^i_{x^i} X^i_t | B^j_t E \rho^{\frac{1}{2}} (v^j_{A^i} X^i_t | B^j_t E)^2 \cdot \]

We now use Claim 3.4 with the substitutions
\begin{itemize}
  \item $A_1 \rightarrow X_1^{i-1} A_1^{i-1}$
  \item $A_2 \rightarrow X_i A_i D_i \tilde{D}_i$
  \item $B_1 \rightarrow B_1^{i-1} E$
  \item $B_2 \rightarrow B_i$
\end{itemize}
and using the Markov property $X_1^{i-1} A_1^{i-1} \leftrightarrow B_1^{i-1} E \leftrightarrow B_i$. Thus, we have
\[ v_{X_i A_i D_i \tilde{D}_i B_i | X_1^{i-1} A_1^{i-1} B_1^{i-1} E} = \tilde{P}_{X_i A_i D_i \tilde{D}_i B_i | X_1^{i-1} A_1^{i-1} B_1^{i-1} E} \cdot \]

As a result, as in the proof of Corollary 3.5, we then get
\[ v^j_{A^i} X^i_t | B^j_t E D_t \tilde{D}_t = (D_i \circ M_i)(\omega^j_{R_{i-1} A_1^{i-1} X_1^{i-1} B_1^{i-1} E}), \]
where
\[ \omega^j_{R_{i-1} A_1^{i-1} X_1^{i-1} B_1^{i-1} E} := T_{X_1^{i-1} A_1^{i-1} B_1^{i-1} E} \rho_{R_{i-1} A_1^{i-1} X_1^{i-1} B_1^{i-1} E} T_{X_1^{i-1} A_1^{i-1} B_1^{i-1} E}^\dagger \cdot \]

with $T_{X_1^{i-1} A_1^{i-1} B_1^{i-1} E} = (v^j_{X_1^{i-1} A_1^{i-1} B_1^{i-1} E})^\frac{1}{2} (\rho_{X_1^{i-1} A_1^{i-1} B_1^{i-1} E})^{-\frac{1}{2}}$. Finally, we get
\[ H_\alpha (A^i D^i | B^i E \tilde{D}^i) \tilde{P}_t \]
\[ = \sum_i H_\alpha (A_i D_i | B^i E \tilde{D}_i A_1^{i-1} X_1^{i-1}) \tau_i (D_i \circ M_i)(\omega^j_i) \]
\[ \geq \sum_i \omega_{R_{i-1} A_1^{i-1} B_1^{i-1} E} H_\alpha (A_i D_i | B^i E \tilde{D}_i A_1^{i-1}) \tau_i (D_i \circ M_i)(\omega^j_i) \cdot \]

where in the inequality we used the fact that $X_1^{i-1}$ is classical together with Lemma B.3. We point out that it is clear from this calculation that if part of the systems $A_1^{i-1} B_1^{i-1} E$ is classical in $\rho$, it remains classical in $\omega^j$. This proves the claims in Remark 4.3.

Considering the right hand side of expression (40), we get for any such state $\omega_{R_{i-1} R}$,
\[ H(A_i D_i | B_i \tilde{D}_i R)(D_i \circ M_i)(\omega) = H(A_i X_i D_i | B_i \tilde{D}_i R)(D_i \circ M_i)(\omega) \]
\[ = H(A_i X_i | B_i R) M_i (\omega) + H(D_i | \tilde{D}_i X_i)(D_i \circ M_i)(\omega) \]
\[ \geq H(A_i | B_i R) M_i (\omega) + \sum_x q(x) H_\alpha (D_i | \tilde{D}_i)(\omega) \]
\[ = H(A_i | B_i R) M_i (\omega) + \sum_x q(x) (\tilde{g} - f(\delta_x)) \]
\[ \geq \tilde{g} \]
where \( q = \mathcal{M}_i(\omega) X_i \) denotes the distribution of \( X_i \) on \( \mathcal{X} \) obtained from the state \( \mathcal{M}_i(\omega) \). The third equality comes from the fact that \( X_i \) is determined by \( A_i B_i \). The first inequality follows from the monotonicity of the Rényi entropies in \( \alpha \) \cite{8,37}. The last equality holds because \( f \) is affine and the final inequality because \( f \) is a min-tradeoff function. Putting everything together, Eq. (37) becomes

\[
H_{\alpha}^{\uparrow}(A_1^n|B_1^n E)_{\rho|\Omega} > nh - n \left( \frac{\alpha - 1}{4} \right) V^2 - \frac{\alpha - 1}{\alpha} \log \frac{1}{\rho|\Omega} .
\]

This concludes the proof of the first inequality (31) of Proposition 4.5.

In order to show the second inequality (32), using the same argument as before, we obtain

\[
H_{\alpha}^{\downarrow}(A_1^n D_i^n|B_1^n E\tilde{D}_i^n)_{\bar{\rho}} \leq \sum_{i} \sup_{\omega_{R_i-1}} H(A_i D_i|B_i \tilde{D}_i R)_{(D_i \circ M_i)(\omega)} + n \left( \frac{\alpha - 1}{4} \right) V^2 ,
\]

where the supremum is over all states \( \omega_{R_i-1} \) with \( R \) constrained as described by Remark 4.3. For any such state and a max-tradeoff function \( f \), we have

\[
H(A_i D_i|B_i \tilde{D}_i R)_{(D_i \circ M_i)(\omega)} \leq H(A_i|B_i R)_{M_i(\omega)} + \sum_{x} q(x) H_{\alpha}^{\downarrow}(D_i|\tilde{D}_i)_{\tau(x)}
\]

\[
= H(A_i|B_i R)_{M_i(\omega)} + \sum_{x} q(x)(\bar{g} - f(\delta_x))
\]

\[
\leq \bar{g} .
\]

It then suffices to combine these inequalities with inequality (38). \( \Box \)

We now prove the claim used in the preceding proof.

**Claim 4.6.** For \( \alpha \in (1,2) \), \( \rho \) and \( \Omega \) as in the statement of Proposition 4.5 and \( \bar{\rho} \) as defined in (34) (see also the preceding text for a definition of \( \bar{g} \)), we have

\[
H_{\alpha}^{\uparrow}(A_1^n|B_1^n E)_{\rho|\Omega} \geq H_{\alpha}^{\uparrow}(A_1^n D_i^n|B_1^n E\tilde{D}_i^n)_{\bar{\rho}|\Omega} - n \bar{g} + nh , \quad (41)
\]

\[
H_{\alpha}^{\downarrow}(A_1^n|B_1^n E)_{\rho|\Omega} \leq H_{\alpha}^{\downarrow}(A_1^n D_i^n|B_1^n E\tilde{D}_i^n)_{\bar{\rho}|\Omega} - n \bar{g} + nh . \quad (42)
\]

**Proof.** We focus on proving inequality (41). The first step is to show that as \( X_1^n \) is a deterministic function of \( A_1^n B_1^n \), we have

\[
H_{\alpha}^{\downarrow}(A_1^n D_i^n|B_1^n E\tilde{D}_i^n)_{\bar{\rho}|\Omega} = H_{\alpha}^{\downarrow}(A_1^n X_1^n D_i^n|B_1^n E\tilde{D}_i^n)_{\bar{\rho}|\Omega} . \quad (43)
\]

In order to do that, observe that for any \( x_1^n \in \mathcal{X}^n \), we have

\[
\bar{\rho}_{A_1^n B_1^n E D_i^n \tilde{D}_i^n, x_1^n} = \rho_{A_1^n B_1^n E, x_1^n} \otimes \tau(x_1^n)_{D_i^n \tilde{D}_i^n} .
\]

where we introduced the notation \( \tau(x_1^n)_{D_i^n \tilde{D}_i^n} = \tau(x_1)_{D_1 \tilde{D}_1} \otimes \cdots \otimes \tau(x_n)_{D_n \tilde{D}_n} . \) This implies that for any \( x_1^n \), we have

\[
\bar{\rho}_{X_1^n A_1^n B_1^n E D_i^n \tilde{D}_i^n, x_1^n} = (T_n \circ \cdots \circ T_1)(\bar{\rho}_{A_1^n B_1^n E D_i^n \tilde{D}_i^n, x_1^n}) .
\]
By taking the sum over $\chi^n_1 \in \Omega$ and then normalising by $\rho[\Omega]$, we get

$$\tilde{\rho}_{X^n_1 A^n_1 B^n_1 E} D^n_1 D^n_1 |\Omega \rangle \otimes \rho A^n_1 B^n_1 E \otimes D^n_1 |\Omega \rangle = (T_1 \circ \cdots \circ T_n)(\tilde{\rho}_{A^n_1 B^n_1 E} D^n_1 D^n_1 |\Omega \rangle) .$$

Thus, we can apply Lemma B.7 and prove the equality (43).

Let now $\sigma_{B^n_1 E} D^n_1$ be a state such that

$$H^\dagger_\alpha(A^n_1 X^n_1 D^n_1 |B^n_1 E D^n_1) \tilde{\rho}_{\Omega} = -D_\alpha(\tilde{\rho}_{A^n_1 X^n_1 D^n_1 B^n_1 E} D^n_1 |\Omega \rangle \otimes \sigma_{B^n_1 E} D^n_1) .$$

Let furthermore $S = S_{D \tilde{D}}$ be the TPCP map that applies a random (according to the Haar measure) unitary to $D$ and its conjugate to $\tilde{D}$ (in such a way that the maximally entangled state on $D \tilde{D}$ used to define $\tau(x)$ is preserved). It is then easy to see that the map $S^{\otimes n}$ applied to the $n$ pairs $D_i \tilde{D}_i$ leaves $\tilde{\rho}_{\Omega}$ invariant. Hence, by the data processing inequality

$$H^\dagger_\alpha(A^n_1 X^n_1 D^n_1 |B^n_1 E D^n_1) \tilde{\rho}_{\Omega} \leq -D_\alpha(S^{\otimes n}(\tilde{\rho}_{A^n_1 X^n_1 D^n_1 B^n_1 E} D^n_1 |\Omega \rangle \otimes S_{B^n_1 E} D^n_1))$$

$$= -D_\alpha(\tilde{\rho}_{A^n_1 X^n_1 D^n_1 B^n_1 E} D^n_1 |\Omega \rangle \otimes \tilde{\sigma}_{B^n_1 E} D^n_1) ,$$

where $\tilde{\sigma}_{B^n_1 E} D^n_1 = \sigma_{B^n_1 E} \otimes \tilde{\rho}_{D^n_1}$. Lemma 3.1 then implies that

$$H^\dagger_\alpha(A^n_1 X^n_1 D^n_1 |B^n_1 E D^n_1) \tilde{\rho}_{\Omega} \leq H^\dagger_\alpha(A^n_1 X^n_1 |B^n_1 E D^n_1) \tilde{\rho}_{\Omega} + H_\alpha(D^n_1 A^n_1 X^n_1 B^n_1 E \tilde{D}^n_1) \nu$$

(44)

where $\nu$ is a state defined by

$$\nu_{A^n_1 X^n_1 B^n_1 E} \tilde{D}^n_1 = \left(\frac{1}{\tilde{\rho}[\Omega]} \sum_{x^n_1 \in \Omega} x^n_1 | x^n_1 \rangle \otimes \sigma_{A^n_1 X^n_1 B^n_1 E} \otimes \tilde{\rho}_{\Omega} \right)^\alpha \quad \text{and}$$

$$\nu_{A^n_1 X^n_1 B^n_1 E} D^n_1 \tilde{D}^n_1 = \nu_{A^n_1 X^n_1 B^n_1 E} \tilde{D}^n_1 \left(\frac{1}{\tilde{\rho}[\Omega]} \sum_{x^n_1 \in \Omega} x^n_1 | x^n_1 \rangle \otimes \sigma_{A^n_1 X^n_1 B^n_1 E} \otimes \tilde{\rho}_{\Omega} \right)^\alpha .$$

We now use properties of $\rho_{\Omega}$ and $\tilde{\sigma}$ to simplify the expression of $\nu$. Observing that

$$\tilde{\rho}_{X^n_1 A^n_1 B^n_1 E} D^n_1 |\Omega \rangle = \frac{1}{\rho[\Omega]} \sum_{x^n_1 \in \Omega} x^n_1 | x^n_1 \rangle \otimes \rho_{A^n_1 B^n_1 E} \otimes \tilde{\rho}_{\Omega} ,$$

(45)

we can write

$$\nu_{A^n_1 X^n_1 B^n_1 E} \tilde{D}^n_1 = \sum_{x^n_1 \in \Omega} x^n_1 | x^n_1 \rangle \otimes \nu_{A^n_1 B^n_1 E} \otimes \tilde{\rho}_{\Omega} ,$$

with

$$\nu_{A^n_1 B^n_1 E} \chi^n_1 = \frac{1}{\rho[\Omega]} \left(\frac{1}{\tilde{\rho}[\Omega]} \sum_{x^n_1 \in \Omega} x^n_1 | x^n_1 \rangle \otimes \rho_{A^n_1 B^n_1 E} \chi^n_1 \otimes \tilde{\rho}_{\Omega} \right)^\alpha .$$

In addition, as $\tilde{\rho}_{\Omega}$ is of the form

$$\tilde{\rho}_{A^n_1 X^n_1 B^n_1 E} D^n_1 \tilde{D}^n_1 |\Omega \rangle = \frac{1}{\rho[\Omega]} \sum_{x^n_1 \in \Omega} x^n_1 \otimes \sigma_{A^n_1 X^n_1 B^n_1 E} \otimes \tau(x^n_1) \tilde{D}^n_1 \tilde{D}^n_1 .$$
we have

\[ \tilde{\rho}_{D^n_1} = \sum_{x^n_1 \in \Omega} |x^n_1 \rangle \langle x^n_1 | \otimes \rho_{A^n_1 B^n_1 E, x^n_1}^0 \otimes \tau(x^n_1) D^n_1 | \tilde{D}_1^n \] ,

where \( \rho_{A^n_1 B^n_1 E, x^n_1}^0 \) is the projector onto the support of \( \rho_{A^n_1 B^n_1 E, x^n_1} \). Hence,

\[ \nu_{A^n_1 X^n_1 E} = \sum_{x^n_1 \in \Omega} |x^n_1 \rangle \langle x^n_1 | \otimes \nu_{A^n_1 B^n_1 E, x^n_1} \otimes \tau(x^n_1) D^n_1 | \tilde{D}_1^n \] .

(46)

Getting back to the inequality (44), we have \( H_{\alpha}(A^n_1 X^n_1 | B^n_1 E \tilde{D}_1^n)_{\rho|\Omega} = H_{\alpha}(A^n_1 | B^n_1 E)_{\tilde{\rho}|\Omega} \) using Eq. (45) to drop \( \tilde{D}_1^n \) and Lemma B.7 to drop \( X^n_1 \). Moreover, using (46), we have that \( H_{\alpha}(D^n_1 | A^n_1 X^n_1 B^n_1 E \tilde{D}_1^n)_{\nu} = H_{\alpha}(D^n_1 | X^n_1 \tilde{D}_1^n)_{\nu} \). Finally, we get

\[ H_{\alpha}^+(A^n_1 D^n_1 | B^n_1 E \tilde{D}_1^n)_{\tilde{\rho}|\Omega} \leq H_{\alpha}^+(A^n_1 | B^n_1 E)_{\tilde{\rho}|\Omega} + H_{\alpha}(D^n_1 | X^n_1 \tilde{D}_1^n)_{\nu} \] .

(47)

It is a direct consequence of the definition of \( \tau(x) \) that

\[ H_{\alpha}(D^n_1 | \tilde{D}_1^n)_{\tau(x^n_1)} = n\tilde{g} - \sum_{i=1}^n f(\delta x_i) \]

\[ = n\tilde{g} - n \sum_{x \in \mathcal{X}} \text{freq}(x^n_1)(x) f(\delta x) \]

\[ = n\tilde{g} - n \int_{x \in \mathcal{X}} \text{freq}(x^n_1)(x) \delta x \]

\[ = n\tilde{g} - n f \left( \sum_{x \in \mathcal{X}} \text{freq}(x^n_1)(x) \delta x \right) = n\tilde{g} - n f(\text{freq}(x^n_1)) \] ,

where we have used that \( f \) is an affine function. Using Lemma B.3 and (46) we can bound the second term on the right hand side of (47) by

\[ H_{\alpha}(D^n_1 | \tilde{D}_1^n X^n_1)_{\nu} \leq \max_{x^n_1 \in \Omega} H_{\alpha}(D^n_1 | \tilde{D}_1^n)_{\tau(x^n_1)} \]

\[ \leq \max_{x^n_1 : f(\text{freq}(x^n_1)) \geq h} n\tilde{g} - n f(\text{freq}(x^n_1)) \leq n\tilde{g} - nh \] .

Inserting this in (47) gives

\[ H_{\alpha}^+(A^n_1 | B^n_1 E)_{\tilde{\rho}|\Omega} \geq H_{\alpha}^+(A^n_1 D^n_1 | B^n_1 E \tilde{D}_1^n)_{\tilde{\rho}|\Omega} - n\tilde{g} + nh \] .

This concludes the proof of inequality (41). For the proof of inequality (42), we can follow similar steps.

In order to prove inequality (42). The first step is to show that as \( X^n_1 \) is a deterministic function of \( A^n_1 B^n_1 \), we have

\[ H_{\alpha}(A^n_1 D^n_1 | B^n_1 E \tilde{D}_1^n)_{\tilde{\rho}|\Omega} = H_{\alpha}(A^n_1 X^n_1 D^n_1 | B^n_1 E \tilde{D}_1^n)_{\tilde{\rho}|\Omega} \] .

(48)

In order to do that, observe that for any \( x^n_1 \in \mathcal{X} \), we have

\[ \tilde{\rho}_{A^n_1 B^n_1 E D^n_1} = \rho_{A^n_1 B^n_1 E, x^n_1} \otimes \tau(x^n_1) D^n_1 | \tilde{D}_1^n \] .
where we introduced the notation $\tau(x^n_1)_{D^n_1D^n_1} = \tau(x_1)_{D_1D_1} \otimes \cdots \otimes \tau(x_n)_{D_nD_n}$. This implies that for any $x^n_i$, we have

$$\tilde{\rho}_{X^n_i A^n_i B^n_i ED^n_i D^n_i, x^n_i} = (T_n \circ \cdots \circ T_1)(\tilde{\rho}_{A^n_i B^n_i ED^n_i D^n_i, x^n_i}).$$

By taking the sum over $x^n_i \in \Omega$ and then normalising by $\rho[\Omega]$, we get

$$\tilde{\rho}_{X^n_i A^n_i B^n_i ED^n_i D^n_i | \Omega} = (T_n \circ \cdots \circ T_1)(\tilde{\rho}_{A^n_i B^n_i ED^n_i D^n_i | \Omega}).$$

Thus, we can apply Lemma B.7 and prove the equality (48). Theorem 3.2 then implies that

$$H_\alpha(A^n_1 X^n_1 D^n_1 | B^n_1 E D^n_1)_{\rho[\Omega]} = H_\alpha(A^n_1 X^n_1 | B^n_1 E D^n_1)_{\rho[\Omega]} + H_\alpha(D^n_1 | A^n_1 X^n_1 B^n_1 E D^n_1)_\nu \quad \text{(49)}$$

where $\nu$ is a state defined by

$$v_{A^n_1 X^n_1 B^n_1 E D^n_1} = \frac{1}{\rho[\Omega]} \sum_{x^n_i \in \Omega} |x^n_i\rangle\langle x^n_i| \otimes \rho_{A^n_1 B^n_1 E, x^n_i} \otimes \tilde{\rho}_{D^n_1},$$

(50)

we can write

$$v_{A^n_1 X^n_1 B^n_1 E D^n_1} = \sum_{x^n_i \in \Omega} |x^n_i\rangle\langle x^n_i| \otimes v_{A^n_1 B^n_1 E, x^n_i} \otimes \tilde{\rho}_{D^n_1},$$

with $v_{A^n_1 B^n_1 E, x^n_i} = \frac{1}{\rho[\Omega]^\alpha} \left( \frac{\rho_{A^n_1 E, x^n_i}^{\frac{1}{\alpha}} \rho_{B^n_1 E}^{\frac{1}{\alpha}} \rho_{A^n_1 B^n_1 E, x^n_i}^{\frac{1}{\alpha}}}{\text{tr} \left( \tilde{\rho}_{A^n_1 X^n_1 B^n_1 E D^n_1 | \Omega} \rho_{A^n_1 E}^{\frac{1}{\alpha}} \rho_{B^n_1 E}^{\frac{1}{\alpha}} \rho_{A^n_1 B^n_1 E, x^n_i}^{\frac{1}{\alpha}} \right)} \right)^\alpha$.

In addition, as $\tilde{\rho}_{\Omega}$ is of the form

$$\tilde{\rho}_{A^n_1 X^n_1 B^n_1 E D^n_1 | \Omega} = \frac{1}{\rho[\Omega]} \sum_{x^n_i \in \Omega} |x^n_i\rangle\langle x^n_i| \otimes \rho_{A^n_1 B^n_1 E, x^n_i} \otimes \tau(x^n_i)_{D^n_1 D^n_1},$$

we have

$$\tilde{\rho}_{D^n_1 | A^n_1 X^n_1 B^n_1 E D^n_1 | \Omega} = \sum_{x^n_i \in \Omega} |x^n_i\rangle\langle x^n_i| \otimes \rho^{0}_{A^n_1 B^n_1 E, x^n_i} \otimes \tau(x^n_i)_{D^n_1 | D^n_1},$$

where $\rho^{0}_{A^n_1 B^n_1 E, x^n_i}$ is the projector onto the support of $\rho_{A^n_1 B^n_1 E, x^n_i}$. Hence,

$$v_{A^n_1 X^n_1 B^n_1 E D^n_1} = \sum_{x^n_i \in \mathcal{X}^n} |x^n_i\rangle\langle x^n_i| \otimes v_{A^n_1 B^n_1 E, x^n_i} \otimes \tau(x^n_i)_{D^n_1 | D^n_1}. \quad \text{(51)}$$
Getting back to the inequality (49), we have $H_\alpha(A^n_1 | X^n_1 | B^n_1 E \tilde{D}^n_1)_{\tilde{\rho}, \Omega} = H_\alpha(A^n_1 | B^n_1 E)_{\tilde{\rho}, \Omega}$ using Eq. (50) to drop $\tilde{D}^n_1$ and Lemma B.7 to drop $X^n_1$. Moreover, using (51), we have that $H_\alpha(D^n_1 | A^n_1 X^n_1 B^n_1 E \tilde{D}^n_1)_v = H_\alpha(D^n_1 | X^n_1 \tilde{D}^n_1)_v$. Finally, we get

$$H_\alpha(A^n_1 D^n_1 | B^n_1 E \tilde{D}^n_1)_{\tilde{\rho}, \Omega} = H_\alpha(A^n_1 | B^n_1 E)_{\tilde{\rho}, \Omega} + H_\alpha(D^n_1 | X^n_1 \tilde{D}^n_1)_v. \quad (52)$$

It is a direct consequence of the definition of $\tau(x)$ that

$$H_\alpha(D^n_1 | \tilde{D}^n_1)_{\tau(x^n_1)} = n \tilde{g} - \sum_{i=1}^{n} f(\delta_{x_i}) = n \tilde{g} - n \sum_{x \in \mathcal{X}} \text{freq}(x^n_1)(x) f(\delta_x) = n \tilde{g} - n f \left( \sum_{x \in \mathcal{X}} \text{freq}(x^n_1)(x) \delta_x \right) = n \tilde{g} - n f(\text{freq}(x^n_1)),$$

where we have used that $f$ is an affine function. Using Lemma B.3 and (51) we can bound the second term on the right hand side of (52) by

$$H_\alpha(D^n_1 | \tilde{D}^n_1 X^n_1)_v \geq \min_{x^n_1 \in \Omega} H_\alpha(D^n_1 | \tilde{D}^n_1)_{\tau(x^n_1)} \geq \min_{x^n_1 : f(\text{freq}(x^n_1)) \leq h} n \tilde{g} - n f(\text{freq}(x^n_1)) \geq n \tilde{g} - nh.$$

Inserting this in (52) and replacing $\alpha$ with $\frac{1}{\alpha}$ gives

$$H_{\frac{1}{\alpha}}(A^n_1 | B^n_1 E)_{\rho, \Omega} \leq H_{\frac{1}{\alpha}}(A^n_1 D^n_1 | B^n_1 E \tilde{D}^n_1)_{\tilde{\rho}, \Omega} - n \tilde{g} + nh.$$

This concludes the proof of inequality (42). \qed

Finally, we prove Theorem 4.4 using Proposition 4.5.

**Proof of Theorem 4.4.** The first step is to use Lemma B.10 to lower-bound the smooth min-entropy by a Rényi entropy:

$$H_{\alpha}^{\min}(A^n_1 | B^n_1 E)_{\rho, \Omega} \geq H_{\alpha}^{\uparrow}(A^n_1 | B^n_1 E)_{\rho, \Omega} - \frac{g(\epsilon)}{\alpha - 1}. \quad (53)$$

Then Proposition 4.5 yields

$$H_{\alpha}^{\min}(A^n_1 | B^n_1 E)_{\rho, \Omega} > nh - n \left( \frac{\alpha - 1}{4} \right) V^2 - \frac{1}{\alpha'} \log \frac{1}{\rho[\Omega]} - \frac{g(\epsilon)}{\alpha - 1}$$

$$> nh - n \left( \frac{\alpha - 1}{4} \right) V^2 - \frac{1}{\alpha'} \log \frac{1}{\rho[\Omega]} - \log(2/\epsilon^2)$$

$$\geq nh - n \left( \frac{\alpha - 1}{4} \right) V^2 - \frac{1}{(\alpha - 1)} \log \frac{2}{\rho[\Omega]^2 \epsilon^2},$$

where we have used the fact that we are constrained to choose $\alpha \leq 1 + \frac{2}{V} \leq 2$ in the last inequality. We now choose

$$\alpha := 1 + 2 \sqrt{\log \frac{2}{\rho[\Omega]^2 \epsilon^2}} \sqrt{n V} \quad (54)$$
and note that, as long as

$$n > \log \frac{2}{\rho[\Omega]^2\varepsilon^2},$$

(55)

the value $\alpha$ is strictly smaller than $1 + \frac{2}{\varepsilon}$ and therefore within the required bounds. Note also that if (55) does not hold then the term $c\sqrt{n}$ in the claim (29) is at least $nV \geq 2n \log(1 + 2d_A) \geq 2n \log d_A$, whereas the min-entropy is always at least $-n \log d_A$ and $nf_{\min}(q)$ is at most $n \log d_A$, which means that the claim is trivial. Finally, inserting (54) into the above yields

$$H_{\min}^\varepsilon(A^n_1|B^n_1 E)_{\rho} > nh - \sqrt{nV} \sqrt{\log \frac{2}{\rho[\Omega]^2\varepsilon^2}},$$

as advertised. Once again, the max-entropy statement (30) holds by switching the direction of the inequalities, flipping the appropriate signs, and replacing every occurrence of $H_{\alpha}$ by $H_{\frac{1}{\alpha}}$. □

It might seem restrictive to assume that the tradeoff function is affine. We next show that we may take a general convex function provided the event $\Omega$ can be described as follows: $x^n \in \Omega$ if and only if $\text{freq}(x^n) \in \hat{\Omega}$, where $\hat{\Omega}$ is a convex subset of $\mathbb{P}$.

**Corollary 4.7.** Let $\mathcal{M}_1, \ldots, \mathcal{M}_n$ and $\rho A^n_1 B^n_1 X^n_1 E$ be such that (26) and the Markov conditions (27) hold, let $h \in \mathbb{R}, \varepsilon \in (0, 1)$, let $\hat{\Omega}$ be a convex set $\hat{\Omega} \subseteq \mathbb{P}$ and define the corresponding event $\Omega \subseteq X^n$ by $x^n_1 \in \Omega \iff \text{freq}(x^n_1) \in \hat{\Omega}$. Then, if $f$ is a differentiable and convex min-tradeoff function for $\mathcal{M}_1, \ldots, \mathcal{M}_n$ satisfying $f(q) \geq h$ for all $q \in \hat{\Omega}$, we have

$$H_{\min}^\varepsilon(A^n_1|B^n_1 E)_{\rho|\Omega} > nh - c\sqrt{n}$$

(56)

where $c = 2(\log(1 + 2d_A) + \|\nabla f\|_{\infty})\sqrt{1 - \frac{1}{2\varepsilon^2\rho[\Omega]}}$. Similarly, if $f$ is a differentiable and concave max-tradeoff function for $\mathcal{M}_1, \ldots, \mathcal{M}_n$ satisfying $f(q) \leq h$ for all $q \in \hat{\Omega}$, we have

$$H_{\max}^\varepsilon(A^n_1|B^n_1 E)_{\rho|\Omega} < nh + c\sqrt{n}.$$  

(57)

**Proof.** Let us denote by $\text{cl}(\hat{\Omega})$ the closure of the set $\hat{\Omega}$. Now as $f$ is continuous on the compact set $\text{cl}(\hat{\Omega})$ (it is even assumed to be differentiable on all of $\mathbb{P}$), we have $\min_{q \in \text{cl}(\hat{\Omega})} f(q) = f(q_0)$ for some $q_0 \in \text{cl}(\hat{\Omega})$. By continuity of $f$ and by definition of $h$, we have $f(q_0) \geq h$. Now consider the affine function $g(q) = (\nabla f)_{q_0} \cdot (q - q_0) + f(q_0)$.

By convexity of $f$, we have that $g(q) \leq f(q)$ for all $q \in \mathbb{P}$ and thus $g$ is a min-tradeoff function. In addition, as $\text{cl}(\hat{\Omega})$ is convex we can apply the first order optimality conditions and get that $(\nabla f)_{q_0} \cdot (q - q_0) \geq 0$ for all $q \in \text{cl}(\hat{\Omega})$. As a result, for all $q \in \text{cl}(\hat{\Omega})$, we have $g(q) \geq f(q_0) \geq h$. This implies that if $x^n_1 \in \Omega$, then $g(\text{freq}(x^n)) \geq h$. We can then apply Theorem 4.4 with the affine tradeoff function $g$ and get the desired result as $\|\nabla g\|_{\infty} \leq \|\nabla f\|_{\infty}$.

The proof for $H_{\max}$ is analogous. □

One natural choice for the event $\Omega$ is that the empirical distribution $\text{freq}(X^n_1)$ takes a particular value $q$. This yields the following special case of Corollary 4.7.
**Corollary 4.8.** Let $\mathcal{M}_1, \ldots, \mathcal{M}_n$ and $\rho_{A_i^n|B_i^n|X_i^n}^q$ be such that (26) and the Markov conditions (27) hold. Then, for any differentiable and convex min-tradeoff function $f$ for $\mathcal{M}_1, \ldots, \mathcal{M}_n$ and for any $q \in \mathbb{P}$, we have

$$H_{\min}^f(A_1^n|B_1^n)_{\rho} > n f(q) - c \sqrt{n}$$

where $c = 2(\log(1 + 2d_A) + \|\nabla f(q)\|_{\infty}) \sqrt{1 - 2 \log(\epsilon p[q])}$, where $\rho_{q}$ denotes the state $\rho$ conditioned on the event that $\text{freq}(X_i^n) = q$, and $\rho[q]$ the probability of this event.

Note that an analogous statement holds of course for the max-entropy, replacing $f$ by a concave max-tradeoff function and changing the inequality accordingly.

The following corollary specialises the above to the formulation (3), in which no statistical test is being done, i.e. the $X_i$ systems are trivial. We provide the statement for the case of the lower boundary.

**Corollary 4.9.** Let $\mathcal{M}_1, \ldots, \mathcal{M}_n$ and $\rho_{A_i^n|B_i^n|X_i^n}^q$ be such that (26) and the Markov conditions (27) hold. Then

$$H_{\min}^f(A_1^n|B_1^n)_{\rho} > \sum_{i=1}^{n} \inf_{\omega R_{i-1} R} H(A_i|B_i R)_{(\mathcal{M}_i \otimes \mathcal{I}_R)(\omega R_{i-1} R)} - c \sqrt{n}$$

where $c = 3(\log(1 + 2d_A)) \sqrt{1 - 2 \log(\epsilon)}$.

**Proof.** Note that the quantity $H_{\min}^f(A_1^n|B_1^n)_{\rho}$ only depends on the marginal of the state $\rho$ on $A_i^n|B_i^n$. Thus, we can modify the maps $\mathcal{M}_i$ in any way that does not affect the reduced state $\rho_{A_i^n|B_i^n}$ before applying Corollary 4.8. In particular, we change $\mathcal{M}_i$ so that the original value of $X_i$ is disregarded and replaced with the constant value $X_i = i$. The values $X_1, \ldots, X_n$ can then be regarded as random variables with alphabet $\mathcal{X} = \{1, \ldots, n\}$. We define the real function $f$ on $\mathbb{P}$ as

$$f(q) = \sum_{i=1}^{n} q(i) \inf_{\omega R_{i-1} R} H(A_i|B_i R)_{(\mathcal{M}_i \otimes \mathcal{I}_R)(\omega R_{i-1} R)} .$$

Note that for any $i \in \{1, \ldots, n\}$ and any $q \in \mathbb{P}$, we have either $q(i) \neq 1$ in which case $\Sigma_i(q) = \emptyset$ (we use the notation in (28)) and the min-tradeoff condition is trivial or $q(i) = 1$, in which case $\Sigma_i(q) = \{(\mathcal{M}_i \otimes \mathcal{I}_R)(\omega R_{i-1} R) : \omega R_{i-1} R \in \mathcal{D}(R_{i-1} \otimes R)\}$. Thus for any $q \in \mathbb{P}$,

$$f(q) \leq \inf_{\omega R_{i-1} R} H(A_i|B_i R)_{(\mathcal{M}_i \otimes \mathcal{I}_R)(\omega R_{i-1} R)} .$$

As a result, $f$ is a min-tradeoff function for all $\mathcal{M}_i$ for $i \in \{1, \ldots, n\}$. We now fix $q \in \mathbb{P}$ such that $q(1) = \cdots = q(n) = \frac{1}{n}$, in which case the event $\text{freq}(X_i^n) = q$ occurs with certainty. Because

$$\|\nabla f(q)\|_{\infty} \leq \log d_A$$

which implies that $\|\nabla f(q)\|_{\infty} \leq \log(1 + 2d_A)$, the claim follows immediately from Corollary 4.8. $\square$
As indicated in the introduction, in the special case where the individual pairs \((A_i, B_i)\) are independent and identically distributed (IID), the entropy accumulation theorem corresponds to the Quantum Asymptotic Equipartition Property [54]. We can therefore formulate the latter as a corollary of Theorem 4.4.14

**Corollary 4.10.** For any bipartite state \(\nu_{AB}\), any \(n \in \mathbb{N}\), and any \(\varepsilon \in (0, 1)\),

\[
\frac{1}{n} H^\varepsilon_{\min}(A^n_1 | B^n_1)_{\nu \otimes^n} > H(A | B)_\nu - 2\sqrt{\frac{1 - 2 \log \varepsilon}{n}} \log(1 + 2d_A).
\]

**Proof.** Let, for any \(i = 1, \ldots, n\), \(\mathcal{M}_i\) be the TPCP map from \(R\) to \(XABR\) which sets \(AB\) to state \(\nu_{AB}\) and where \(X\) and \(R\) are trivial (one-dimensional) systems. The concatenation of these maps thus generates the state \(\rho_{A^nB^n} = \nu_{AB} \otimes^n\). The claim is then obtained from Theorem 4.4 with the trade-off function \(f\) being a constant equal to \(h = H(A | B)_\nu\) and with \(\Omega\) as the certain event. \(\square\)

5. Applications

Entropy is a rather general notion and, accordingly, entropy accumulation has applications in various areas of physics, information theory, and computer science. An example from physics is the phenomena of thermalisation. It is known that a system can only thermalise if its smooth min-entropy is sufficiently large [18]. To illustrate how Theorem 4.4 could give an estimate of this entropy, consider a system of interest (e.g., a cup of coffee) which is in contact with a large environment (the air around it). Suppose that, for an appropriately chosen discretisation of the evolution, the system interacts at each time step with a different part of the environment (e.g., with different air molecules bouncing off the coffee cup). Theorem 4.4 then provides a bound on the total entropy that is transferred to the environment in terms of the von Neumann entropy transferred in each time step. Because the joint time evolution of system and environment is unitary, this entropy flow to the environment could be expressed in terms of the entropy change of the system itself. The argument would therefore prove that the total entropy acquired by the system over many time steps is bounded by the sum of the entropies produced in each individual time step.

Another area where the notion of entropy plays a crucial role is quantum cryptography. Many proofs of security of cryptographic protocols involve lower-bounding the uncertainty that a dishonest adversary has about some system of interest. The state-of-the-art is to derive such bounds using a combination of de Finetti-type theorems as well as the Quantum Asymptotic Equipartition Property [4,15,41,42]. However, the use of de Finetti theorems comes with various disadvantages. Firstly, they are only applicable under certain assumptions on the symmetry of the protocols. Secondly, they introduce additional error terms that can be large in the practically relevant finite-size regime [47]. Finally, it is not known how to apply de Finetti theorems in a device-independent scenario (see [21] for an overview and references on device-independent cryptography). These problems can all be circumvented by the use of entropy accumulation, as demonstrated in [5] for the case of device-independent quantum key distribution and randomness expansion. The resulting security statements are valid against general attacks and essentially optimal in the finite-size regime.

---

14 In the version of [54], the term \(1 + 2d_A\) in the logarithm is replaced by an expression that depends on the entropy of \(A\) conditioned on \(B\).

15 In statistical mechanics this is formally described by the Repeated Interaction System (RIS) model [7,14].
In the remainder of this section, we illustrate the use of entropy accumulation with two concrete examples. The first is a security proof for a basic quantum key distribution protocol. The second is a novel derivation of an upper bound on the fidelity of fully quantum random access codes.

5.1. Sample application: security of quantum key distribution. A Quantum Key Distribution (QKD) protocol enables two parties, Alice and Bob, to establish a common secret key, i.e., a string of random bits unknown to a potential eavesdropper, Eve. The setting is such that Alice and Bob can communicate over a quantum channel, which may however be fully controlled by Eve. In addition, Alice and Bob have a classical communication link which is assumed to be authenticated, i.e., Eve may read but cannot alter the classical messages exchanged between Alice and Bob. The protocol is said to be secure against general attacks if any attack by Eve is either detected (in which case the protocol aborts) or does not compromise the secrecy of the final key. Here, we will show that our main theorem can be directly applied to show security against general attacks for a fairly standard QKD protocol. As a bonus, our proof still holds even if we do not make any assumptions about Bob’s measurement device: the POVM applied by Bob at every step of the protocol can be arbitrary, and may vary from one step to the next (thereby achieving one-sided measurement device independence as in [58], but without the restriction to memoryless devices; see also [56]). In fact, as shown in [5], the entropy accumulation theorem can be used to prove the security of fully device-independent quantum key distribution.

For concreteness, we consider here a variant of the E91 QKD protocol [22] (and note that any security proof for this protocol also implies security of the BB84 protocol [10, 11]). The protocol consists of a sequence of instructions for Alice and Bob, as described in the box below. These depend on certain parameters, including the number, $n$, of qubits that need to be transmitted over the quantum channel, the maximum tolerated noise level, $e$, of this channel, as well as the key rate, $r$, which is defined as the number of final key bits divided by $n$. In the first protocol step, Alice and Bob need to measure their qubits at random in one of two mutually unbiased bases, which we term the computational and the diagonal basis. These are chosen with probability $1 - \mu$ and $\mu$, respectively, for some $\mu \in (0, 1)$. The protocol also invokes an error correction scheme termed EC, which allows Bob to infer the measurement outcomes obtained by Alice for the set of indices $S$ where the basis choices of Alice and Bob were the same. Note that if the protocol was implemented without any noise, then Bob’s outcomes would match exactly with Alice’s outcomes on the indices $S$ and no error correction would be required. However, in the presence of noise, such an error correction step is needed. For this, Alice needs to send classical error correcting information to Bob, whose maximum relative length is characterised by another parameter, $\vartheta_{EC}$. We assume that EC is reliable. This means that, except with negligible probability, Bob either obtains a correct copy of Alice’s string or he is notified that the string cannot be inferred.\footnote{Any error correction scheme can be turned into a reliable one by appending a test where Alice and Bob compare a hash value computed from their (corrected) strings.}

The security of QKD against general attacks has been established in a sequence of works [13, 32, 33, 41, 49]. Specifically, for the E91 protocol, the following result has been shown.
fying

The E91 protocol is secure for any choice of protocol parameters satisfying

\[ r < 1 - H_{\text{Sh}}(e) - \vartheta_{\text{EC}} - 2\mu , \]

provided that \( n \) is sufficiently large.

Note that, because \( \mu > 0 \) can be chosen arbitrarily small, the theorem implies that the E91 protocol can generate secret keys at an asymptotic rate of \( 1 - H_{\text{Sh}}(e) - \vartheta_{\text{EC}} \). We now show how this result can be obtained using the notion of entropy accumulation.

Proof. According to a standard result on two-universal hashing (see, for instance, Corollary 5.6.1 of [41]), the key \( F(A_S) \) computed in the privacy amplification step is secret to an adversary holding information \( E' \) if the smooth min-entropy of \( A_S \) conditioned on \( E' \) is sufficiently larger than the output size of the hash function \( F \). Since, in our case, this size is \( \lfloor r n \rfloor \), the condition reads

\[ nr \leq H_{\text{min}}^\varepsilon(A_S|E')_{\rho|\Omega} - O(1) , \]

where the entropy is evaluated for the joint state \( \rho|\Omega \) of \( A_S \) and \( E' \) conditioned on the event \( \Omega \) that the protocol is not aborted and that Bob’s guess \( \hat{A}_S \) of \( A_S \) is correct. The smoothing parameter \( \varepsilon \in (0, 1) \) specifies the desired level of secrecy,\(^{18} \) and we assume here that it is constant (independent of \( n \)). Because conditioning the smooth min-entropy of a classical variable on an additional bit cannot decrease its value by more than 1 (see, e.g., Proposition 5.10 of [51]), we may bound the smooth min-entropy in (59) by

\[ H_{\text{min}}^\varepsilon(A_S|E')_{\rho|\Omega} \geq H_{\text{min}}^\varepsilon(A_S|B^n_1 \tilde{B}_1^n E)_{\rho|\Omega} - |S|\vartheta_{\text{EC}} \geq H_{\text{min}}^\varepsilon(A_S|B^n_1 \tilde{B}_1^n E)_{\rho|\Omega} - n\vartheta_{\text{EC}} , \]

\(^{17} H_{\text{Sh}}(e) = -e \log e - (1 - e) \log(1 - e) \) is the binary Shannon entropy.

\(^{18} \)Roughly, \( \varepsilon \) corresponds to the maximum probability by which one could encounter a deviation from perfect secrecy [40].
where $E$ denotes all information held by Eve after the distribution step, and where $|S|\theta_{EC}$ is the maximum number of bits exchanged for error correction. Note that we also included the basis information $B^n_1$ and $\bar{B}^n_1$ in the conditioning part because Eve may obtain this information during the sifting and information reconciliation step. We are thus left with the task of lower bounding $H_{\min}^\varepsilon(AS|B^n_1\bar{B}^n_1E)_{\rho}\Omega_1$, which is usually the central part of any security proof. Since it is also the part where entropy accumulation is used, we formulate it separately as Claim 5.2 below. Inserting this claim into (60), we conclude that the secrecy condition (59) is fulfilled whenever

$$nr \leq n(1 - H_{Sh}(e) - \theta_{EC} - 2\mu) - o(n)$$

holds. But this is clearly the case for any choice of parameters satisfying (58), provided that $n$ is sufficiently large. □

It remains to show the separate claim, which we do using entropy accumulation.

**Claim 5.2.** Let $A^n_1$, $B^n_1$, $\bar{B}^n_1$, and $S$ be the information held by Alice and Bob as defined by the protocol, let $E$ be the information gathered by Eve during the distribution step, and let $\Omega$ be the event that the protocol is not aborted and that Bob’s guess $\hat{A}_S$ of $A_S$ is correct. Then, provided that $\Omega$ has a non-negligible probability (i.e., it does not decrease exponentially fast in $n$),

$$H_{\min}^\varepsilon(AS|B^n_1\bar{B}^n_1E)_{\rho}\Omega_1 > n(1 - 2\mu - H_{Sh}(e)) - o(n).$$

(61)

**Proof.** Let $\rho_{Q^n_1\bar{Q}^n_1E}$ be the joint state of Alice and Bob’s qubit pairs before measurement, together with the information $E$ gathered by Eve during the distribution step, and let

$$\rho_{A^n_1\bar{A}^n_1B^n_1\bar{B}^n_1X^n_1E} = (M_n \circ \cdots \circ M_1 \otimes I_E)(\rho_{Q^n_1\bar{Q}^n_1E}^0).$$

where $M_i$, for any $i \in \{1, \ldots, n\}$, is the TPCP map from $Q^n_i \bar{Q}^n_i$ to $Q^n_{i+1} \bar{Q}^n_{i+1} A_i \bar{A}_i B_i \bar{B}_i X_i$ defined as follows:

(i) $B_i, \bar{B}_i$: random bits chosen independently according to the distribution $(1 - \mu, \mu)$

(ii) $A_i = \begin{cases} \text{outcome of measurement of } Q_i \text{ in computational basis} & \text{if } B_i = \bar{B}_i = 0 \\ \perp & \text{if } B_i \neq \bar{B}_i \end{cases}$

(iii) $\bar{A}_i = \begin{cases} \text{outcome of measurement of } \bar{Q}_i \text{ in diagonal basis} & \text{if } B_i = \bar{B}_i = 1 \\ \perp & \text{otherwise} \end{cases}$

(iv) $X_i = \begin{cases} A_i \oplus \bar{A}_i & \text{if } B_i = \bar{B}_i = 1 \\ \perp & \text{otherwise} \end{cases}$

(v) $Q^n_{i+1}$ and $\bar{Q}^n_{i+1}$ are left untouched.

Note that the values $B^n_1$ and $\bar{B}^n_1$ correspond to the ones generated during the distribution step of the protocol. The same is true for $A^n_1$, with the modification that $A_i$ holds the measurement outcome only if $B_i = \bar{B}_i$. That is, $A_i \neq \perp$ if and only if $i \in S$, where $S$ is the set determined in the sifting step. We can therefore rewrite (61) as

$$H_{\min}^\varepsilon(A^n_1B^n_1E)_{\rho}\Omega_1 > n(1 - 2\mu - H_{Sh}(e)) - o(n).$$

(62)
To prove this inequality, we use Theorem 4.4 with the replacements \( A_i \to A_\perp, B_i \to B_i \tilde{B}_i, X_i \to X_i, \) and \( R_i \to Q_{i+1}^n \tilde{Q}_{i+1}^n. \) We note that \( X_i \) is a deterministic function of the classical registers \( A_i \tilde{A}_i \) and \( B_i \tilde{B}_i. \) To obtain the bound in (62), we need to define a min-tradeoff function. Let \( i \in \{1, \ldots, n\} \) and consider the state
\[
v_{X_i A_i \tilde{A}_i B_i \tilde{B}_i R} = \text{tr}_{Q_{i+1}^n \tilde{Q}_{i+1}^n} (\mathcal{M}_i \otimes \mathcal{I}_R)(\omega Q_i^n \tilde{Q}_i^n R),
\]
where \( \omega Q_i^n \tilde{Q}_i^n R \) is an arbitrary state. Let furthermore \( v_{b} = v_{X_i A_i \tilde{A}_i R | b} \) be the corresponding state obtained by conditioning on the event that \( B_i = \tilde{B}_i = b, \) for \( b \in \{0, 1\}. \) We may now bound the entropy of \( A_i \) using the entropic uncertainty relation proved in [12], which asserts that
\[
H(A_i | R)_{v_0} \geq 1 - H(A_i | \tilde{A}_i)_{v_{\perp 1}}.
\]
By the definition of \( X_i, \) we also have
\[
H(A_i | \tilde{A}_i)_{v_{\perp 1}} = H(X_i)_{v_{\perp 1}} = H_{\text{Sh}} \left( \frac{v_{X_i(1)}}{v_{X_i(0)} + v_{X_i(1)}} \right) = H_{\text{Sh}} \left( \frac{v_{X_i(1)}}{\mu^2} \right),
\]
where we wrote \( v_{X_i} \) to denote the probability distribution on \( \{0, 1, \perp\} \) defined by the state \( v, \) and where we have used that \( v_{X_i(0)} + v_{X_i(1)} = \mu^2. \) Furthermore, because \( A_i \) is classical, its von Neumann entropy cannot be negative, which implies that
\[
H(A_i | B_i \tilde{B}_i R)_{v} \geq v_{B_i \tilde{B}_i (0,0)} H(A_i | R)_{v_0} = (1 - \mu^2)H(A_i | R)_{v_0} \geq H(A_i | R)_{v_0} - 2\mu + \mu^2.
\]
Combining this with the above, we find that
\[
H(A_i \tilde{A}_i | B_i \tilde{B}_i R)_{v} \geq H(A_i | B_i \tilde{B}_i R)_{v} \geq \tilde{f}(v_{X_i})
\]
holds for
\[
\tilde{f}(q) = \begin{cases} 
1 - 2\mu + \mu^2 - H_{\text{Sh}} \left( \frac{q(1)}{\mu^2} \right) & \text{if } q(0) + q(1) = \mu^2 \\
1 & \text{otherwise}.
\end{cases}
\]
In other words, \( \tilde{f} \) is a min-tradeoff function for \( \mathcal{M}_i. \) Furthermore, because the binary Shannon entropy \( H_{\text{Sh}} \) is concave, \( \tilde{f} \) is convex. We may thus define a linearised min-tradeoff function \( f \) as a tangent hyperplane to \( \tilde{f} \) at the point \( q_0 \) given by \( q_0(0) = (1 - e)\mu^2, q_0(1) = e\mu^2, \) and \( q_0(\perp) = 1 - \mu^2. \) Furthermore, we define
\[
h = f(q_0) = \tilde{f}(q_0) = 1 - 2\mu + \mu^2 - H_{\text{Sh}}(e).
\]
Finally, note that the event \( \Omega \) that Bob’s guess of \( A_i \) is correct and that the protocol is not aborted implies that \( q = \text{freq}(X^n_i) \) is such that \( \frac{q(1)}{\mu^2} \leq e \) and, hence, \( f(\text{freq}(X^n_i)) \geq h. \) Since we assumed that \( \Omega \) has non-negligible probability, Theorem 4.4 implies that
\[
\frac{n}{\mu^2} - H_{\text{min}}(A^n_1 A^n_1 | B^n_1 \tilde{B}^n_1 E)_{\rho_\Omega} > nh - o(n) = n(1 - 2\mu + \mu^2 - H_{\text{Sh}}(e)) - o(n).
\]
(Note that the Markov chain conditions are satisfied because \( B_i \) and \( \tilde{B}_i \) are chosen at random independently of any other information.) Furthermore, because \( A_i \) equals \( \perp \) unless \( B_i = \tilde{B}_i = 1, \) which occurs with probability \( \mu^2, \) we have
\[
\frac{n}{\mu^2} H_{\text{max}}(A^n_1 A^n_1 | B^n_1 \tilde{B}^n_1 E)_{\rho_\Omega} \leq H_{\text{max}}(A^n_1 | B^n_1 \tilde{B}^n_1)_{\rho_\Omega} \leq \mu^2 n + o(n).
\]
Combining these inequalities with the chain rule for smooth entropies (see Theorem 15 of [60]),
\[
H_{\min}^{\epsilon}(A_1^n | B_1^n \tilde{B}_1^n E)_{\rho_{\Omega}} \geq H_{\min}^{\epsilon/4}(A_1^n \tilde{A}_1^n | B_1^n \tilde{B}_1^n E)_{\rho_{\Omega}} - H_{\max}(\tilde{A}_1^n | A_1^n B_1^n \tilde{B}_1^n E)_{\rho_{\Omega}} - O(1),
\]
proves (62) and, hence, Claim 5.2. \(\square\)

5.2. Sample application: fully quantum random access codes. One relatively simple application of our main result is to give upper bounds on the fidelity achieved by so-called Fully Quantum Random Access Codes (FQRAC). An FQRAC is a method for encoding \(m\) message qubits into \(n < m\) code qubits, such that any subset of \(k\) message qubits can be retrieved with high fidelity. Limits on the performance of random access codes with classical messages are rather well understood: the case \(k = 1\) was studied in [1,2,38], and upper bounds on the success probability that decay exponentially in \(k\) were derived in [9,20,65]. In the fully quantum case, [20] gives similar upper bounds on the fidelity that decay exponentially in \(k\). Here, we show that such exponential bounds for the fully quantum case can be obtained in a relatively elementary fashion via the concept of entropy accumulation. The example also highlights that entropy accumulation is already useful in its basic form (3), which does not involve statistics information \(X_i\). Indeed, here the bound on the entropy produced at every step comes from the bound on the number of code qubits.

Definition 5.3. A \((\epsilon, m, n, k)\)-Fully Quantum Random Access Code (FQRAC) consists of an encoder \(\mathcal{E}_{M_1 \rightarrow C_1^n}\) and a decoder \(\mathcal{D}_{C_1^n S \rightarrow \tilde{M}_S}\), where \(M_1^m\) represents the \(m\) message qubits, \(C_1^n\) represents the \(n\) code qubits, \(S\) represents a classical description of a subset of \(\{1, \ldots, m\}\) of size \(k\), and \(\tilde{M}_S\) represents the output of the decoder, corresponding to the \(k\) positions of \(M_1^m\) listed in \(S\). Such a code must satisfy the following: for any state \(\rho_{RM_1^m S}\) which is classical on \(S\), we must have that
\[
F(S(\rho_{RM_1^m S}), (\mathcal{D} \circ \mathcal{E})(\rho_{RM_1^m S}))^2 \geq 1 - \epsilon,
\]
where \(R\) is a reference system of arbitrary dimension, and where \(S_{M_1^m S \rightarrow \tilde{M}_S}\) is a TPCP map that selects the \(k\) positions of \(M_1^m\) corresponding to those in \(S\) and outputs them into \(\tilde{M}_S\). Moreover, \(F(\rho, \sigma) := \|\sqrt{\rho} \sqrt{\sigma}\|_1\) refers to the fidelity between two states \(\rho\) and \(\sigma\).

Entropy accumulation gives the following constraint on FQRACs:

Theorem 5.4. A \((\epsilon, m, n, k)\)-FQRAC satisfies
\[
1 - \epsilon = f^2 < 2^{-k\left(\frac{m-n-k+1}{m}\right)^2 + 3}.
\]

Compared to the previously derived bound (Theorem 9 of [20]), the one obtained here is tighter for small \(k\),\(^{19}\) whereas it is weaker for large \(k\).

\(^{19}\) The bound of Theorem 9 of [20] has a pre-factor of the order of \(m\) and is therefore only non-trivial if \(k > \log m\).
Proof. Since the fidelity bound must be true for any state \( \rho \), it must in particular be true for the state consisting of \( m \) maximally entangled pairs and a uniform distribution over subsets \( S \). For every \( i \in \{1, \ldots, k\} \), define

\[
\mathcal{M}_i : M_1^{m-i+1} \to M_1^{m-i} \tilde{J}_i \hat{M}_i
\]

as a TPCP map that does the following:

1. Generate an index \( \tilde{J}_i \) at random from \( \{1, \ldots, m-i+1\} \).
2. Move the contents of \( M_{\tilde{J}_i} \) into \( \hat{M}_i \), and set \( M_{m-i+1} \) to the contents of \( M_{m-i} \) with the \( \tilde{J}_i \)th position removed.

Finally, define the state

\[
\rho^i_{M_1^{m-i} M_i^c} := (\mathcal{M}_i \circ \cdots \circ \mathcal{M}_1 \circ \mathcal{E})(\Phi \otimes \mathcal{M}_1^{m} |M_1^{m} M_1^{m}\rangle),
\]

where \( |\Phi\rangle := \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle) \). The next step is to use Theorem 4.4 on the state \( \rho_{k} \) with the identifications

\[
A_i \to \hat{M}_i \quad B_i \to \tilde{J}_i \quad E \to C^n_1
\]

and the tradeoff function \( f \) being the constant function equal to

\[
\inf_{i,\nu^i} H(\hat{M}_1 | \hat{M}_1^{i-1} \tilde{J}_1^n C_1^n)_{\nu^i},
\]

where the infimum is taken over states \( \nu^i \) of the form

\[
\nu^i_{M_1^{m-i} \tilde{J}_i^n C_1^n} = \mathcal{M}_i \left( \omega^i_{M_1^{m-i} \tilde{J}_1^{i-1} M_1^{m-i+1} C_1^n} \right),
\]

for some state \( \omega^i \). Here we also used Remark 4.3, which asserts that the system \( R \) that is used when defining the min-tradeoff function can be chosen isomorphic to \( A_1^{m-1} B_1^{i-1} E \).

Note that the Markov chain condition is immediate from the fact that \( \tilde{J}_i \) is chosen at random. As the systems \( X_i \) are trivial, we naturally take \( \Omega \) to be the certain event. We find that

\[
H_{\min}^{f/2} (\hat{M}_1 | \tilde{J}_1^n C_1^n)_{\rho^k} \geq k \inf_{i, \nu^i} H(\hat{M}_1 | \hat{M}_1^{i-1} \tilde{J}_1^n C_1^n)_{\nu^i} - \sqrt{4k \log \frac{8}{f^2} \log 5}.
\]

Furthermore, again by Remark 4.3, if part of \( B \) is classical in \( \rho \), then it remains classical in \( \nu \). As a result, we can assume in the following that \( \tilde{J}_1^{i-1} \) is a classical system in \( \nu^i \).

We continue by computing the expectation over the choice of \( \tilde{J}_i \):

\[
H(\hat{M}_1 | \hat{M}_1^{i-1} \tilde{J}_1^n C_1^n)_{\nu^i} = \frac{1}{m-i+1} \sum_{j_i=1}^{m-i+1} H(M_{j_i} | \hat{M}_1^{i-1} C_1^n \tilde{J}_1^{i-1})_{\nu^i} \geq \frac{1}{m} \sum_{j_i=1}^{m-i+1} H(M_{j_i} | \hat{M}_1^{i-1} C_1^n \tilde{J}_1^{i-1})_{\nu^i} = \frac{1}{m} H(M_1^{m-i+1} | \hat{M}_1^{i-1} C_1^n \tilde{J}_1^{i-1})_{\nu^i}
\]

\[
= \frac{1}{m} H(M_1^{m-i+1} | \hat{M}_1^{i-1} C_1^n \tilde{J}_1^{i-1})_{\nu^i}
\]
\[ \begin{align*}
&= \frac{1}{m} \left( H(M_{i}^{m-i+1} \hat{M}_{i}^{i-1} C_{1}^{i} | \bar{J}_{i}^{i-1})_{\alpha_{\omega}} - H(\hat{M}_{i}^{i-1} C_{1}^{i} | \bar{J}_{i}^{i-1})_{\alpha_{\omega}} \right) \\
\geq & \frac{-n - k + 1}{m},
\end{align*} \]

(67)

(68)

where the last inequality holds because $\bar{J}_{i}$ is classical, which implies that the first entropy in the bracket of the penultimate expression is non-negative, and because the second entropy in the bracket is upper bounded by $n + k - 1$.

We now use Proposition 5.5 and Remark 5.6 of [51], which imply that

\[ H_{\text{max}}^{\sqrt{1 - f^2}} (\hat{M}_{1}^{k} | \bar{J}_{1}^{k} C_{1}^{k})_{\rho_{k}} \geq H_{\text{min}}^{f/2} (\hat{M}_{1}^{k} | \bar{J}_{1}^{k} C_{1}^{k})_{\rho_{k}} - \log \frac{1}{1 - (f^2/2 + \sqrt{1 - f^2} \sqrt{1 - f^2}/4)^2} \]

\[ \geq H_{\text{min}}^{f/2} (\hat{M}_{1}^{k} | \bar{J}_{1}^{k} C_{1}^{k})_{\rho_{k}} - \log \frac{3}{f^3}, \]

where the second inequality holds because the denominator in the logarithm is lower bounded by $f^3/3$, as can be readily verified. Combining this with the above gives

\[ H_{\text{max}}^{\sqrt{1 - f^2}} (\hat{M}_{1}^{k} | \bar{J}_{1}^{k} C_{1}^{k})_{\rho_{k}} \geq -k \left( \frac{n + k - 1}{m} \right) - \sqrt{4k \log \frac{8}{f^2} \log 5 - \log \frac{3}{f^3}}. \]

Conversely, note that, by assumption, the purified distance between $\rho_{k}$ and the state consisting of $k$ maximally entangled qubit pairs is upper bounded by $\sqrt{1 - (1 - \varepsilon)} = \sqrt{1 - f^2}$. Since the max-entropy of $k$ maximally entangled qubit pairs equals $-k$, we have

\[ H_{\text{max}}^{\sqrt{1 - f^2}} (\hat{M}_{1}^{k} | \bar{J}_{1}^{k} C_{1}^{k})_{\rho_{k}} \leq -k. \]

We have thus derived the condition

\[ \sqrt{4k \log \frac{8}{f^2} \log 5} \geq k \left( \frac{m - n - k + 1}{m} \right) - \log \frac{3}{f^3}. \]

(69)

It is easy to verify that this condition is violated whenever

\[ \log \frac{8}{f^2} > k \left( \frac{m - n - k + 1}{m} \right)^2 \]

(70)

is violated. In fact, if $\log \frac{8}{f^2} \leq k \left( \frac{m - n - k + 1}{5m} \right)^2$, then we have

\[ 4k \log \frac{8}{f^2} \log^2 5 \leq \frac{4 \log^2 5}{25} k^2 \left( \frac{m - n - k + 1}{m} \right)^2, \]

and

\[ \log \frac{3}{f^3} \leq \frac{3}{2} \log \frac{8}{f^2} \leq \frac{3}{50} k \left( \frac{m - n - k + 1}{m} \right)^2 \leq \frac{3}{50} k \left( \frac{m - n - k + 1}{m} \right). \]

Adding the square root of the first inequality and the second one, we get that inequality (69) is violated. Thus, the condition (70) must hold, and therefore also (63). \[ \square \]

\[ 20 \]  Because $\arcsin(f/2) + \arcsin(\sqrt{1 - f^2}) < \arcsin(f) + \arcsin(\sqrt{1 - f^2}) = \pi/2$, the condition of Remark 5.6 of [51] is satisfied.
6. Conclusions

Informally speaking, entropy accumulation is the claim that the operationally relevant entropy (the smooth min- or max-entropy) of a multipartite system is well approximated by the sum of the von Neumann entropies of its individual parts. This has ramifications in various areas of science, ranging from quantum cryptography to thermodynamics.

As described in Sect. 5, current cryptographic security proofs have various fundamental and practical limitations [46]. That these can be circumvented using entropy accumulation has already been demonstrated in [5] for the case of device-independent cryptography. We anticipate that the approach can be applied similarly to other cryptographic protocols. Examples include quantum key distribution protocols such as DPS and COW [27,50], for which full security has not yet been established.\(^{21}\) One may also expect to obtain significantly improved security bounds for protocols that involve high-dimensional information carriers and, in particular, continuous-variable protocols [25,62].\(^{22}\) A strengthening of current security claims may as well be obtained for other cryptographic constructions, such as bit commitment and oblivious transfer protocols (see, for example, [16,20,29]).

Entropy accumulation can also be used in statistical mechanics, e.g., to characterise thermalisation processes. At the beginning of Sect. 5 we outlined an argument that could confirm and make precise the intuition that entropy production (in terms of von Neumann entropy) is relevant for thermalisation. However, to base such arguments on physically realistic assumptions, it may be necessary to generalise Theorem 4.4 to the case where the Markov conditions (27) do not hold exactly. One possibility, permitted by the main result of [23], could be to replace them by the less stringent conditions

\[
H(B_i | B_i^{i-1} E) \approx H(B_i | A_i^{i-1} B_i^{i-1} E). 
\]

Another promising direction would be to apply entropy accumulation to estimate the entropy of low-energy states of many-body systems. One may expect that, under appropriate physical assumptions, these states possess a structure that permits a decomposition of the form described by Fig. 1 such that the Markov conditions required for Theorem 4.4, or at least some relaxations of them such as (71), hold. This may for example be the case for systems whose states are well approximated by matrix products states (see, e.g., [59]). We leave the investigation of such applications, as well as the development of corresponding extensions of the entropy accumulation theorem, for future work.

Acknowledgments. The authors would like to thank Martin Müller-Lennert, for allowing us to recopy Lemma 3.9 from his Master’s thesis into this paper (as Lemma B.4). We also thank Rotem Arnon-Friedman and Thomas Vidick for useful discussions about applications of the entropy accumulation theorem, Carl Miller for discussions on randomness expansion and Marco Tomamichel as well as the anonymous reviewers for comments on the manuscript. FD acknowledges the financial support of the Czech Science Foundation (GAČR) project no GA16-22211S and of the European Commission FP7 Project RAQUEL (Grant No. 323970). OF acknowledges support from the French National Research Agency via Project No. ANR-18-CE47-0011 (ACOM) and from LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007). RR acknowledges funding from the RAQUEL project, from the Swiss National Science Foundation (via Grant No. 200020-135048 and the National Centre of Competence in Research “Quantum Science and Technology”), from the European Research Council (Grant No. 258932), and from the US Air Force Office of Scientific Research (Grant Nos. FA9550-16-1-0245 and FA9550-19-1-0202).

\(^{21}\) These protocols do not have the required symmetries to employ standard techniques such as de Finetti-type theorems [42].

\(^{22}\) The security of continuous-variable protocols against general attacks has been proved [43], but the bounds have an unfavourable scaling in the finite-size regime.
Appendix A: The Function $\| \cdot \|_\alpha$

We use an extension of the Schatten $\alpha$-norm to the regime where $\alpha > 0$, which is defined for any operator $X = X_{B \leftarrow A}$ from a space $A$ to a space $B$ by

$$\|X\|_\alpha = \text{tr} \left( (X^\dagger X)^{\frac{\alpha}{2}} \right)^{\frac{1}{\alpha}}.$$

It follows from the Singular Value Theorem that $\|X\|_\alpha = \|X^\dagger\|_\alpha = \|X^T\|_\alpha = \|\overline{X}\|_\alpha$ (see, e.g, Section 2 of [61]), from which it also follows that

$$\|X\|_\alpha = \text{tr} \left( (XX^\dagger)^{\frac{\alpha}{2}} \right)^{\frac{1}{\alpha}}.$$

Note also that

$$\|X\|_\alpha^2 = \left\| X^\dagger X \right\|_\alpha = \left\| XX^\dagger \right\|_\alpha.$$  \hfill (72)

The following is Lemma 12 from [37].

**Lemma A.1.** For any non-negative operator $X$ and for any $\alpha \in \mathbb{R}^+$

$$\|X\|_\alpha = \begin{cases} \sup_Z \text{tr}(XZ^{\alpha'}) & \text{if } \alpha' \geq 0 \\ \inf_Z \text{tr}(XZ^{\alpha'}) & \text{if } \alpha' \leq 0 \end{cases}$$  \hfill (73)

where the supremum and infimum range over density operators $Z$.

Appendix B: Properties of the Sandwiched Rényi Entropies

The sandwiched Rényi entropy from Definition 2.3 is a special case of the sandwiched Rényi relative entropy, which is defined as follows.

**Definition B.1.** For two density operators $\rho$ and $\sigma$ on the same Hilbert space and for $\alpha \in (0, 1) \cup (1, \infty)$ the sandwiched relative Rényi entropy of order $\alpha$ is defined as

$$D_\alpha(\rho \| \sigma) = \frac{1}{\alpha'} \log \left\| \rho^{\frac{1}{\alpha'}} \sigma^{-\frac{\alpha'}{2}} \right\|_{2\alpha}^2,$$

where $\alpha' = \frac{\alpha - 1}{\alpha}$.
In particular, for a bipartite density operator $\rho_{AB}$, the sandwiched $\alpha$-Rényi entropy of $A$ conditioned on $B$ is related to this relative entropy by

$$H_\alpha(A|B)_\rho = -D_\alpha(\rho_{AB}\|\text{id}_A \otimes \rho_B) .$$  \hfill (74)

It turns out that this is not the only way to define a conditional entropy based on a relative entropy. One popular alternative is to replace the marginal $\rho_B$ by a maximisation over arbitrary density operators on $B$:

$$H_\alpha^\uparrow(A|B)_\rho = -\inf_{\sigma_B} D_\alpha(\rho_{AB}\|\text{id}_A \otimes \sigma_B) .$$  \hfill (75)

We refer to [53] for a comparison of the different notions.

The following Lemma corresponds to Eq. 19 of [37]. For its proof, it is convenient to represent vectors of product systems as matrices. Let $\{|i\>_A\}$ and $\{|j\>_B\}$ be fixed orthonormal bases of $A$ and $B$, respectively. For any vector

$$|\psi\>_A = \sum_{i,j} \gamma_{i,j} |i\>_A \otimes |j\>_B \in A \otimes B$$

we define the linear operator

$$\text{op}_{B\leftarrow A}(|\psi\>) = \sum_{i,j} \gamma_{i,j} |j\>_B \langle i|_A .$$

We emphasise that this definition is basis-dependent. Therefore, in expressions that involve this operator as well as the transpose operation $Z \mapsto Z^\top$, it is understood that both are taken with respect to the same basis. It is straightforward to prove the following properties (see, e.g., Section 2.4 of [61]). For any operators $X_{A'\leftarrow A}$ and $Y_{B'\leftarrow B}$,

$$Y_{B'\leftarrow B}\text{op}_{B\leftarrow A}(|\psi\>) X_{A'\leftarrow A} = \text{op}_{B\leftarrow A}((X_{A'\leftarrow A} \otimes Y_{B'\leftarrow B})|\psi\>) .$$  \hfill (76)

Furthermore,

$$\text{op}_{B\leftarrow A}(|\psi\>) \text{op}_{B\leftarrow A}(|\psi\>)^\top = \text{tr}_A(|\psi\><\psi|) \quad \text{and} \quad \text{op}_{B\leftarrow A}(|\psi\>)^\top \text{op}_{B\leftarrow A}(|\psi\>) = \text{tr}_B(|\psi\><\psi|)^\top ,$$  \hfill (77)

and, hence,

$$\|\text{op}_{B\leftarrow A}(|\psi\>)\|_2 = |||\psi|| .$$  \hfill (78)

Lemma B.2. For any density operators $\rho$ and $\sigma$ on the same Hilbert space and for $\alpha \in (0, 1) \cup (1, \infty)$ we have

$$D_\alpha(\rho\|\sigma) = \sup_{\tau} \frac{1}{\alpha'} \log \left\| \left( \frac{\tau^{\frac{\alpha'}{2}}}{\tau^{\frac{\alpha'}{2}}} \otimes \frac{\tau^{\frac{\alpha'}{2}}}{\tau^{\frac{\alpha'}{2}}} \right)|\psi\> \right\|^2$$

where $|\psi\><\psi|$ is a purification of $\rho$ and where the supremum ranges over all density operators $\tau$ on the purifying system. In particular, for any pure $\rho_{ABE} = |\psi\><\psi|$ we have

$$H_\alpha(A|B)_\rho = -\sup_{\tau_E} \frac{1}{\alpha'} \log \left\| \left( \frac{\tau^{\frac{\alpha'}{2}}}{\tau^{\frac{\alpha'}{2}}} \otimes \frac{\tau^{\frac{\alpha'}{2}}}{\tau^{\frac{\alpha'}{2}}} \right)|\psi\> \right\|^2 .$$
Proof. Let us denote by $A$ the Hilbert space on which $\rho$ and $\sigma$ act and by $E$ the purifying space, so that $|\psi\rangle$ is a vector on $A \otimes E$. Then, using (72) and (77), the sandwiched Rényi entropy can be written as

$$D_\alpha(\rho \parallel \sigma) = \frac{1}{\alpha'} \log \left\| \sigma^{\frac{-\alpha'}{2}} \rho \sigma^{\frac{-\alpha'}{2}} \right\|_{\alpha}$$

$$= \frac{1}{\alpha'} \log \left\| \sigma^{\frac{-\alpha'}{2}} \text{op}_{A \leftarrow E}(|\psi\rangle) \text{op}_{A \leftarrow E}(|\psi\rangle)^\dagger \sigma^{\frac{-\alpha'}{2}} \right\|_{\alpha}$$

$$= \frac{1}{\alpha'} \log \left\| \text{op}_{A \leftarrow E}(|\psi\rangle)^\dagger \sigma^{\frac{-\alpha'}{2}} \sigma^{\frac{-\alpha'}{2}} \text{op}_{A \leftarrow E}(|\psi\rangle) \right\|_{\alpha}.$$ 

Using Lemma A.1 as well as (72) and (76) we obtain

$$D_\alpha(\rho \parallel \sigma) = \sup_{\tau} \frac{1}{\alpha'} \log \text{tr} \left( \text{op}_{A \leftarrow E}(|\psi\rangle)^\dagger \sigma^{\frac{-\alpha'}{2}} \sigma^{\frac{-\alpha'}{2}} \text{op}_{A \leftarrow E}(|\psi\rangle) \tau^{\alpha'} \right)$$

$$= \sup_{\tau} \frac{1}{\alpha'} \log \left\| \sigma^{\frac{-\alpha'}{2}} \text{op}_{A \leftarrow E}(|\psi\rangle) \tau^{\frac{\alpha'}{2}} \right\|_2^2$$

$$= \sup_{\tau} \frac{1}{\alpha'} \log \left\| \text{op}_{A \leftarrow E}(\sigma^{\frac{-\alpha'}{2}} \otimes \tau^{\frac{\alpha'}{2}} |\psi\rangle) \right\|_2^2,$$

where the supremum is taken over density operators $\tau$ on $E$. The first equality of the lemma then follows by (78). Finally, the second equality is obtained via (74). \qed

The next lemma concerns the conditioning on classical information.

**Lemma B.3** (Proposition 5.1 of [52]). For any density operator $\rho_{ABX}$ which is classical on $X$, i.e.,

$$\rho_{ABX} = \sum_{x \in \mathcal{X}} p_x \rho_{AB|x} \otimes |x\rangle \langle x|_X,$$

where $\rho_{AB|x}$ are density operators on $A \otimes B$ and $\{|x\rangle\}_{x \in \mathcal{X}}$ is an orthonormal basis of $X$, we have

$$H_\alpha(A|BX)_{\rho} = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} p_x 2^{(1-\alpha)H_\alpha(A|B)_{\rho|x}}.$$

**Proof.** Using the explicit form of $\rho_{ABX}$, it is straightforward to verify that

$$\left( \rho_{B|X}^{\frac{1-\alpha}{2}} \rho_{ABX} \rho_{B|X}^{\frac{1-\alpha}{2}} \right)^\alpha = \sum_{x \in \mathcal{X}} p_x \left( \rho_{B|x}^{\frac{1-\alpha}{2}} \rho_{AB|x} \rho_{B|x}^{\frac{1-\alpha}{2}} \right)^\alpha \otimes |x\rangle \langle x|_X.$$

Taking the trace on both sides, the equality can be rewritten in terms of $\alpha$-entropies as

$$2^{(1-\alpha)H_\alpha(A|BX)_{\rho}} = \sum_{x \in \mathcal{X}} p_x 2^{(1-\alpha)H_\alpha(A|B)_{\rho|x}},$$

which concludes the proof. \qed
The following lemma can be found as Lemma 3.9 in [36]; the statement and its proof are given here for the convenience of the reader. The statement and proof can also be found in [52, Proposition 6.5].

Lemma B.4 (Lemma 3.9 from [36], a variant of Proposition 6.2 of [51]). Let $\rho \in D_{\leq}(A)$ and $\sigma \in \text{Pos}(A)$ with $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, and define $\varepsilon_{\max} := \sqrt{2 \text{tr} \rho - (\text{tr} \rho)^2}$. For $\varepsilon \in (0, \varepsilon_{\max})$ and $\alpha \in (1, 2]$, we have

$$D_{\max}^\varepsilon(\rho \| \sigma) \leq D_\alpha(\rho \| \sigma) + \frac{g(\varepsilon)}{\alpha - 1},$$

where $g(\varepsilon) = -\log \left( 1 - \sqrt{1 - \varepsilon^2} \right)$, and $D_{\max}^\varepsilon(\rho \| \sigma) = \inf \hat{\rho} \inf \{ \lambda : \hat{\rho} \leq 2^\lambda \sigma \}$, with the infimum ranging over all $\hat{\rho}$ within $\varepsilon$ of $\rho$ in purified distance.

Proof. Assume without loss of generality that $\sigma$ has full support. By Lemma 6.1 in [51], we can find a $\lambda$ such that $\lambda \geq D_{\max}^\varepsilon(\rho \| \sigma)$ where

$$\varepsilon = \sqrt{2 \text{tr}[\Delta] - \text{tr}[\Delta]^2} \quad (79)$$

and $\Delta$ is the positive part of $\rho - 2^\lambda \sigma$. It suffices to upper-bound $\lambda$ by $D_\alpha(\rho \| \sigma) + g(\varepsilon)/(\alpha - 1)$. Now, let $\{|e_i\}_{i \in S}$ be an orthonormal basis consisting of eigenvectors of $\rho - 2^\lambda \sigma$. Let $S_+$ be the subset of $S$ corresponding to positive eigenvalues. Define the non-negative numbers $r_i = \langle e_i | \rho | e_i \rangle$ and $s_i = \langle e_i | \sigma | e_i \rangle$. Note that for $i \in S_+$, we have $r_i - 2^\lambda s_i = \langle e_i | (\rho - 2^\lambda \sigma) | e_i \rangle \geq 0$ and therefore $r_i/2^{-\lambda} \geq 1$. We use this to bound

$$\text{tr}[\Delta] = \sum_{i \in S_+} (r_i - 2^\lambda s_i) \leq \sum_{i \in S_+} r_i \leq \sum_{i \in S_+} \left( \frac{r_i}{2^{\lambda}} \right)^{\alpha - 1}$$

$$= 2^{-\lambda(\alpha - 1)} \sum_{i \in S_+} r_i^{\alpha} s_i^{1-\alpha} \leq 2^{-\lambda(\alpha - 1)} \sum_{i \in S} r_i^{\alpha} s_i^{1-\alpha}.$$

Hence,

$$\frac{1}{\alpha - 1} \log \text{tr}[\Delta] \leq \frac{1}{\alpha - 1} \log \sum_{i \in S} r_i^{\alpha} s_i^{1-\alpha} - \lambda.$$

Now, we solve Eq. (79) for $\text{tr}[\Delta]$ and bound

$$\lambda \leq \frac{1}{\alpha - 1} \log \sum_{i \in S} r_i^{\alpha} s_i^{1-\alpha} - \frac{1}{\alpha - 1} \log \left( 1 - \sqrt{1 - \varepsilon^2} \right).$$

It remains to upper-bound $\frac{1}{\alpha - 1} \log \sum_{i \in S} r_i^{\alpha} s_i^{1-\alpha}$ by $D_\alpha(\rho \| \sigma)$. To this end, we define the TPCP map $F(X) = \sum_{i \in S} P_i X P_i$, where $P_i$ denotes the projector onto the subspace spanned by $e_i$. Note that

$$D_\alpha(F(\rho) \| F(\sigma)) = \frac{1}{\alpha - 1} \log \sum_{i \in S} r_i^{\alpha} s_i^{1-\alpha}.$$

The theorem then follows from the data processing inequality. □

\footnote{The positive part of a Hermitian operator $X$ is defined as $\{X > 0\}X$, where $\{X > 0\}$ is the projector onto the span of the eigenspaces of $X$ with positive eigenvalues.}
The following two lemmas relate the entropy conditioned on a classical value $x$ to the unconditioned entropy.

**Lemma B.5.** Let $\rho_{AB}$ be a quantum state of the form $\rho = \sum_x p_x \rho_{AB|x}$, where $\{p_x\}$ is a probability distribution over $\mathcal{X}$. Then, for any $x \in \mathcal{X}$ and any $\alpha \in (1, \infty)$,

$$H_{\alpha}^\uparrow (A|B)_\rho - \frac{\alpha}{\alpha - 1} \log \left( \frac{1}{p_x} \right) \leq H_{\alpha}^\uparrow (A|B)_{\rho_x} \quad (80)$$

and for $\alpha \in (0, 1)$,

$$H_{\alpha}^\uparrow (A|B)_\rho - \frac{\alpha}{\alpha - 1} \log \left( \frac{1}{p_x} \right) \geq H_{\alpha}^\uparrow (A|B)_{\rho_x} \quad (81)$$

**Proof.** For any $\alpha$, we have

$$D_\alpha (\rho_{AB} || \text{id}_A \otimes \sigma_B) = \frac{1}{\alpha - 1} \log \text{tr} \left( \left( \sigma_B^{\frac{1-\alpha}{\alpha}} \rho_{AB} \sigma_B^{\frac{1-\alpha}{\alpha}} \right)^\alpha \right) \geq \frac{1}{\alpha - 1} \log \text{tr} \left( \left( \sigma_B^{\frac{1-\alpha}{\alpha}} p_x \rho_{AB|x} \sigma_B^{\frac{1-\alpha}{\alpha}} \right)^\alpha \right) = \frac{1}{\alpha - 1} \log p_x^\alpha + \frac{1}{\alpha - 1} \log \text{tr} \left( \left( \sigma_B^{\frac{1-\alpha}{\alpha}} \rho_{AB|x} \sigma_B^{\frac{1-\alpha}{\alpha}} \right)^\alpha \right) = \frac{\alpha}{\alpha - 1} \log p_x + D_\alpha (\rho_{AB|x} || \text{id}_A \otimes \sigma_B) .$$

For the first inequality, we used the fact that $\rho_{AB} = \sum_x p_x' \rho_{AB|x'}$ implies that $\sigma_B^{\frac{1-\alpha}{\alpha}} \rho_{AB} \sigma_B^{\frac{1-\alpha}{\alpha}} \geq \sigma_B^{\frac{1-\alpha}{\alpha}} p_x \rho_{AB|x} \sigma_B^{\frac{1-\alpha}{\alpha}}$. Then used the fact that $y \mapsto y^\alpha$ is a monotone function on $[0, \infty)$. Taking the infimum over $\sigma_B$ and then multiplying both sides by $-1$, we get the desired result. The proof is the same for $\alpha \in (0, 1)$ except that the direction of the inequality is reversed. $\square$

**Lemma B.6.** Let $\rho_{AB}$ be a quantum state of the form $\rho = \sum_x p_x \rho_{AB|x}$, where $\{p_x\}$ is a probability distribution over $\mathcal{X}$. Then, for any $x \in \mathcal{X}$ and any $\alpha \in (1, 2)$,

$$H_{\frac{1}{\alpha}} (A|B)_\rho + \frac{\alpha}{\alpha - 1} \log \left( \frac{1}{p_x} \right) \geq H_{\frac{1}{\alpha}} (A|B)_{\rho_x} \quad (82)$$

**Proof.** We define the state $\rho_{ABX} = \sum_{x'} p_{x'} \rho_{AB|x'} \otimes |x'\rangle \langle x'|_X$. Note that it is legitimate to use the notation $\rho$ as the reduced state on $A \otimes B$ corresponds to $\rho_{AB}$. As conditioning can only decrease the entropy, we obtain

$$H_{\frac{1}{\alpha}}^\uparrow (A|B)_\rho \geq H_{\frac{1}{\alpha}}^\uparrow (A|BX)_\rho = \frac{\alpha}{\alpha - 1} \log \text{tr} \left( \left( \rho_{BX}^{\frac{\alpha - 1}{\alpha}} \rho_{ABX}^{\frac{\alpha - 1}{\alpha}} \right)^{\frac{1}{\alpha}} \right)$$

$$= \frac{\alpha}{\alpha - 1} \log \text{tr} \sum_{x'} p_{x'} \left( \rho_{B|x'}^{\frac{\alpha - 1}{\alpha}} \rho_{AB|x'}^{\frac{\alpha - 1}{\alpha}} \right)^{\frac{1}{\alpha}} \geq \frac{\alpha}{\alpha - 1} \log p_{x'} \left( \rho_{B|x'}^{\frac{\alpha - 1}{\alpha}} \rho_{AB|x'}^{\frac{\alpha - 1}{\alpha}} \right)^{\frac{1}{\alpha}}$$
where we used multiple times the orthogonality of the family \( \{ | x \rangle \}_{x \in \mathcal{X}} \) and of the family \( \{ | y \rangle \}_{y \in \mathcal{Y}} \). Similarly, \( M = W A B \) with \( A \). Similarly, \( M = W A B \) with \( A \) defined by \( \mathcal{M}(W_B) = \sum \Pi_{z,B} W_B \Pi_{z,B} \). Using the data processing inequality and the fact that \( (I_{AX} \otimes \mathcal{M})(\rho_{ABX}) = \rho_{ABX} \), we have

\[
H_{\alpha}^{\uparrow}(A|B)_{\rho} = - \inf_{\sigma} D_{\alpha}(\rho_{ABX} \parallel \sigma_B) = - \inf_{\sigma} D_{\alpha}(\rho_{ABX} \parallel \mathcal{M}(\sigma_B)) .
\]

We now show that for any state \( \sigma_B \), we have \( D_{\alpha}(\rho_{ABX} \parallel \mathcal{M}(\sigma_B)) = D_{\alpha}(\rho_{AB} \parallel \mathcal{M}(\sigma_B)) \). To make the notation lighter, we use the following \( \Pi_{y} \) for \( \Pi_{z,B} \) and \( \Pi_{x} \) for \( \Pi_{y,A} \). The relative entropy \( D_{\alpha}(\rho_{ABX} \parallel \sigma_B) \) is defined in terms of

\[
\begin{align*}
\text{tr} \left( \mathcal{M}(\sigma_B)^{1-\alpha} \rho_{ABX} \mathcal{M}(\sigma_B)^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} \\
= \text{tr} \left( \sum_{y,z} (\Pi_{y} \sigma_B \Pi_{z})^{1-\alpha} (\Pi_{y} \otimes \Pi_{z}) \omega_{AB}(\Pi_{y} \otimes \Pi_{z}) \otimes | t(y,z) \rangle \langle t(y,z) | (\Pi_{y} \sigma_B \Pi_{z})^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} \\
= \sum_{y,z} \text{tr} \left( (\Pi_{y} \sigma_B \Pi_{z})^{1-\alpha} (\Pi_{y} \otimes \Pi_{z}) \omega_{AB}(\Pi_{y} \otimes \Pi_{z}) \otimes | t(y,z) \rangle \langle t(y,z) | (\Pi_{y} \sigma_B \Pi_{z})^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} \\
= \sum_{y,z} \text{tr} \left( (\Pi_{y} \sigma_B \Pi_{z})^{1-\alpha} (\Pi_{y} \otimes \Pi_{z}) \omega_{AB}(\Pi_{y} \otimes \Pi_{z}) (\Pi_{y} \sigma_B \Pi_{z})^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} ,
\end{align*}
\]

where we used multiple times the orthogonality of the family \( \{ \Pi_{y} \} \) and of the family \( \{ \Pi_{z} \} \). Similarly, \( D_{\alpha}(\rho_{AB} \parallel \mathcal{M}(\sigma_B)) \) is defined in terms of

\[
\begin{align*}
\text{tr} \left( \mathcal{M}(\sigma_B)^{1-\alpha} \rho_{AB} \mathcal{M}(\sigma_B)^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} \\
= \sum_{y,z} \text{tr} \left( (\Pi_{z} \sigma_B \Pi_{y})^{1-\alpha} (\Pi_{y} \otimes \Pi_{z}) \omega_{AB}(\Pi_{y} \otimes \Pi_{z}) (\Pi_{z} \sigma_B \Pi_{y})^{\frac{1-\alpha}{2\alpha}} \right)^{\alpha} .
\end{align*}
\]

And this concludes the proof. \( \square \)
In the subsequent arguments we will use the quantity

\[ D'_\alpha(\rho \parallel \sigma) = \frac{1}{\alpha - 1} \log \text{tr}(\rho^\alpha \sigma^{1-\alpha}) \]

which is defined for any non-negative operators \( \rho \) and \( \sigma \) on the same space and for any \( \alpha \in [0, 1) \cup (1, \infty) \). As observed in [17, 24, 64], it follows from the Araki-Lieb-Thirring inequality [3, 31] that

\[ D'_\alpha(\rho \parallel \sigma) \leq D'_\alpha(\rho \parallel \sigma). \tag{87} \]

Furthermore, we can define a conditional entropy based on this quantity:

\[ H'_\alpha(A \mid B)_\rho := -\inf_{\sigma} D'_\alpha(\rho_{AB} \parallel \text{id}_A \otimes \sigma_B). \tag{88} \]

In [53, Theorem 2], it is shown that \( H' \) and \( H \) are duals of each other, in the sense that

\[ H_\alpha(A \mid B)_\rho = -H'_\frac{1}{\alpha}(A \mid C)_\rho \tag{89} \]

for any pure state \( \rho_{ABC} \).

The following lemma is another variant of Lemma 8 of [54] (see also Lemma 6.3 of [51]).

**Lemma B.8.** Let \( \rho \) be a density operator and \( \sigma \) a non-negative operator, let \( \eta = \max(4, 2D'_2(\rho \parallel \sigma) + 2D'_0(\rho \parallel \sigma) + 1) \), and let \( \alpha \in (1, 1 + 1/\log \eta) \). Then

\[ D'_\alpha(\rho \parallel \sigma) < D(\rho \parallel \sigma) + (\alpha - 1)(\log \eta)^2. \]

**Proof sketch.** The proof proceeds in the same way as the proof of Lemma 8 of [54]. The idea is to consider, for any \( \beta > 0 \), the functions \( r_\beta \) and \( s_\beta \) from \( \mathbb{R}^+ \) to \( \mathbb{R}^+ \) defined by

\[ r_\beta(t) = t^\beta - \beta \ln t - 1 \]
\[ s_\beta(t) = 2(\cosh(\beta \ln t) - 1). \]

It can be readily verified that \( r_\beta(t) \leq s_\beta(t) \) for all \( t > 0 \), that \( s_\beta(t) = s_\beta(1/t) \), that \( s_\beta(t) \) is monotonically increasing for \( t > 1 \), and that \( s_\beta(t) \) is concave for \( \beta < 1/2 \) and \( t \geq 3 \).

It is then shown that

\[ D'_\alpha(\rho \parallel \sigma) \leq D(\rho \parallel \sigma) + \frac{1}{\beta \ln 2} \langle \phi | r_\beta(X) | \phi \rangle, \]

where \( \beta = \alpha - 1 \), \( X = \rho \otimes \sigma^{-1T} \), and \( |\phi\rangle = (\sqrt{\rho} \otimes \text{id} \otimes |\gamma\rangle \) with \( |\gamma\rangle = \sum_i |i\rangle \otimes |i\rangle \).

From there, we proceed in a slightly different way, noting that

\[ s_\beta(t) = s_\beta(\max(t, \frac{1}{t})) \leq s_\beta(\max(t, \frac{1}{t}) + 1) \leq s_\beta(t + \frac{1}{t} + 1). \]

Using this as well as Lemma 11 of [54], we obtain

\[ \langle \phi | r_\beta(X) | \phi \rangle \leq \langle \phi | s_\beta(X) | \phi \rangle \leq \langle \phi | s_\beta(X + \frac{1}{X} + \mathbb{I}) | \phi \rangle \leq s_\beta(\langle \phi | X + \frac{1}{X} + \mathbb{I} | \phi \rangle), \]
which holds because \( s_\beta \) is concave for \( \beta \leq \frac{1}{\log \eta} \leq \frac{1}{2} \), and because the eigenvalues of \( X + \frac{1}{X} + I \) lie in the interval \([3, \infty)\). Using that
\[
\langle \phi | X + \frac{1}{X} + I | \phi \rangle = 2^{D'_2(\rho \| \sigma)} + 2^{-D'_0(\rho \| \sigma)} + 1 \leq \eta
\]
and combining the inequalities above, we find
\[
D'_\alpha(\rho \| \sigma) \leq D(\rho \| \sigma) + \frac{1}{\beta \ln 2} s_\beta(\eta) .
\]
Applying Taylor’s theorem to an expansion around \( \beta = 0 \) gives
\[
\frac{1}{\beta \ln 2} s_\beta(\eta) \leq \left( \frac{\ln \eta}{\beta} \right)^2 \cosh(\frac{\ln \eta}{\beta}) \leq \frac{\beta (\log \eta)^2}{\beta \ln 2} 2 \cosh(\beta \ln \eta) < \beta (\log \eta)^2 ,
\]
where for the last inequality, we use the fact that \( \ln 2 \cosh(\ln 2) < 1 \).

The following lemma is a generalisation of Proposition 3.10 of [36]. In [52, Section 6.4.2], using a Taylor approximation, the factor in front of \((\alpha - 1)\) can be improved although at the price of having the error term containing non-explicit constants.

**Lemma B.9.** For any density operator \( \rho_{AB} \) and \( 1 < \alpha < 1 + 1/\log(1 + 2d_A) \)
\[
H(A|B)_\rho - (\alpha - 1) \log^2(1 + 2d_A) < H_\alpha(A|B)_\rho \\
\leq H_{\frac{1}{\alpha}}(A|B)_\rho < H(A|B)_\rho + (\alpha - 1) \log^2(1 + 2d_A) ,
\]
where \( d_A = \dim A \).

**Proof.** We start with the proof of the first inequality. Lemma B.8 implies that
\[
H(A|B)_\rho - (\alpha - 1) \log^2(1 + 2d_A))^2 < -D'_\alpha(\rho_{AB} \| \id_A \otimes \rho_B)
\]
holds for all \( 1 < \alpha < 1 + \frac{1}{\log(1 + 2d_A)} \), where we have used that
\[
1 + 2d_A \geq 1 + 2D'_0(\rho_{AB} \| \id_A \otimes \rho_B) + 2D'_2(\rho_{AB} \| \id_A \otimes \rho_B) .
\]
Furthermore, because of (87) we have
\[
-D'_\alpha(\rho_{AB} \| \id_A \otimes \rho_B) \leq -D_\alpha(\rho_{AB} \| \id_A \otimes \rho_B) = H_\alpha(A|B)_\rho .
\]
Combining this with the above concludes the proof of the first inequality.

The second inequality follows directly from the monotonicity of the relative Rényi entropy in \( \alpha [8,37] \).

To prove the last inequality, we again use the duality relation (88):
\[
H_{\frac{1}{\alpha}}(A|B)_\rho = -H'_\alpha(A|C)_\rho \\
= \inf_{\sigma} D'_\alpha(\rho_{AC} \| \id_A \otimes \sigma_C) \\
\leq D'_\alpha(\rho_{AC} \| \id_A \otimes \rho_C) .
\]
We may now again use Lemma B.8 to obtain
\[
D'_\alpha(\rho_{AC} \| \id_A \otimes \rho_C) < -H(A|C)_\rho + (\alpha - 1) \log^2(1 + 2d_A)) .
\]
Combining the two inequalities with the fact that \(-H(A|C) = H(A|B) \) concludes the proof. \(\square\)
The following lemma generalises a classical result originally proposed in [44]. It follows rather directly from similar statements proved in [34, 36, 51, 54].

**Lemma B.10.** For any density operator $\rho$, any non-negative operator $\sigma$, any $\alpha \in (1, 2]$, and any $0 < \varepsilon < 1$,

$$
H_{\min}^{\varepsilon}(A|B)_{\rho} \geq H_{\alpha}^{\uparrow}(A|B)_{\rho} - \frac{g(\varepsilon)}{\alpha - 1}
$$

$$
H_{\max}^{\varepsilon}(A|B)_{\rho} \leq H_{\alpha}^{\downarrow}(A|B)_{\rho} + \frac{g(\varepsilon)}{\alpha - 1},
$$

where $g(\varepsilon) = -\log(1 - \sqrt{1 - \varepsilon^2}) < \log(2/\varepsilon^2)$.

**Proof.** For the first inequality, we use Lemma B.4, which directly implies that

$$
H_{\min}^{\varepsilon}(A|B)_{\rho} \geq \sup_{\sigma_B} \left(-D_{\alpha}(\rho_{AB}||\text{id}_A \otimes \sigma_B)\right) - \frac{g(\varepsilon)}{\alpha - 1}.
$$

The desired inequality then follows because

$$
\sup_{\sigma_B} \left(-D_{\alpha}(\rho_{AB}||\text{id}_A \otimes \sigma_B)\right) = H_{\alpha}^{\uparrow}(A|B)_{\rho}.
$$

To prove the second inequality we use the duality between smooth min- and max-entropy [55], which asserts that

$$
H_{\max}^{\varepsilon}(A|B)_{\rho} = -H_{\min}^{\varepsilon}(A|C)_{\rho}
$$

holds for any purification $\rho_{ABC}$ of $\rho_{AB}$. We can then employ Proposition 6.2 of [51],

$$
-H_{\min}^{\varepsilon}(A|C)_{\rho} \leq \inf_{\sigma_C} D_{\alpha}'(\rho_{AC}||\text{id}_A \otimes \sigma_C) + \frac{g(\varepsilon)}{\alpha - 1}.
$$

The claim then follows from (88):

$$
\inf_{\sigma_C} D_{\alpha}'(\rho_{AC}||\text{id}_A \otimes \sigma_C) = -H_{\alpha}^{\uparrow}(A|C)_{\rho}
$$

$$
= H_{\alpha}^{\downarrow}(A|B)_{\rho}.
$$

\[\square\]

**Appendix C: Necessity of the Markov Chain Conditions**

The aim of this section is to illustrate that the Markov chain conditions in Theorem 4.4 are important, in the sense that dropping them completely would render the statement invalid.

We first recall that a tri-partite density operator $\rho_{ABC}$ has the Markov state property $A \leftrightarrow B \leftrightarrow C$ if and only if the mutual information between $A$ and $C$ conditioned on $B$ equals zero, i.e., $I(A : C|B)_{\rho} = 0$ (see [26, 39], as well as [23] for a robust version). Using the properties of the conditional mutual information, one can easily derive the following claims:

---

24 Alternatively, one may again use Lemma 3.9 of [36], which asserts that the relation is also true for $D_{\alpha}$ instead of $D_{\alpha}'$, together with (87).
• Symmetry: $A \leftrightarrow B \leftrightarrow C$ implies $C \leftrightarrow B \leftrightarrow A$
• Local processing of endpoints: $A A' \leftrightarrow B \leftrightarrow C$ implies $A \leftrightarrow B \leftrightarrow C$
• Centering of information: $A A' \leftrightarrow B \leftrightarrow C$ implies $A \leftrightarrow A' B \leftrightarrow C$
• Composition: $A \leftrightarrow B \leftrightarrow C$ and $A' \leftrightarrow A B \leftrightarrow C$ imply $AA' \leftrightarrow B \leftrightarrow C$.

By the standard properties of Markov chains described above, it is straightforward to show that the set of Markov chain conditions (27) for a trivial $E$ system is equivalent to the set of conditions

$$A_i^i \leftrightarrow B_i^i \leftrightarrow B_{i+1}^n \quad (i = 1, \ldots, n-1). \quad (90)$$

Similarly, using the composition property, one can show that this set of conditions is equivalent to the set of conditions

$$A_i \leftrightarrow A_i^{i-1} B_i^1 \leftrightarrow B_{i+1}^n \quad (i = 1, \ldots, n-1).$$

The latter can be expressed in terms of the entropy equalities

$$H(A_i|A_i^{i-1} B_i^1)_{\rho} = H(A_i|A_i^1 B_i^1)_{\rho}.$$

The entropy accumulation statement for the smooth min-entropy in the simplified form (3) can thus be rewritten as

$$H_{\text{min}}^{\varepsilon}(A_1^n|B_1^n)_{\rho} > \sum_{i=1}^{n} \inf_{a_{i-1},b_{i-1},b_{i+1}} H(A_i|A_1^{i-1} B_i^n)_{\rho_{A_i B_i A_1^{i-1} B_i^{i-1} B_i^{i-1}}} - c_{\varepsilon} \sqrt{n}. \quad (91)$$

Note that, if one replaces the smooth min-entropy on the left hand side by the von Neumann entropy then this expression looks similar to the usual chain rule for von Neumann entropies,

$$H(A_i^n|B_i^n)_{\rho} = \sum_{i=1}^{n} H(A_i|A_1^{i-1} B_i^1)_{\rho},$$

which holds for arbitrary $\rho_{A_1^n B_1^n}$. One may therefore wonder whether (91) may also hold without the Markov conditions (90). This is however not the case, as we are going to show with a specific example.

The example is classical, in the sense that $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ correspond to random variables and the map $\mathcal{M}_i$ takes $a_1^{i-1} b_1^{i-1}$ as input and outputs $a_1^{i-1} b_1^{i-1}$ as is, together with $A_i B_i$ generated from the conditional distribution $\rho_{A_i B_i A_1^{i-1} = a_1^{i-1}, b_1^{i-1}}$. With this setup, (91) can be written as

$$H_{\text{min}}^{\varepsilon}(A_1^n|B_1^n)_{\rho} > \sum_{i} \inf_{a_{i-1},b_{i-1},b_{i+1}} H(A_i|B_i)_{\rho_{A_i B_i A_1^{i-1} = a_1^{i-1}, b_1^{i-1}}} - c_{\varepsilon} \sqrt{n}. \quad (92)$$

Actually, (92) is even weaker than what would follow from (91) as we are taking the infimum also over $b_{i+1}^n$ but we will see that even this weaker inequality is false. In fact, consider the following construction, let $B_1, B_2, \ldots, B_n$ be $n$ mutually independent and uniformly distributed $n$-bit strings. Furthermore let $C$ be a uniform random bit and let $A = A_1^n$ be an $n$-bit-string defined by

$$A = \begin{cases} B_1 \oplus B_2 \oplus \cdots \oplus B_n & \text{if } C = 0 \\ \text{uniform and independent} & \text{if } C = 1, \end{cases}$$
where $\oplus$ denotes the bit-wise addition modulo 2. In other words, with probability $1/2$, the string $A$ is fully determined by $B_1, \ldots, B_n$, and with probability $1/2$, $A$ is completely random. We then have, for $\varepsilon \ll 1$,

$$H_{\min}^{\varepsilon}(A^n_1 | B^n_1)_{\rho} = H_{\min}^{\varepsilon}(A^n_1 | B_1 \oplus B_2 \oplus \cdots \oplus B_n)_{\rho} \approx 1.$$  

Furthermore, for any $i$ and for any fixed $a^n_{i-1}, b^n_{i-1},$ and $b^n_{i+1}$, we have

$$H(A_i | B_i)_{\rho_{a^n_{i-1}b^n_{i-1}b^n_{i+1}}} \geq \frac{1}{2},$$

for the bit $A_i$ is random with probability $1/2$. Since there are $n$ such terms in the sum on the right hand side of (92), it scales linearly in $n$. But we have just seen that the left hand side is roughly equal to 1. This shows that this inequality, and hence also the quantum version (91), cannot hold in general if we drop the Markov chain conditions (90).

References

1. Ambainis, A., Nayak, A., Ta-Shma, A., Vazirani, U.: Dense quantum coding and a lower bound for 1-way quantum automata. In: Proceedings of the 31st Annual ACM Symposium on Theory of Computing, STOC ’99, pp. 376–383, New York, NY, USA. ACM (1999)
2. Ambainis, A., Nayak, A., Ta-Shma, A., Vazirani, U.: Dense quantum coding and quantum finite automata. J. ACM 49(4), 496–511 (2002). arXiv:quant-ph/9804043
3. Araki, H.: On an inequality of Lieb and Thirring. Lett. Math. Phys. 19(2), 167–170 (1990)
4. Arnon-Friedman, R., Renner, R.: de Finetti reductions for correlations. J. Math. Phys. 56(5) (2015) arXiv:1308.0312
5. Arnon-Friedman, R., Renner, R., Vidick, T.: Simple and tight device-independent security proofs. SIAM J. Comput. 48(1), 181–225 (2019)
6. Asorey, M., Kossakowski, A., Marmo, G., Sudarshan, E.G.: Relations between quantum maps and quantum states. Open Syst. Inf. Dyn. 12(04), 319–329 (2005). arXiv:quant-ph/0602228
7. Attal, S., Pautrat, Y.: From repeated to continuous quantum interactions. Ann. Henri Poincaré 7(1), 59–104 (2006)
8. Beigi, S.: Sandwiched Rényi divergence satisfies data processing inequality. J. Math. Phys. 54(12), 122202 (2013). arXiv:1306.5920
9. Ben-Aroya, A., Regev, O., de Wolf, R.: A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In: Proceedings of the FOCS (2008). arXiv:0705.3806
10. Bennett, C.H., Brassard, G.: Quantum cryptography: public key distribution and coin tossing. In: Proceedings of the International Conference on Computers, Systems and Signal Processing (1984)
11. Bennett, C.H., Brassard, G., Mermin, N.D.: Quantum cryptography without Bell’s theorem. Phys. Rev. Lett. 68(5), 557 (1992)
12. Berta, M., Christandl, M., Colbeck, R., Renes, J.M., Renner, R.: The uncertainty principle in the presence of quantum memory. Nat. Phys. 6, 659 (2010). arXiv:0909.0950
13. Bihain, E., Boyer, M., Boykin, P.O., Mor, T., Roychowdhury, V.: A proof of the security of quantum key distribution (extended abstract). In: Proceedings of the ACM STOC, pp. 715–724, New York, NY, USA. ACM (2000)
14. Bruneau, L., Joye, A., Merkli, M.: Repeated interactions in open quantum systems. J. Math. Phys. 55(7), 1 (2014)
15. Christandl, M., König, R., Renner, R.: Postselection technique for quantum channels with applications to quantum cryptography. Phys. Rev. Lett. 102(2), 020504 (2009). arXiv:0809.3019
16. Damgård, I.B., Fehr, S., Salvail, L., Schaffner, C.: Cryptography in the bounded quantum-storage model. In: Proceedings of the FOCS, pp. 449–458 (2005). arXiv:quant-ph/0508222
17. Datta, N., Leditzky, F.: A limit of the quantum Rényi divergence. J. Phys. A: Math. Theor. 47(4), 045304 (2014). arXiv:1308.5961
18. del Rio, L., Hutter, A., Renner, R., Wehner, S.: Relative thermalization. arXiv:1401.7997 (2014)
19. Dupuis, F., Fawzi, O.: Entropy accumulation with improved second-order term. IEEE Trans. Inform. Theory 65(11), 7596–7612 (2019). arXiv:1805.11652
20. Dupuis, F., Fawzi, O., Wehner, S.: Entanglement sampling and applications. IEEE Transactions on Information Theory 61(2), 1093–1112 (2015). arXiv:1305.1316
21. Ekert, A., Renner, R.: The ultimate physical limits of privacy. Nature 507(7493), 443–447 (2014)
22. Ekert, A.K.: Quantum cryptography based on Bell’s theorem. Phys. Rev. Lett. 67(6), 661 (1991)
23. Fawzi, O., Renner, R.: Quantum conditional mutual information and approximate Markov chains. Commun. Math. Phys. 340(2), 575–611 (2015). arXiv:1410.0664
24. Frank, R.L., Lieb, E.H.: Monotonicity of a relative Rényi entropy. J. Math. Phys. 54(12), 122201 (2013). arXiv:1306.5358
25. Grosshans, F., Grangier, P.: Continuous variable quantum cryptography using coherent states. Phys. Rev. Lett. 88, 057902 (2002)
26. Hayden, P., Jozsa, R., Petz, D., Winter, A.: Structure of states which satisfy strong subadditivity of quantum entropy with equality. Commun. Math. Phys. 246(2), 359–374 (2004). arXiv:quant-ph/0304007
27. Inoue, K., Honjo, T.: Robustness of differential-phase-shift quantum key distribution against photon-number-splitting attack. Phys. Rev. A 71, 042305 (2005)
28. Koashi, M., Imoto, N.: Operations that do not disturb partially known quantum states. Phys. Rev. A 66(2), 022318 (2002)
29. König, R., Wehner, S., Wolf, S.: Unconditional security from noisy quantum storage. IEEE Trans. Inform. Theory 58(3), 1962–1984 (2012). arXiv:0906.1030
30. Leifer, M.S.: Conditional density operators and the subjectivity of quantum operations. AIP Conf. Proc. 889(1), 172–186 (2007). arXiv:quant-ph/0611233
31. Lieb, E., Thirring, W.: Inequalities for the moments of the eigenvalues of the schrodinger equation and their relation to sobolev inequalities. In: Lieb, E., Simon, B., Wightman, A.S. (eds.) Studies in Mathematical Physics: Essays in honor of Valentine Bargman pp. 269–303 (1976)
32. Lo, H.-K., Chau, H.F.: Unconditional security of quantum key distribution over arbitrarily long distances. Science 283, 2050–2056 (1999)
33. Mayers, D.: Unconditional security in quantum cryptography. J. ACM 48, 351–406 (2001)
34. Miller, C.A., Shi, Y.: Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. In: Proceedings of the ACM STOC, pp. 417–426. ACM, (2014). arXiv:1402.0489
35. Miller, C.A., Shi, Y.: Universal security for randomness expansion. arXiv:1411.6608v3 (2014)
36. Müller-Lennert, M.: Quantum Relative Rényi Entropies. Master’s thesis, ETH Zurich (2013)
37. 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). arXiv:1306.3142
38. Nayak, A.: Optimal lower bounds for quantum automata and random access codes. In: Proceedings of the 40th Annual Symposium on Foundations of Computer Science, FOCS ’99, p. 369, Washington, DC, USA, (1999). IEEE Computer Society
39. Petz, D.: Sufficiency of channels over von Neumann algebras. Q. J. Math. 39(1), 97–108 (1988)
40. Portmann, C., Renner, R.: Cryptographic security of quantum key distribution. arXiv:1409.3525 (2014)
41. Renner, R.: Security of quantum key distribution. PhD thesis, ETH Zurich (2005). arXiv:quant-ph/0512258
42. Renner, R.: Symmetry of large physical systems implies independence of subsystems. Nat. Phys. 3, 645–649 (2007). arXiv:quant-ph/0703069
43. Renner, R., Cirac, J.I.: de Finetti representation theorem for infinite-dimensional quantum systems and applications to quantum cryptography. Phys. Rev. Lett. 102, 110504 (2009)
44. Renner, R., Wolf, S.: Smooth Rényi entropy and applications. In: Proc. IEEE ISIT (2004)
45. Renner, R., Wolf, S.: Simple and tight bounds for information reconciliation and privacy amplification. In: Roy, B. (ed.) Proceedings of the ASIACRYPT, volume 3788 of LNCS, pp. 199–216. Springer, Berlin (2005)
46. Scarani, V., Bechmann-Pasquinucci, H., Cerf, N.J., Dušek, M., Lütkenhaus, N., Peev, M.: The security of practical quantum key distribution. Rev. Mod. Phys. 81(3), 1301 (2009). arXiv:0802.4155
47. Scarani, V., Renner, R.: Quantum cryptography with finite resources: unconditional security bound for discrete-variable protocols with one-way postprocessing. Phys. Rev. Lett. 100, 200501 (2008)
48. Shannon, C.: A mathematical theory of communications. Bell Syst. Tech. J. 27, 379–423 (1948)
49. Shor, P.W., Preskill, J.: Simple proof of security of the bb84 quantum key distribution protocol. Phys. Rev. Lett. 85(2), 441–444 (Jul 2000). arXiv:quant-ph/0003004
50. Stucki, D., Brunner, N., Gisin, N., Scarani, V., Zbinden, H.: Fast and simple one-way quantum key distribution. Applied Physics Letters 87(19), 194108 (2005). arXiv:quant-ph/0506097
51. Tomamichel, M.: A Framework for Non-Asymptotic Quantum Information Theory. PhD thesis, ETH Zurich, (2012). arXiv:1203.2142
52. Tomamichel, M.: Quantum Information Processing with Finite Resources: Mathematical Foundations, vol. 5. Springer, Berlin (2015)
53. Tomamichel, M., Berta, M., Hayashi, M.: Relating different quantum generalizations of the conditional Rényi entropy. J. Math. Phys. 55(8), 082206 (2014). arXiv:1311.3887
54. Tomamichel, M., Colbeck, R., Renner, R.: A fully quantum asymptotic equipartition property. IEEE Trans. Inform. Theory 55, 5840–5847 (2009). arXiv:0811.1221
55. Tomamichel, M., Colbeck, R., Renner, R.: Duality between smooth min- and max-entropies. IEEE Trans. Inform. Theory, 56, 4674 (2010). arXiv:0907.5238v2
56. Tomamichel, M., Fehr, S., Kaniewski, J., Wehner, S.: A monogamy-of-entanglement game with applications to device-independent quantum cryptography. New J. Phys. 15(10), 103002 (2013)
57. Tomamichel, M., Hayashi, M.: A hierarchy of information quantities for finite block length analysis of quantum tasks. IEEE Trans. Inform. Theory 59(11), 7693–7710 (2013)
58. Tomamichel, M., Renner, R.: Uncertainty relation for smooth entropies. Phys. Rev. Lett. 106(11), 110506 (2011). arXiv:1009.2015
59. Verstraete, F., Cirac, J.I.: Matrix product states represent ground states faithfully. Phys. Rev. B 73(9), 094423 (2006)
60. Vitanov, A., Dupuis, F., Tomamichel, M., Renner, R.: Chain rules for smooth min- and max-entropies. IEEE Trans. Inform. Theory 59(5), 2603–2612 (2013). arXiv:1205.5231
61. Watrous, J.: Theory of quantum information (2011). https://cs.uwaterloo.ca/~watrous/LectureNotes.html
62. Weedbrook, C., Lance, A.M., Bowen, W.P., Symul, T., Ralph, T.C., Lam, P.K.: Quantum cryptography without switching. Phys. Rev. Lett. 93, 170504 (2004)
63. Wegman, M.N., Carter, J.L.: New hash functions and their use in authentication and set equality. J. Comput. Syst. Sci. 22(3), 265–279 (1981)
64. Wilde, M., Winter, A., Yang, D.: Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. Comm. Math. Phys. 331(2), 593–622 (2014). arXiv:1306.1586
65. Wullschleger, J.: Bitwise quantum min-entropy sampling and new lower bounds for random access codes (2010). arXiv:1012.2291

Communicated by M. M. Wolf