Sampling of min-entropy relative to quantum knowledge

Robert König and Renato Renner
rkoenig@caltech.edu, renner@phys.ethz.ch

February 2, 2008

Abstract

Let $X_1, \ldots, X_n$ be a sequence of $n$ classical random variables and consider a sample $X_{s_1}, \ldots, X_{s_r}$ of $r \leq n$ positions selected at random. Then, except with (exponentially in $r$) small probability, the min-entropy $H_{\text{min}}(X_{s_1} \cdots X_{s_r})$ of the sample is not smaller than, roughly, a fraction $\frac{r}{n}$ of the overall entropy $H_{\text{min}}(X_1 \cdots X_n)$, which is optimal.

Here, we show that this statement, originally proved in [S. Vadhan, LNCS 2729, Springer, 2003] for the purely classical case, is still true if the min-entropy $H_{\text{min}}$ is measured relative to a quantum system. Because min-entropy quantifies the amount of randomness that can be extracted from a given random variable, our result can be used to prove the soundness of locally computable extractors in a context where side information might be quantum-mechanical. In particular, it implies that key agreement in the bounded-storage model—using a standard sample-and-hash protocol—is fully secure against quantum adversaries, thus solving a long-standing open problem.

Contents

1 Introduction 2
2 Basic definitions and known results 3
  2.1 Randomness extractors 3
  2.2 Randomness condensers 4
  2.3 Constructing locally computable extractors: The sample-and-hash approach 5
  2.4 Averaging samplers are condensers: preservation of min-entropy rates 6
  2.5 Extractors, condensers and prior classical information 6
  2.6 Extractors, condensers and prior quantum information 7
3 Our contribution 8
  3.1 Main result: samplers are quantum condensers 8
  3.2 Related work 9
4 Proof sketch 10
  4.1 Proof idea 11
  4.2 Towards a modified chain-rule: Entropy-splitting 12
  4.3 (Partial) recombination of split states 12
  4.4 Putting it together: splitting, sampling and recombining 13
5 Rules and tools for min-entropy 15
  5.1 Preliminaries 15
  5.2 Definition of min-entropy 16
  5.3 Some basic rules and properties 18
  5.4 Entropy-splitting: A splitting-chain-rule for min-entropy 19
    5.4.1 A warm-up 20
1 Introduction

Let $X$ be a classical random variable and let $E$ be a (generally quantum-mechanical) system whose state might be correlated to $X$. The min-entropy of $X$ given $E$, denoted $H_{\text{min}}(X|E)$, is a natural measure for the uncertainty on the value of $X$ given access to the side information $E$. More precisely, $H_{\text{min}}(X|E)$ corresponds to the maximum length of a bitstring $R$ which is (a) uniquely determined by $X$ and (b) virtually uniform and independent of $E$.

Here, we study the following question initiated by Nisan and Zuckerman [NZ96].

Given a sequence $X_1, \ldots, X_n$ of $n$ classical random variables with min-entropy (relative to side information $E$) at least $H_{\text{min}}(X_1 \cdots X_n|E) \geq n\nu$, for some $\nu \geq 0$, what is the min-entropy $H_{\text{min}}(X_{s_1} \cdots X_{s_r}|E)$ of a randomly selected sample $X_{s_1}, \ldots, X_{s_r}$ of $r$ positions? In other words, we are starting with a sequence $X_1, \ldots, X_n$ which contains at least $n\nu$ bits of uniform (relative to $E$) randomness, and we are interested in the amount of uniform (again relative to $E$) randomness of the subsequence $X_{s_1}, \ldots, X_{s_r}$.

As a main result, we show that the min-entropy per position is preserved under sampling, i.e.,

$$\frac{1}{n} H_{\text{min}}(X_1 \cdots X_n|E) \geq \nu \quad \text{implies} \quad \frac{1}{r} H_{\text{min}}(X_{s_1} \cdots X_{s_r}|E) \geq \nu + o(1)$$

(except with probability exponentially small in $r$). This generalizes a result by Vadhan [Vad03] who considered the case where $E$ is purely classical.

A main application of this result is in the context of randomness extraction. It relies on the leftover-hash lemma [ILL89] (see also [BBCM95]), or, more precisely, its quantum generalization [Ren05] (see also [KMR05, RK05]), saying that the randomness of a classical random variable $X$, measured in terms of the min-entropy, can be extracted by applying a suitable hash function. That is, $X$ can be mapped to a string $Z$ of size (roughly) $H_{\text{min}}(X|E)$ which is virtually uniform and independent of $E$. Our result now implies that, given a long sequence $X_1, \ldots, X_n$ with sufficient min-entropy, random bits can be obtained by the sample-and-hash technique, i.e., first sampling a subsequence $X_{s_1}, \ldots, X_{s_r}$ and then applying a two-universal hash function.

The sample-and-hash technique is of interest in cryptography, in particular in the context of the bounded storage model [Man92]. Here, the security of cryptographic schemes is based on the assumption that a string of random variables $X_1, \ldots, X_n$, called randomizer, is temporarily available for public access, but too long to be stored on a computer, even by a potential adversary. The idea then is to use this string as a source of secret randomness.

---

1See Lemma 5.1 of Section 5.2 for a mathematically precise statement.
2Nisan and Zuckerman considered the special case where $E$ is classical.
3If the system $E$ is purely classical, it can generally be omitted in the analysis, as explained in Section 2.6.
Based on the original work by Maurer [Mau92], various schemes for key expansion in the bounded storage model have been proposed [DM02, DM04, Lu02, Vad03]. These are mostly based on the sample-and-hash technique described above. More precisely, a short initial string is used for selecting positions of the randomizer $X_1, \ldots, X_n$. Then a hash function is applied to extract a key $Z$.

Because the min-entropy of the randomizer $X_1, \ldots, X_n$ given the information $E$ stored by an adversary, $H_{\min}(X_1 \cdots X_n|E)$, is necessarily large, our result implies that the final key $Z$ is indeed uniform relative to $E$ and, hence, secret. In other words, our result proves that key expansion in the bounded storage model is possible in the context of a quantum adversary. It generalizes previous results [DM04, Lu02, Vad03] where security has been proved under the assumption that the adversary is purely classical.

Outline

The paper is organized as follows: We first cover some background material on randomness extraction in Section 2. In Section 3, we discuss our main result and its relation to prior work. Section 4 provides an informal overview of the central ideas involved in the proof. The remainder of the paper is devoted to a formal derivation of our main results; in Section 5, we establish the required properties of min-entropy. We subsequently apply these to the problem at hand in Section 6 where we derive our main result. We conclude in Section 7 by giving explicit parameters for key expansion in the bounded storage model.

2 Basic definitions and known results

2.1 Randomness extractors

Randomness extraction, i.e., the process of transforming partially random data $X$ into a uniformly distributed string $Z$, plays an important role in computer science and, in particular, cryptography. For example, it is used to generate secure keys, given only partially secret raw data. One of the most fundamental results in the area of randomness extraction is the leftover-hash lemma [ILL89]. It states that the number of uniform bits that can be extracted from a given random variable $X$ by two-universal hashing (i.e., by applying a function chosen at random from a two-universal set of hash functions) is roughly equal to the min-entropy $H_{\min}$ of $X$ defined by

$$H_{\min}(X) := - \log \max_x P_X(x).$$

We can express this result more formally by saying that two-universal hashing is an extractor. A $(k, \varepsilon)$-extractor is a function $\text{Ext} : X \times Y \to Z$ with the property that the random variable $Z = \text{Ext}(X,Y)$ is $\varepsilon$-close to uniform, i.e.,

$$\frac{1}{2} \| P_{\text{Ext}(X,Y)} - P_{U_Z} \| \leq \varepsilon,$$

whenever $X$ is a random variable $X$ with min-entropy at least $H_{\min}(X) \geq k$ and $Y$ is an independent and uniform seed, i.e., $P_Y \equiv P_{U_Y}$. (Here $P_{U_Z}$ denotes the uniform distribution on $Z$.) A strengthening of this notion is the concept of a strong extractor, whose output is required to be uniform even conditioned on the seed $Y$. A strong $(k, \varepsilon)$-extractor satisfies the inequality

$$\| P_{\text{Ext}(X,Y)|Y} - P_{U_Z} \cdot P_{U_Y} \| \leq \varepsilon$$

for all $P_{X|Y} \equiv P_X \cdot P_{U_Y}$ with $H_{\min}(X) \geq k$. Two-universal hashing corresponds to a strong $(k, \varepsilon)$-extractor with $\ell$ bits of output, for any $\varepsilon \geq 0$ and $k \geq \ell + 2 \log 1/\varepsilon$.

---

4In the literature, the quantity $H_{\min}$ is also denoted $H_\infty$ and called Rényi entropy of order $\infty$.

5The $L_1$-norm of a function $f : Z \to R$ is defined as $\| f \| := \sum_{z \in Z} | f(z) |$. 

3
While two-universal hashing is optimal in the number $H_0(Z) := \log |Z|$ of bits it can extract, it is not usable in certain applications. For example, computing the output $Z = \text{Ext}(X, Y)$ might be infeasible, e.g., if the initial number $H_0(X) = n$ of bits is too large to be processed by a limited computational device. Also, in cryptographic scenarios, the seed $Y$ is sometimes a (secret) key of limited size (e.g., $H_0(Y) = O(\log n)$) compared to the length of $X$. Thus it is natural to try to find extractors with additional properties, such as efficient computability or limited seed length. An example of such a requirement which is important for applications in the bounded storage model is local computability; in other words, if $X = (X_1, \ldots, X_n)$ consists of a large number $n$ of blocks (or bits), the output $\text{Ext}(X, Y)$ should only depend on a small subset $X_S = (X_{s_1}, \ldots, X_{s_r})$ of these values, where $S = \{s_1, \ldots, s_r\} = S(y) \subset [n] = \{1, \ldots, n\}$ specifies the subset for every $y \in \mathcal{Y}$. In other words, these extractors are of the form $\text{Ext}(X, Y) = f(X_{S(y)}, Y)$.

### 2.2 Randomness condensers

With the aim of finding other constructions of extractors, it is natural to consider weaker notions of randomness generation. One natural way to generalise the concept of a randomness extractor is to require that the output is only close to a random variable with high min-entropy (instead of being close to a uniform random variable). This leads to the definition of a $(k, k', \varepsilon)$-condenser: This is a function $\text{Cond} : \mathcal{X} \times \mathcal{Y} \to Z$ such that for all random variables $X$ with $H_{\min}(X) \geq k$, there is a random variable $Z$ with $H_{\min}(Z) \geq k'$ such that

$$\frac{1}{2}\|P_{\text{Cond}(X,Y)} - P_Z\| \leq \varepsilon,$$

where $Y$ is a uniform and independent seed on $\mathcal{Y}$. In terms of the so-called smooth min-entropy $H_{\min}^\varepsilon(Z) := \sup_{\nu \parallel \leq \varepsilon} P_{\nu} - P_{\nu}^{\parallel \leq \varepsilon}$, this requirement is simply expressed by

$$H_{\min}^\varepsilon(\text{Cond}(X, Y)) \geq k'.$$

The notion of a condenser is a strict generalisation of the notion of an extractor. Indeed, a $(k, \varepsilon)$-extractor $\text{Ext} : \mathcal{X} \times \mathcal{Y} \to Z$ is a $(k, \log |Z|, \varepsilon)$-condenser and vice versa.

Again, a stronger version of condensers is obtained by requiring that $\text{Cond}(X, Y)$ has high smooth entropy with high probability over $y$. The analog of \footnote{The supremum ranges over all subnormalised probability distributions $P_\nu$, that is functions $P_\nu : Z \to [0,1]$ satisfying $\sum_{z \in Z} P_\nu(z) = 1$.} defining a strong $(k, \varepsilon')$-condenser is the requirement that for every $X$ with $H_{\min}(X) \geq k$, there exists a joint distribution $P_{\tilde{Z}Y}$ such that

$$\frac{1}{2}\|P_{\text{Cond}(X,Y)Y} - P_{\tilde{Z}Y}\| \leq \varepsilon,$$

where $Y$ is independent of $X$ with uniform distribution $P_Y = P_{\text{dy}}$ on $\mathcal{Y}$, and $H_{\min}(\tilde{Z}|\tilde{Y}) \geq k'$. Here, the conditional min-entropy is defined as

$$H_{\min}(\tilde{Z}|\tilde{Y}) := -\log \max_{y \in \mathcal{Y}} P_Y(y) \max_{z} P_{\tilde{Z}|\tilde{Y}=y}(z).$$

As before, this requirement is equivalent to demanding that

$$H_{\min}^\varepsilon(\text{Cond}(X, Y)|Y) \geq k',$$

where $H_{\min}^\varepsilon(Z|Y) := \sup_{\nu} \|P_{\nu Y} - P_{\nu Y}\| \leq \varepsilon$ is the conditional smooth min-entropy. With this definition, a function $\text{Ext} : \mathcal{X} \times \mathcal{Y} \to Z$ is a strong $(k, \varepsilon)$-extractor if and only if it is a strong $(k, \log |Z|, \varepsilon)$-condenser.
2.3 Constructing locally computable extractors: The sample-and-hash approach

Condensers can be used as a building block for constructing extractors. A possible way of obtaining a new construction is by applying an extractor to the output of a condenser. More precisely, suppose that

\[
\text{Cond} : \mathcal{X}_C \times \mathcal{Y}_C \rightarrow \mathcal{X}_E \quad \text{is a } (k_C, k_E, \varepsilon_C) \rightarrow \text{condenser, and}
\]
\[
\text{Ext} : \mathcal{X}_E \times \mathcal{Y}_E \rightarrow \mathcal{Z}_E \quad \text{is a } (k_E, \varepsilon_E) \rightarrow \text{extractor}.
\]

It is easy to see that in this situation, the function

\[
\hat{\text{Ext}} : \mathcal{X}_C \times (\mathcal{Y}_C \times \mathcal{Y}_E) \rightarrow \mathcal{Z}_E
\]

\[
(x_C, (y_C, y_E)) \mapsto \text{Ext}(\text{Cond}(x_C, y_C), y_E)
\]

is a \((k_C, \varepsilon_C + \varepsilon_E)\)-extractor. This is because the condenser \(\text{Cond}\) generates a random variable with a sufficient amount of min-entropy for \(\text{Ext}\). This conclusion is also true for the strong versions of these notions: if \(\text{Cond}\) and \(\text{Ext}\) are a strong condenser and a strong extractor, respectively, then the function \(\hat{\text{Ext}}\) is a strong extractor.

Let us now return to the problem of constructing locally computable extractors. Clearly, if \(\text{Cond}(X, Y) = \text{Cond}((X_1, \ldots, X_n), Y)\) is of the form \(\text{Cond}(X, Y) = X_{S(Y)}\), where \(S(y) \subset [n]\) is a subset of indices for every \(y \in \mathcal{Y}\), then the previous construction results in an extractor of the form \(\hat{\text{Ext}}(X, Y) = \text{Ext}(X_1, \ldots, X_n, (Y_C, Y_E)) = \text{Ext}(X_{S(Y_C)}, Y_E)\). This extractor is clearly locally computable. This way of building a locally-computable extractor by first sampling a few indices specified by \(S(Y_C)\) at random and then applying an extractor is called the sample-and-hash approach. Building locally computable extractors is thus reduced to the problem of constructing condensers of the form \(\text{Cond}(X, Y) = X_{S(Y)}\).

2.4 Averaging samplers are condensers: preservation of min-entropy rates

Consider a sequence of random variable \(X = (X_1, \ldots, X_n)\) on \(\mathcal{X}^n\) and assume that the min-entropy rate \(R_{\min}(X) := \frac{H_{\min}(X)}{\log |\mathcal{X}|}\) is lower bounded by \(\mu\), i.e.,

\[
\mu \leq R_{\min}(X) = \frac{1}{n \log |\mathcal{X}|} H_{\min}(X_1 \cdots X_n).
\]

We will call the quantity \(\frac{H_{\min}(X)}{\log |\mathcal{X}|}\) on the lhs the min-entropy rate of \(X\). Suppose further that we select \(r\) of these random variables at random, resulting in a subset \(X_S = (X_{s_1}, \ldots, X_{s_r})\) corresponding to indices \(S = \{s_1, \ldots, s_r\}\). Intuitively, one would expect that with high probability over the choice of \(S\), the amount of randomness contained in such a sample is proportional to its size \(r = |S|\), i.e.,

\[
\mu - \delta \leq R_{\min}(X_S) = \frac{1}{|S| \log |\mathcal{X}|} H_{\min}(X_S)
\]

for some small \(\delta > 0\). In other words, we expect the min-entropy rate to be preserved under sampling. Indeed, as shown by Vadhan [Vad03] (improving on previous work by Nisan and Zuckerman [NZ96]), inequality (3) is correct with high probability (over the choice of the sample \(S = \{s_1, \ldots, s_r\}\)). In the terminology of condensers, this is saying that the function

\[
\text{Cond} : \mathcal{X}^n \times \binom{[n]}{r} \rightarrow \mathcal{X}^r
\]

\[
((X_1, \ldots, X_n), S) \mapsto X_S
\]

is a \((\mu \log |\mathcal{X}|, (\mu - \delta)r \log |\mathcal{X}|, \varepsilon)\)-condenser for some small \(\delta, \varepsilon > 0\). We call this function the \(\binom{n}{r}\)-subset condenser.
Neglecting issues related to computational complexity, a condenser of the form \( \text{Cond}(X, Y) = X_{S(Y)} \) is fully specified by the distribution \( P_S = P_{S(Y)} \) over subsets of \([n]\). The \( \binom{n}{r} \)-subset condenser is simply represented by the uniform distribution over all subsets \( S \subset [n] \) of size \(|S| = r\).

It is natural to ask which distributions over subsets \( S \) give rise to good condensers. Intuitively, a necessary condition is that the set of subsets \( S \) covers \([n]\) well in some sense. In fact, Vadhan [Vad03] showed that it suffices for \( S \) to be a so-called averaging sampler; i.e., a distribution over subsets of \([n]\) which can be used to approximate the average of any \( n \) values. Formally, such a sampler is defined as follows:

**Definition 2.1.** An \((n, \xi, \varepsilon)\)-sampler is a probability distribution \( P_S \) over subsets \( S \subset [n] \) with the property that

\[
\Pr_{S \in \mathcal{S}} \left[ \frac{1}{|S|} \sum_{i \in S} \beta_i - \xi \leq \varepsilon \right] \leq \varepsilon \quad \text{for all } \beta_1, \ldots, \beta_n \in [0, 1]^n .
\]

For simplicity, we will assume that \( P_S \) is 

observe that we only consider a one-sided error. We will call \( \xi \) the accuracy of the sampler, and \( \varepsilon \) its failure probability. Returning to our example, the uniform distribution over subsets of a fixed size is an averaging sampler with the following parameters.

**Lemma 2.2.** Let \( r < n \) and let \( P_S \) be the uniform distribution over subsets \( S \subset [n] \) of size \(|S| = r\). This defines a \((n, \xi, e^{-r \xi^2/2})\)-sampler for every \( r > 0 \) and \( \xi \in [0, 1] \).

This statement is a consequence of the Hoeffding-Azuma inequality and given as Lemma 5.5 of [BH05]. We call this sampler simply the \( \binom{n}{r} \)-subset sampler. It will be sufficient for our purposes, but our results hold more generally for arbitrary averaging samplers.

Vadhan showed that in the same way as the \( \binom{n}{r} \)-subset sampler gives rise to the \( \binom{n}{r} \)-condenser, any averaging sampler defines a corresponding condenser (with appropriate parameters). In other words, a probability distribution \( P_S \) over subsets of \([n]\) with the sampler property \( \xi \) preserves the min-entropy rate when picking a random subset, in the sense of [3].

### 2.5 Extractors, condensers and prior classical information

In cryptographic settings, it is often desirable to generate randomness which is not only (close to) uniform, but also independent of an adversary’s prior information. We first consider the case where the adversary is classical, such that her information is described by a random variable \( E \). In other words, the task is to generate a key \( Z \) satisfying \( \frac{1}{2} \| P_{Z^E} - P_{U^E} \cdot P_E \| \leq \varepsilon \), where \( \tilde{E} \) summarises the adversary’s knowledge.

Suppose the initial situation is described by a joint distribution \( P_{XE} \), where \( X \) is held by the honest parties, and the adversary holds \( E \). We will assume that the adversary’s information about \( X \) is limited; this is expressed by a lower bound on the conditional entropy \( H_{\text{min}}(X|E) \).

Conveniently, a strong \((k, \varepsilon)\)-extractor achieves key extraction in this setup, when invoked with (public) independent randomness \( Y \). That is, we have

\[
\frac{1}{2} \| P_{\text{Ext}(X,Y)Y^E} - P_{U^E} \cdot P_{U^Y} \cdot P_E \| \leq 2 \varepsilon
\]

for all \( P_{XE} \) with \( H_{\text{min}}(X|E) \geq k + \log 1/\varepsilon \). In other words, if the adversary’s initial prior information \( E \) about \( X \) is limited, the extracted key \( Z = \text{Ext}(X,Y) \) will look uniform to the adversary.

---

\(^7\) We note that the notion of samplers is usually defined differently in the computer science literature. There, a sampler is an algorithm which efficiently approximates the average of a large number of values. The aim is to give an estimate of the average \( \frac{1}{r} \sum_{i=1}^{n} \beta_i \) of an (arbitrary) vector \( \beta_1, \ldots, \beta_n \in [0, 1]^n \) whose entries are accessible in the form of an oracle. Here we restrict our attention to so-called averaging samplers: These output the value \( \frac{1}{|S|} \sum_{i \in S} \beta_i \) of a (randomly) chosen subset \( S \subset [n] \) of values. For a more detailed discussion of samplers and their computational aspects, see [Gol97].
even if he is given the seed of the extractor (i.e., $\tilde{E} = (E, Y)$). This procedure of using public (independent) randomness to generate secret keys from partially secret information is well-known as privacy amplification [BBCM95] (usually in conjunction with two-universal hashing as an extractor).

Inequality (5) is a trivial application of Markov’s inequality; it is obtained by applying the extractor property to the conditional distributions $P_{X|E=\epsilon}$. A similar conclusion holds more generally for any strong $(k, k', \epsilon)$-condenser: Here we have

$$H_{\min}^{z_{\epsilon}}(\text{Cond}(X, Y)|YE) \geq k'$$

for all joint distributions $P_{XE}$ with $H_{\min}(X|E) \geq k + \log 1/\epsilon$. This means that the problem of randomness extraction in the context of prior classical information essentially reduces to the randomness generation problem without any side-information.

### 2.6 Extractors, condensers and prior quantum information

The mentioned property of extractors and condensers fails to be true in cases where the adversary’s prior information $E$ is quantum. Indeed, in this case, the conditional distributions $P_{X|E=\epsilon}$ are no longer defined, and the analysis of randomness extraction has to be done differently.

The relevant concepts in this modified setup are sufficiently straightforward to define: Consider a classical random variable $X$ and a quantum system $E$ which is correlated to this variable. This situation is completely described by a classical-quantum state $\rho_{XE} = \sum_{x \in X} P_X(x)|x\rangle\langle x| \otimes \rho_E^x$ (where $\{|x\rangle\}_x$ is an orthonormal basis), or equivalently the ensemble $\{P_X(x), \rho_E^x\}_{x \in X}$ on $E$.

For the purpose of randomness extraction, the relevant measure of min-entropy is the conditional min-entropy $H_{\min}(X|E)$ introduced in [Ren05]; this quantity is defined by

$$H_{\min}(X|E) := -\log \min_{\sigma_E} \{\lambda : \rho_{XE} \leq \lambda \cdot \text{id}_X \otimes \sigma_E\}.$$  

The conditional min-entropy generalizes the classical min-entropy [1]. For classical-quantum states $\rho_{XE}$, the min-entropy $H_{\min}(X|E)$ characterises the amount of uniform randomness $Z = f(X)$ that can be extracted from $X$ such that $Z$ is independent of $E$.

In terms of this measure of prior information, a $(k, \epsilon)$-strong quantum extractor is a function $\text{Ext} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ with the property that (cf. (5))

$$\frac{1}{2} \|\rho_{\text{Ext}(X,Y)YE} - \rho_{\text{Ext}} \otimes \rho_{\text{Ext}} \otimes \rho_{E}\| \leq \epsilon$$

for all classical-quantum-states $\rho_{XE}$ with $H_{\min}(X|E) \geq k$. In this expression, $Y$ is an independent and uniform seed on $\mathcal{Y}$, and $\rho_{\text{Ext}}$ denotes the completely mixed state on $\mathcal{Z}$, i.e., the state $\frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} |z\rangle\langle z|$. Clearly, a $(k, \epsilon)$-strong quantum extractor is a $(k, \epsilon)$-strong extractor in the original (classical) sense. The converse is not true in general (see [GKK+07] for a particularly striking example in the bounded storage model). However, the leftover hash lemma can be generalised to the quantum case: the two-universal hashing construction $\text{Ext} : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^\ell$ is a $(k, \epsilon)$-strong quantum extractor for any $k \geq \ell + 2 \log 1/\epsilon$, as shown by Renner [Ren05]. (The optimality of this extractor with respect to the number of extracted bits is shown below in Lemma 6.1) As with classical extractors, an important goal is to find constructions which are more randomness-efficient, and satisfy additional properties such as local computability.

Similarly, a $(k, k', \epsilon)$-strong quantum condenser $\text{Cond} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ is defined by the requirement (cf. (6))

$$H_{\min}^{z_{\epsilon}}(\text{Cond}(X, Y)|YE) \geq k'$$

for all $\rho_{XE}$ with $H_{\min}(X|E) \geq k$. In this expression, the smooth min-entropy $H_{\min}^{z_{\epsilon}}(X|E)$ is defined by a maximisation over a set of operators in the vicinity of $\rho_{XE}$. Note that there is

---

*This definition is meaningful arbitrary bipartite states $\rho_{XE}$ even with non-classical part $X$.\*
a certain freedom in these definitions (the only constraint is the preservation of the desirable compositability properties). We choose to define the smooth min-entropy as

\[ H_{\min}^\varepsilon(X|E) = \sup_{\bar{\rho}_{XE}: \|\bar{\rho}_{XE} - \rho_{XE}\| \leq \varepsilon} H_{\min}(X|E)_{\bar{\rho}_{XE}}, \]

where the maximisation is over all subnormalised nonnegative operators \( \bar{\rho}_{XE} \) in an \( \varepsilon \)-ball around \( \rho_{XE} \), and the quantity on the rhs is the min-entropy of the corresponding operator (see below for a formal definition). As shown in [Ren05], our main result states that samplers can be used to “condense” min-entropy even in a quantum context, in the same way as they give rise to randomness condensers for classical distributions (as discussed in Section 2.4). More precisely, we consider an \( n \)-tuple \( X^n = (X_1, \ldots, X_n) \) of random variables on \( \mathcal{X}^n \), where \( \mathcal{X} \) is a (large) alphabet. We show that relative to a quantum system \( E \), the min-entropy rate is preserved when picking a random subset \( X_S \) (using a sampler).

To express this in a concise form, we introduce the min-entropy rates

\[ R_{\min}^\varepsilon(A|B)_\rho := \frac{H_{\min}^\varepsilon(A|B)_\rho}{H_0(A)_\rho}, \]

where \( H_0(A) = \log |A| \) is the alphabet size of \( A \). Our main result states that this quantity is approximately preserved under sampling. Clearly, when applied to a \((n, \xi, \varepsilon)\)-sampler, such a statement must depend on the accuracy \( \xi \) of the sampler and its failure probability \( \varepsilon \). For such a sampler and the situation described above, our main result is given by the inequality

\[ R_{\min}^\varepsilon(X_S|SE)_\rho \geq R_{\min}(X^n|E)_\rho - 3\xi - 2\kappa \log 1/\kappa, \]

where the parameters \( \varepsilon' \) and \( \kappa \) are equal to

\[ \varepsilon' = 2 \cdot 2^{-\xi n \log |\mathcal{X}|} + 3\varepsilon^{1/4} \quad \text{and} \quad \kappa = \frac{n}{|S| \log |\mathcal{X}|}. \]

(This result is stated as Corollary 6.19 below.) This inequality shows that (for appropriate alphabet sizes) the min-entropy rate is preserved, up to the accuracy of the sampler. As expected, the failure probability \( \varepsilon \) of the sampler is reflected in the distance (i.e., the smoothness parameter \( \varepsilon' \)). In fact, this distance mainly depends on the failure probability of the sampler, and the term \( 2 \cdot 2^{-\xi n \log |\mathcal{X}|} \) is usually negligible.

Observe that the expression \( 2\kappa \log 1/\kappa \) on the rhs of (9) goes to zero as \( \kappa \to 0 \). The parameter \( \kappa \) captures the alphabet sizes in the problem; our result applies to regions where \( \kappa \) is small. As \( |S| \leq n \), this is equivalent to demanding that \( \mathcal{X} \) is a large alphabet. Thus we will henceforth assume that the random variables \( X_S \) are large “blocks” (instead of individual bits).

It is instructive to apply this result to the \( \binom{n}{r} \)-subset sampler: Here the error probability \( \varepsilon \) decays exponentially with \( r \) for any fixed \( \xi \in [0, 1] \). More precisely, the following reformulation of (9) is obtained by setting \( \Delta = 3\xi + 2\kappa \log 1/\kappa \). We then have

\[ R_{\min}^\varepsilon(X_S|SE)_\rho \geq R_{\min}(X^n|E)_\rho - \Delta \quad \text{for any } \Delta \geq 2\kappa \log 1/\kappa, \]

\[ \varepsilon = e^{-\Omega(r(\Delta - 2\kappa \log 1/\kappa)^2)}. \]

Thus (smooth) min-entropy-rate is preserved up to a constant, with an exponentially small error \( \varepsilon \).
3.2 Related work

We briefly explain how our contribution relates to other known results. We stress that giving a comprehensive review of all the relevant areas is not the aim of this section. Nor do we attempt to provide a complete list of references; the pointers given here are mainly intended to facilitate access to further literature. We identify the following broad points of contact with previous work:

Quantum information about classical random variables: Random access encodings

Our main result is an upper bound on the amount of information a quantum system gives about certain classical values. As such, it fits into a long line of work, the most prominent example of which is Holevo’s upper bound on the accessible information [Hol73].

More specifically, our result bounds the information about a (randomly selected) substring $X_S = (X_{s_1}, \ldots, X_{s_m})$ of a classical string $X^n = (X_1, \ldots, X_n)$. In this sense, it is structurally identical to the random access encodings studied by Ambainis, Nayak, Ta-Shma and Vazirani [ANTSv99]. Formally, an $\binom{n}{r} \overset{\rho}{\rightarrow} m$ random access encoding maps $n$-bit strings $X^n = (X_1, \ldots, X_n)$ into $m$-qubit states $\rho_X$ in a way that allows to retrieve any (single) bit $X_i$ with probability at least $p$ by a measurement.

Strengthening the result of [ANTSv99], Nayak [Nay99] showed that at least $m \geq (1 - h(p)) n$ qubits are needed for this kind of encoding. (Here $h(\cdot)$ is the binary entropy function.) This can be understood as a precise expression of the qualitative statement that $m$ qubits cannot be used to store more than $m$ classical bits.

Recently, this result has been significantly generalized by Ben-Aroya, Regev and de Wolf [BAR07]. They studied $\binom{n}{r} \overset{\rho}{\rightarrow} m$ encodings, where the aim is to be able to retrieve each substring $X_S$ of length $r = |S|$ with probability at least $p$ from the $m$-qubit state. They showed that the success probability $p$ decreases exponentially in $r$ when $m < 0.7n$. The result [BAR07] of Ben-Aroya et al. is of the same form as ours. Indeed, as explained below, in terms of entropies, it expresses the fact that in the studied situation, the entropy-rate is preserved. However, there are at least three major differences to our work.

Firstly, [BAR07] provides an upper bound on the guessing probability $p(X_S | E)$, i.e., the probability of retrieving the correct value $X_S$ given quantum information $E$, which is the figure of merit in the context of random access encodings. By virtue of the identity $p(X_S | E) = 2^{-H_{\min}(X_S | E)}$ (see [KSR07] for more details), their result implies a lower bound on the min-entropy (and, hence, also on the smooth min-entropy for any $\varepsilon \geq 0$). In contrast, we derive a lower bound on the smooth min-entropy $H_{\min}^\varepsilon(X_S | E)$, which is the relevant quantity in the context of randomness extraction (e.g., in the bounded storage model). This, in turn, implies an upper bound on the guessing probability $p(X_S | E) \leq 2^{-H_{\min}^\varepsilon(X_S | E)} + \varepsilon$. Because our result is not optimized for very small $\varepsilon$, the upper bound on the guessing probability following from our result might be far below the bound of [BAR07]. On the other hand, the bound on the smooth min-entropy implied by the result of [BAR07] is below our bound, which is asymptotically optimal.

A second, apparently insignificant yet important difference between [BAR07] and our work is the alphabet size of the random variables $X_i$ in the tuple $X^n = (X_1, \ldots, X_n) \in \mathcal{X}^n$. While these are single bits in [BAR07], they may be random variables over a large alphabet in our work, i.e., every $X_i \in \{0, 1\}^c$ is itself a $c$-bit string for some (usually large) $c$. In the latter case, choosing a random subset $S \subset [n]$ of size $r = |S|$ effectively generates a substring $X_S = (X_{s_1}, \ldots, X_{s_r})$ of length $\ell = cr$ by blockwise sampling. For $c \gg \log n$, this procedure consumes only $\log \binom{n}{r} \leq r \log n \ll \ell$ random bits, in contrast to $\log \binom{n}{cr} \geq \ell$ when the individual bits are chosen at random.

9The notation used here is slightly different from the original papers.

10Note that our main result as stated in Corollary 6.19 does not directly apply to cases where the alphabet of the random variables $X_i$ is too small, e.g., if they are single bits. However, our result can be extended to these cases in the following way. Given, for instance, a $c$-bit string $B = (B_1, \ldots, B_N)$, the permuted string $B_\sigma := (B_{\sigma(1)}, \ldots, B_{\sigma(N)})$, for any permutation $\sigma \in S_N$, has the same min-entropy as $B$. We can therefore apply Corollary 6.19 to the permuted string $B_\sigma$ for a randomly chosen $\sigma$, and appropriately chosen partitioning $B_\sigma = (X_1, \ldots, X_n)$ into $n$ blocks, resulting in a substring $B' = X_S$ with high min-entropy. Since, after undoing the permutation on $B'$, this string is identically distributed as a bitstring chosen at random from $B$, we conclude that the min-entropy rate is essentially conserved under random sampling of bits.
When applied to the bounded storage model, this means that we can extract more bits than the number of initial (shared) key bits. On the other hand, while the sample-and-hash approach can in principle be applied using the result of [BARd07], the number of extracted bits is much smaller than the number of initial key bits, that is, no significant key expansion can be achieved.

Thirdly, the result of [BARd07] measures the initial quantum information about the string $X$ in terms of the number $m$ of qubits used in the encoding. More precisely, it is assumed that $X$ is uniformly distributed and that at most $m$ qubits containing information about $X$ are stored in a quantum system $E$ (formally, $H_0(E) \leq m$, where $H_0(E)$ denotes the logarithm of the dimension of $E$). In contrast, our result applies more generally to situations where merely a lower bound on the quantity $H_{\text{min}}(X|E)$ is known, while the quantum system $E$ may be arbitrarily large. The above special case where the dimension of $E$ is bounded follows from the general fact that $H_{\text{min}}(X|E) \geq H_{\text{min}}(X) - H_0(E)$.

**Key extraction: Extractors and privacy amplification**

The study of key extraction in the presence of a classical adversary is, as argued above, equivalent to the question of constructing randomness extractors (see [Sha02] for a survey of this intensely studied subject). More specifically, two-universal hashing was first applied to privacy amplification in [BBR88, BBCM95]. Maurer and Dziembowski [DM02, DM04] obtained optimal protocols for key extraction in the (classical) bounded storage model. Lu [Lu02] made the connection to locally (or on-line) computable strong extractors. Vadhan subsequently gave essentially optimal constructions by showing that sampling preserves min-entropy [Vad03]; the sampling approach for extracting randomness can be traced back to the work of Nisan and Zuckerman [NZ96] and abounds in the randomness extractor literature.

The situation in the presence of an adversary with prior quantum information is more intricate, and much less is known to date. On the negative side, Gavinsky, Kempe, Kerenidis, Raz and de Wolf [GKK+07] gave a surprising example of a classical extractor which fails to extract randomness in the presence of a quantum adversary (with a similar amount of quantum memory). On the positive side, Renner [Ren05] showed that two-universal hashing is optimal in the amount of extracted key (see also [RK05]). König and Terhal [KT07] showed that strong extractors with binary output also extract secure bits against quantum adversaries; this provides quantum extractors with short seeds, but does not achieve significant key expansion in the bounded storage model. Recently, new constructions of quantum extractors were proposed by Fehr and Schaffner [FS07]. While these extractors can be used for privacy amplification, their parameters are not suitable for the bounded storage model.

**4 Proof sketch**

In this section, we give an informal overview of the main ideas involved in the proof of the result (9). In Section 4.1, we give a simple proof of an analogous statement for the (classical) Shannon entropy. Our proof for (quantum) min-entropy mimics this line of argument, but differs in a few major points, as discussed below.

A few of our techniques may be of independent interest. A central idea is the splitting of a state into several components based on conditional operators; it leads to a modified chain-rule for min-entropies. We explain this in Section 4.2. The converse procedure which we call recombining is especially interesting when only subsets of the split states are used in the recombination. The outcome of such a partial recombination is a state which approximates the original state. By selecting split states in a systematic fashion, we can single out the high-entropy components of a state. As we explain in Section 4.3, this is a fundamental tool for showing that a given state has a certain amount of (smooth) min-entropy.

We will conclude this part of the paper with an overview of how these two procedures – the splitting and the recombining – can be combined with an argument about samplers to give the result we seek.
We stress that this section is introductory in nature, and the technical details are left to later sections. In particular, we will only argue qualitatively, and the formulas in Sections 4.2 and 4.3 are not meant to be taken literally. However, the basic structure of our arguments will be exactly as sketched here.

4.1 Proof idea

We show how to derive a modified statement related to (9), where we restrict our attention to probability distributions and where the min-entropy $H_{\text{min}}(A|B)$ is replaced by the (conditional) Shannon entropy $H(A|B) = H(AB) - H(B)$. (Here $H(X) = -\sum_{x \in X} P_X(x) \log P_X(x)$ denotes the usual Shannon entropy.) This kind of proof is sketched in [NZ96] to give an intuition why samplers are good condensers. However, neither the proofs in [NZ96] nor Vadhan's proof [Vad03] proceed along these lines.

The essential properties of the Shannon entropy used are the subadditivity property

$$H(A|BC) \leq H(A|B),$$

i.e., the fact that further conditioning can only reduce the entropy, and the chain-rule

$$H(AB|C) = H(A|BC) + H(B|C).$$

Consider a probability distribution $P_{X^n|E}$, where $X^n = (X_1, \ldots, X_n)$ is an $n$-tuple of random variables. Our aim is to show that with high probability over a randomly chosen subset $S \subset [n]$ of size $|S| = r$, the entropy of $H(X_S|E)$ is approximately equal to $\frac{r}{n} H(X^n|E)$.

To abbreviate the notation, we will define

$$X_{>j} = X_{j+1}X_{j+2} \cdots X_n$$
$$X_{<j} = X_1X_2 \cdots X_j \quad \text{for } j \in [n]$$
$$X_{>n} = \emptyset$$

for any such $n$-tuple. The first step is what we call a splitting step: The chain-rule (11) implies that the entropy $H(X^n|E)$ can be decomposed into its constituents,

$$H(X^n|E) = \sum_{i=1}^{n} \alpha_i \quad \text{where} \quad \alpha_i := H(X_i|X_{>i}E) \text{ for any } i \in [n].$$

In other words, we have split the entropy into a sum of individual components.

If we now select a subset $S \subset [n]$ of $r = |S|$ indices at random, then Chernoff’s inequality implies that the inequality

$$\frac{1}{r} \sum_{s \in S} \alpha_s \geq \frac{1}{n} H(X^n|E) - O(1/\sqrt{r}),$$

holds except with probability exponentially small in $r$. Note that this holds more generally for any $(n, \xi, \varepsilon)$-sampler $S$ with corresponding adaptations.

By strong subadditivity, we have

$$\alpha_j = H(X_j|X_{>j}E) \leq H(X_j|X_{>j}\cap S E) \quad \text{for any } j \in [n],$$

where $X_{>j}\cap S$ is the concatenation of all variables $X_i$ with $i > j$ and $i \in S$. With this inequality we can essentially eliminate all variables $X_i$ with $i \notin S$ from our inequalities.

The final step is what we call a recombination step: Using the chain rule once again, we obtain with (13)

$$H(X_S|E) = \sum_{s \in S} H(X_s|X_{>s}\cap S E) \geq \sum_{s \in S} \alpha_s.$$
In other words, we can get a lower bound on the joint entropy \( H(X_S|E) \) by combining the individual contributions \( s \in S \).

With (12), we conclude that with all but exponentially small probability, the (Shannon)-entropy rate is preserved when selecting a random subset.

The proof of our main result for min-entropy follows the same lines, with a modified chain-rule for min-entropies. Notice that the chain-rule in the form (11) can be seen as the combination of two inequalities,

\[
H(AB|C) \leq H(A|BC) + H(B|C), \quad (14)
\]
\[
H(AB|C) \geq H(A|BC) + H(B|C), \quad (15)
\]
both of which are used in the proof sketch. Indeed, the first inequality (14) allows us to divide the joint entropy \( H(X^n|E) \) into a sum of individual contributions, whereas the second inequality (15) provides a lower bound on the joint entropy \( H(X_S|E) \) in terms of its components. We refer to the first application as a splitting and the second application as a recombination step. For the min-entropy, these two steps are more involved; we do not only split and recombine entropies, but corresponding quantum states, as explained in the next section.

4.2 Towards a modified chain-rule: Entropy-splitting

The subadditivity property (10) is easily shown to hold for the min-entropy. Similarly, a recombination-chain-rule (15) can be proved for min-entropy. However, the splitting-chain rule (14) is no longer true for min-entropies and has to be replaced by a more subtle statement. This can be seen as a quantum version of the entropy splitting lemma proposed in [Wul07]. It is a major component of our proof and may be of independent interest.

To state this modified splitting-chain-rule, consider a state \( \rho_{ABC} \) with purification \( |\Psi_{ABCD}\rangle \).

We will construct a decomposition

\[
|\Psi_{ABCD}\rangle = \sum_{\alpha} |\Psi^\alpha_{ABCD}\rangle
\]

of \( |\Psi_{ABCD}\rangle \) into mutually orthogonal subnormalised states \( \{ |\Psi^\alpha_{ABCD}\rangle \}_\alpha \) such that

\[
H_{\text{min}}(A|BC)_{\rho^\alpha} + H_{\text{min}}(B|C)_{\rho^\alpha} \geq H_{\text{min}}(AB|C)_{\rho} \quad (17)
\]
for every \( \alpha \). In contrast to (14), this statement splits the entropy into a sum of individual entropies of states which are different from the original state \( |\Psi_{ABCD}\rangle \). They are, however, directly related to \( |\Psi_{ABCD}\rangle \) by (16); we call these states split states.

For technical reasons, it will be convenient to have a version of (17) which decomposes \( |\Psi_{ABCD}\rangle \) into a fixed number \( m \in \mathbb{N} \) of states. The indices \( \alpha \) are then from the set \( [m] := \{1, \ldots, m\} \), and (17) is replaced by

\[
H_{\text{min}}(A|BC)_{\rho^\alpha} + H_{\text{min}}(B|C)_{\rho^\alpha} \geq H_{\text{min}}(AB|C)_{\rho} - \frac{\Delta}{m} \quad \text{for all } \alpha \in [m], \quad (18)
\]

where \( \Delta \) is function of \( |\Psi_{ABCD}\rangle \) which can be bounded in situations of interest. (The exact statement is given as Corollary 5.5 below.) An important property of the split states is that each \( |\Psi^\alpha_{ABCD}\rangle \) is the result of applying a projection \( Q^\alpha_{AD} \) to \( |\Psi_{ABCD}\rangle \), where \( Q^\alpha_{AD} \) only acts non-trivially on systems \( A \) and \( D \).

4.3 (Partial) recombination of split states

Decomposing a state \( |\Psi_{ABC}\rangle \) into a sum of mutually orthogonal states \( \{ |\Psi^\alpha_{ABC}\rangle \}_{\alpha \in [m]} \) gives us a convenient way of bounding the (smooth) min-entropy of \( |\Psi_{ABC}\rangle \). The general procedure is as follows: Suppose for example that our aim is to bound the quantity \( H(A|B)_{\rho} \) from below. We will
show that if the entropy $H(A|B)_{ρ^α}$ is large for every split state $|Ψ^α⟩$, then the same is true for the quantity $H_{min}(A|B)_{ρ}$ (up to a correction of size log $m$, see Lemma 5.8 for a precise statement).

We can use this fact to show that a state $|Ψ_{ABC}⟩$ is close to a state $|Ψ^{'ABC}⟩$ with large min-entropy $H(A|B)_{ρ}$. We start from an arbitrary orthogonal decomposition of $|Ψ_{ABC}⟩$ of the form (16) into $m$ states $\{|Ψ^α⟩\}_{α∈[m]}$. We then identify a subset $Γ(λ)⊂[m]$ with the property that

$$H_{min}(A|B)_{ρ^α} ≥ λ \quad \text{for all } α ∈ Γ(λ).$$

We define the partially recombined state

$$|Ψ_{ABC}⟩ = \sum_{α∈Γ(λ)} |Ψ^α⟩.$$  

We can show that $H(A|B)_{ρ^α} ≥ λ$ is large. Moreover, since the states $\{|Ψ^α⟩\}_{α∈[m]}$ are assumed to be orthogonal, we can bound the distance of $|Ψ_{ABC}⟩$ to the original state $|Ψ_{ABC}⟩$ by an expression of the form $2\sqrt{1−ω(Γ(λ))}$, where $ω(Γ(λ))$ is the weight of $Γ(λ)$ under the probability distribution $ω(α) = tr[|Ψ^α⟩⟨Ψ^α|]_{ρ}$ on $[m]$. In this way, showing that the smooth min-entropy $H^*(A|B)_{ρ}$ of $|Ψ_{ABC}⟩$ is lower bounded by a value $λ$ reduces to showing that the corresponding set $Γ(λ)$ has a large weight under $ω$.

4.4 Putting it together: splitting, sampling and recombining

Let us now return to our original problem: Given a quantum state $ρ_{X^n E} = ρ_{X_1⋯X_n E}$ with purification $|Ψ⟩$, we would like to show that $H_{min}(X|S|E)_{ρ}$ is large with high probability over the choice of $S⊂[n]$. To illustrate the required steps in the proof, let us consider a simple example where $n = 4$.

The first step is to apply the splitting rule to $|Ψ⟩$, dividing the joint entropy $H_{min}(X^n|E)_{ρ}$ into a contribution from $X_1$ and the remainder. This gives $m$ states $|Ψ^α⟩_{α∈[m]}$ with the property that for all $α^1∈[m]$,

$$H_{min}(X^4|E)_{ρ} ≲ H_{min}(X_1|X_2|E)_{ρ^{α^1}} + H_{min}(X_3|X_4|E)_{ρ^{α^1}}.$$  

(Here $ρ^{α^1} = |Ψ^α⟩⟨Ψ^α|$ denotes the density operator corresponding to $|Ψ^α⟩$.) We then apply the splitting-chain-rule to each of these states in order to split $H_{min}(X_2|E)_{ρ^{α^1}} = H_{min}(X_3|X_4|E)_{ρ^{α^1}}$, into the contribution of $X_2$ and the remaining part. This results, for each $α^1∈[m]$, in a collection of states $\{Ψ^{α^1|α^2}|α^2∈[m]\}$ satisfying

$$H_{min}(X_2|E)_{ρ^{α^1|α^2}} ≲ H_{min}(X_3|X_4|E)_{ρ^{α^1|α^2}}.$$  

Finally, dividing the last term into contributions from $X_3$ and $X_4$, we get, for each $(α^1, α^2)∈[m]^2$, a family of states $\{Ψ^{α^1α^2|α^3}|α^3∈[m]\}$ such that

$$H_{min}(X_4|E)_{ρ^{α^1|α^2|α^3}} ≲ H_{min}(X_3|X_4|E)_{ρ^{α^1|α^2|α^3}}.$$  

This completes the splitting step. Summarising, we have obtained a collection of states starting from $|Ψ⟩$: Those states $\{Ψ^{α^1}\}_{α^1∈[m]}$ obtained by applying the splitting-chain-rule once, the states $\{Ψ^{α^1α^2}\}_{α^1,α^2∈[m]^2}$ corresponding to states that are the result of splitting twice and so on.

A useful geometric visualisation (which is, however, not essential for the proof) is obtained by placing these states at the vertices of an $m$-ary tree (in this case of depth 3). We place $|Ψ⟩$ at the root, and the descendants of each vertex are the split states obtained by splitting. Thus every 3-tuple $(α^1, α^2, α^3)∈[m]^3$ specifies a path with vertex labels $(|Ψ⟩, |Ψ^α⟩, |Ψ^{α^1|α^2}⟩, |Ψ^{α^1α^2|α^3}⟩)$ from the root to a leaf.

Let us combine inequalities (19)–(21) into

$$H_{min}(X^4|E)_{ρ} ≲ H_{min}(X_1|X_2|E)_{ρ^{α^1}} + H_{min}(X_2|X_3|E)_{ρ^{α^1|α^2}} + H_{min}(X_3|X_4|E)_{ρ^{α^1|α^2|α^3}}$$  

for all $α^3 = (α^1, α^2, α^3)∈[m]^3$. 

13
By attaching the entropies of interest to the edges of the mentioned tree, we can interpret this inequality as expressing the fact that the sum of the values of the edges along each path of the tree from the root to a leaf is lower bounded by $H_{\min}(X_4|E)_{\rho}$.

The next things to consider are the sampling- and recombination step. Our aim is to show that the smooth entropy $H_{\min}^K(X_S|E)_{\rho}$ is large (with high probability over the choice of the subset $S \subset [4]$). We follow the procedure outlined in the previous section. That is, we define the recombined state

$$|\tilde{\Psi}\rangle = \sum_{\alpha^3 \in \Gamma(\lambda, S)} |\Psi^{\alpha^3}\rangle$$

where $\Gamma(\lambda, S)$ is the set of paths $\alpha^3 \in [m]^3$ with the property that

$$\delta_{1 \in S} H_{\min}(X_1|X_{>1}E)_{\rho^{\alpha_1}} + \delta_{2 \in S} H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha_1, \alpha_2}} + \delta_{3 \in S} H_{\min}(X_3|X_{>3}E)_{\rho^{\alpha_1, \alpha_2, \alpha_3}} + \delta_{4 \in S} H_{\min}(X_4|E)_{\rho^{\alpha_1, \alpha_2, \alpha_3}} \geq \lambda.$$

In other words, we restrict our attention to paths (and corresponding states) which (when restricted to $S$), have large entropy. We then need to show the following:

(i) with high probability over the choice $S$, the state $|\tilde{\Psi}\rangle$ is close to $|\Psi\rangle$

(ii) the entropy $H(X_S|E)_{\tilde{\rho}}$ is large.

The proof of (i) again involves a bound of the form

$$\frac{1}{2} \|\langle \Psi | - | \tilde{\Psi} \rangle \| \leq \sqrt{1 - \omega(\Gamma(\lambda, S))},$$

where $\omega(\alpha^3) = \text{tr} |\Psi^{\alpha^3}\rangle \langle \Psi^{\alpha^3}|$, $\alpha^3 \in [m]^3$ is a (fixed) probability distribution on the leaves. We will show that a sampler has the following property, when applied to the situation described above (see Section 6.3): With high probability over the choice of $S$, the weight $\omega(\Gamma(\lambda, S))$ is large. More generally, we show how the sampler-property extends from a single sequence of values to the case of a matrix of values (in our case corresponding to edges of a tree).

The proof of (i) is done inductively using subadditivity, the recombination-chain-rule, and the recombination argument outlined above. For concreteness, suppose for example that $S = \{2, 4\}$. Then we have

$$H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha_1, \alpha_2}} + H_{\min}(X_4|E)_{\rho^{\alpha_1, \alpha_2, \alpha_3}} \geq \lambda$$

for all $\alpha^3 = (\alpha_1, \alpha_2, \alpha_3) \in \Gamma(\lambda, S) \subset [m]^3$. It is convenient to rephrase this as follows, writing $\alpha^3 = (\alpha^2, \alpha_3)$. We then have for all $\alpha^2 \in [m]^2$

$$H_{\min}(X_4|E)_{\rho^{\alpha_2, \alpha_3}} \geq \lambda - H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha_2}} \quad \text{for all } \alpha_3 \text{ with } (\alpha^2, \alpha_3) \in \Gamma(\lambda, S).$$

In particular, when we apply this to the (intermediate) partially recombined states

$$|\tilde{\Psi}^{\alpha^2}\rangle = \sum_{\alpha_3: (\alpha^2, \alpha_3) \in \Gamma(\lambda, S)} |\Psi^{(\alpha^2, \alpha_3)}\rangle,$$

we obtain

$$H_{\min}(X_4|E)_{\rho^{\alpha^2}} \geq \lambda - H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha^2}} \quad \text{for all } \alpha^2.$$

We will also use the fact that the recombined states satisfy $H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha^2}} \geq H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha^2}}$ (see Lemma 6.6(v)). Subadditivity gives $H_{\min}(X_2|X_4E)_{\rho^{\alpha^2}} \geq H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha^2}} \geq H_{\min}(X_2|X_{>2}E)_{\rho^{\alpha^2}}$ for all $\alpha^2 \in [m]$. With the previous two inequalities, we therefore get

$$H_{\min}(X_4|E)_{\rho^{\alpha^2}} + H_{\min}(X_2|X_4E)_{\rho^{\alpha^2}} \geq \lambda.$$
This in turn implies

\[ H_{\min}(X_2X_4|E)_{\rho^2} \gtrsim \lambda \quad \text{for all } \alpha^2 \in [m]^2 \]

by the recombination-chain-rule. Because \( |\bar{\Psi}\rangle \) can be written as sum of the states \( |\Psi^{\alpha^j}\rangle \), the recombination-procedure then gives

\[ H_{\min}(X_2X_4|E)_{\bar{\rho}} \gtrsim \lambda , \]

as claimed.

This line of argument can be followed more generally for a general subset \( S \subset [n] \). We will need intermediate (partially) recombined states \( \{|\hat{\Psi}^{\alpha^j}\rangle\}_{\alpha^j \in [m]^j, j \in [n]} \); these can again be thought of as being attached to the vertices of a tree. They are defined recursively, by recombining “good” states (i.e., those corresponding to prefixes of elements in \( \Gamma(\lambda, S) \)). In other words, when recombining, we work our way up the tree (omitting “bad” states, i.e., those with small entropies.)

This concludes our sketch proof; it is now time to elaborate on the details.

5 Rules and tools for min-entropy

In this section, we set the ground for our result concerning samplers. In particular, we formally introduce the conditional min-entropy \( H_{\min}(A|B)_{\rho} \) in Section 5.2. This will be done via an intermediate quantity \( H(A|B)_{\sigma} \). Most of our rules for min-entropy, the most basic of which are stated in Section 5.3, apply to these intermediate quantities; they will be our main object of study. In Section 5.4, we establish our central splitting-chain-rule.

5.1 Preliminaries

Throughout, we consider nonnegative operators acting on finite-dimensional Hilbert spaces (or systems) \( H_A, H_B, \ldots \) and their tensor products. We use subscripts to indicate which systems an operator acts on. We also use subscripts when we trace out systems, but sometimes make use of superscripts to denote “tracing out everything but”, in the following sense: for a tripartite state \( \rho_{ABC} \), we write \( \text{tr}_{BC}(\rho_{ABC}) = \rho_A \) for the reduced density operator on \( A \). As explained above, we sometimes abuse notation by omitting identities. For example, we will write operator inequalities such as

\[ \rho_{AB} \leq \sigma_B , \]

for a bipartite operator \( \rho_{AB} \) and an operator \( \sigma_B \) on \( H_B \). By this inequality, we simply mean \( \rho_{AB} \leq \id_A \otimes \sigma_B \) (which is defined by the condition that \( \id_A \otimes \sigma_B - \rho_{AB} \) is a nonnegative operator). More generally, when writing operators on multipartite systems, we omit identities whenever a unique meaningful statement can be obtained by tensoring corresponding identities to the operators. To give an example, we will write expressions such as

\[ Q_B \rho_{AB} Q_B \leq P_D \rho_{ABCD} P_D , \]

where the operators act on the spaces indicated by subscripts, instead of

\[ (\id_A \otimes Q_B \otimes \id_{CD})(\rho_{AB} \otimes \id_{CD})(\id_A \otimes Q_B \otimes \id_{CD}) \leq (\id_{ABC} \otimes P_D)\rho_{ABCD}(\id_{ABC} \otimes P_D) . \]

Basic properties of operator inequalities we need are their preservation under partial traces and the application of operators, i.e., the fact that \( \rho_{AB} \leq \sigma_{AB} \) implies that

\[ \rho_A \leq \sigma_A \]

and

\[ T_{AB} \rho_{AB} T_{AB}^\dagger \leq T_{AB} \sigma_{AB} T_{AB}^\dagger \]
for any operator $T_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$.

For two operators $\rho_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ and $\sigma_B$ on $\mathcal{H}_B$ such that the support of $\rho_B$ is contained in the support of $\sigma_B$, the conditional operator $\frac{\rho_{AB}}{\sigma_B}$ is defined as

$$
\frac{\rho_{AB}}{\sigma_B} := \sigma_B^{-1/2} \rho_{AB} \sigma_B^{-1/2} .
$$

(23)

Here $\sigma_B^{-1/2} := \sqrt{\sigma_B^{-1}}$, where $\sigma_B^{-1}$ is the generalised inverse of $\sigma_B$. An important property of conditional operators is that

$$
\frac{\rho_{AB}}{\sigma_B} = \text{tr}_C \left( \frac{\rho_{ABC}}{\sigma_B} \right)
$$

(24)

for any tripartite operator $\rho_{ABC}$.

We will say that a bipartite operator $\rho_{AE}$ on $\mathcal{H}_A \otimes \mathcal{H}_E$ is classical on $A$ (relative to an orthonormal basis $\{|a\rangle\}_{a}$ of $\mathcal{H}_A$) if it has the form $\rho_{AE} = \sum_a |a\rangle \langle a| \otimes \rho_{E}$, where $\rho_E$ is classical on $E$. Clearly, if $\rho_{AE}$ is classical on $A$ relative to $\{|a\rangle\}_{a}$, then so is $\rho'_{AE} = O_{AE} \rho_{AE} O_{AE}^\dagger$, for any operator of the form $O_{AE} = \sum_a |a\rangle \langle a| \otimes O_{E}$. It is easy to verify that this statement is still true when considering purifications and additional classical systems: If $|\Psi_{ABEF}\rangle$ is such that the reduced density operator $\rho_{ABE}$ is classical on both $A$ and $B$ (relative to some orthonormal bases) then the same is true for the state $O_{AE}|\Psi_{ABEF}\rangle$.

5.2 Definition of min-entropy

As already mentioned, every pair of operators $\rho_{AB}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ and $\sigma_B$ on $\mathcal{H}_B$ such that the support of $\rho_B$ is contained in the support of $\sigma_B$ give rise to a conditional operator $\frac{\rho_{AB}}{\sigma_B}$ We define the quantity $H(A|B)_{\sigma}$ as minus the logarithm of the maximal eigenvalue of this conditional operator, that is

$$
H(A|B)_{\sigma} := - \log \lambda_{\text{max}} \left( \frac{\rho_{AB}}{\sigma_B} \right) .
$$

In some sense, this can be read as “the entropy of $\rho_A$ when it is conditioned on $\sigma_B$”; in the case where $A$ is classical, the operator $\sigma_B$ is related to a measurement on $\mathcal{H}_B$ (which is supposed to reproduce the value on $A$, see [KSR07]).

Maximising this quantity over all nonnegative trace-one operators $\sigma_B$ whose support contains the support of $\rho_B$ gives the min-entropy of $A$ given $B$, defined as

$$
H_{\text{min}}(A|B)_{\rho} := \sup_{\sigma_B} H(A|B)_{\sigma} .
$$

(25)

This quantity has a simple operational interpretation, as will be shown in a forthcoming publication [KSR07]: it is equivalent to the maximal probability of guessing $A$ given $B$, in the case where $A$ is classical.

\footnote{Note that for $\sigma_B = \rho_B$, definition (23) coincides with the conditional operator $\rho_{A|B} = \frac{\rho_{AB}}{\rho_B}$ discussed, e.g., in [Le07].}

\footnote{The generalised inverse $\sigma^{-1}$ of an operator $\sigma$ is defined as the operator which has the same eigenspaces as $\sigma$ and eigenvalues $\lambda^{-1}$ on the eigenspace of $\sigma$ corresponding to the eigenvalue $\lambda > 0$.}

\footnote{This can be seen by decomposing the state as $|\Psi_{ABEF}\rangle = \sum_{a,b} |a\rangle \langle b| \phi_{E}^{a,b}$. Classicality of the state $\rho_{ABE}$ on $A$ and $B$ then implies that $\text{tr}_F (|\phi_{E}^{a,b}\rangle \langle \phi_{E}^{a',b'}|) = 0$ whenever $(a,b) \neq (a',b')$. The claim can then be deduced from the fact that $\text{tr}_F (O_{K}^{a,b} |\phi_{E}^{a,b}\rangle \langle \phi_{E}^{a',b'}| (O_{K}^{a,b})^\dagger) = O_{K}^{a,b} \text{tr}_F (|\phi_{E}^{a,b}\rangle \langle \phi_{E}^{a',b'}|) (O_{K}^{a,b})^\dagger$.}

\footnote{In the following, we will always assume that the support of $\rho_B$ is contained in the support of $\sigma_B$, such that the conditional operator is well defined.}

\footnote{All logarithms log are binary; natural logarithms will be denoted by ln.}

\footnote{In Section 2.4 the quantity $H_{\text{min}}(A|B)_{\rho}$ was introduced without explicit reference to the intermediate quantities $H(A|B)_{\sigma}$.}
While [25] is ultimately the quantity of interest, the intermediate quantities $H(A|B)_\rho$ are easier to manipulate, and satisfy various useful rules. As we will see below, most of these follow more or less directly from the alternative characterisation

$$H(A|B)_\rho := - \log \min\{ \lambda : \frac{\rho_{AB}}{\sigma_B} \leq \lambda \cdot \id_{AB} \}$$

in terms of a family of operator inequalities.

For consistency reasons, it is convenient to set

$$H(A|B)_\rho := H(A|B)_\rho$$

$$H(\emptyset|B)_\rho := - \log \min\{ \lambda : \frac{\rho_B}{\sigma_B} \leq \lambda \cdot \id \},$$

where $\rho_B = \sigma_B^{-1/2} \rho_B \sigma_B^{-1/2}$. Note that the latter quantity is equal to $H(A|B)_\rho$ if the Hilbert space $\mathcal{H}_A$ corresponding to system $A$ is trivial, i.e., $\mathcal{H}_A \cong \mathbb{C}$. We can think of the quantity $H(B)_\rho$ as a conditional entropy obtained by adjoining a trivial system to $B$ using the isomorphism $\mathcal{H}_B \cong \mathcal{H}_B \otimes \mathbb{C}$. Informally, this corresponds to a situation where we condition “nothing” on $\sigma_B$; formally, it will turn out to be convenient to define $H(\emptyset|B)_\rho := H(\emptyset|B)_\rho$.

Finally, we will also (formally) encounter situations where $\rho_{AB} = 0$; in these cases, we formally set $H(A|B)_\rho = \infty$, $H(A|B)_\rho = \infty$, meaning that an arbitrarily large value can be assigned to these quantities in any identity where they appear.

For a parameter $\varepsilon \geq 0$, the $\varepsilon$-smooth min-entropy of $A$ given $B$ is equal to (cf. [Ren05])

$$H^{\varepsilon}_{\min}(A|B)_\rho = \sup_{\rho_{AB}: \|\rho_{AB} - \rho_B\|_{tr} \leq \varepsilon} H_{\min}(A|B)_\rho,$$

where the supremum is over all nonnegative operators $\tilde{\rho}_{AB}$ with trace bounded by 1 in an $\varepsilon$-ball around $\rho_{AB}$. (Here $\|A\| = \text{tr} \sqrt{AA^\dagger}$ is the $L_1$-norm.)

As already mentioned, the (smooth) entropy $H^{\varepsilon}_{\min}(X|E)$ captures the number of secret bits extractable from $X$ with respect to an adversary holding $E$. The following lemma justifies this operational interpretation.

**Lemma 5.1.** Consider a state $\rho_{XE}$ where $X$ is classical. Let $\rho_{U(0,1)\ell}$ denote the completely mixed state on $\{0,1\}^\ell$. Then

(i) For any $\ell \leq H_{\min}(X|E) - 2 \log 1/\varepsilon$, there is a function $f : \mathcal{X} \times \mathcal{X} \to \{0,1\}^\ell$ (independent of $\rho_{XE}$) which extracts an $\ell$-bit string $Z = f(X,Y)$ from $X$, such that $Z$ is $\varepsilon$-close to uniform and independent of $(E,Y)$, where $Y$ is a uniform and independent seed. In formulae, we have

$$\frac{1}{2} \| \rho_{f(X,Y)E} - \rho_{U(0,1)\ell} \otimes \rho_Y \otimes \rho_E \| \leq \varepsilon.$$  

(ii) For any function $f : \mathcal{X} \to \{0,1\}^\ell$ and $\varepsilon \geq 0$, the inequality

$$\frac{1}{2} \| \rho_{f(X)E} - \rho_{U(0,1)\ell} \otimes \rho_E \| \leq \varepsilon$$

implies

$$H^{2\varepsilon}_{\min}(X|E)_\rho \geq \ell.$$  

**Proof.** Statement (i) is a reformulation of the fact that the two-universal hashing construction is a quantum extractor, as shown by Renner [Ren05].

For the proof of (ii), let $\rho_{SE} := \rho_{U(0,1)\ell} \otimes \rho_E$. Then, obviously

$$H_{\min}(S|E)_\rho \geq H(S|E)_\rho = \ell.$$
Because $\frac{1}{2}\|\bar{\rho}_{SE} - \rho_{f(X)E}\| \leq \epsilon$, this implies
\[
H_{\min}^{2\epsilon}(f(X)|E)_\rho \geq H_{\min}(S|E)_{\bar{\rho}} \geq \ell.
\]
Since the min-entropy can only decrease when applying a function (see [Ren05]), we conclude that
\[
H_{\min}^{2\epsilon}(X|E)_\rho \geq H_{\min}^{2\epsilon}(f(X)|E)_\rho \geq \ell,
\]
as desired. 

\[\square\]

5.3 Some basic rules and properties

We now summarise a few basic rules for the quantities $H(A|B)_\varrho$, which directly follow from using standard properties of operator inequalities, as described in Section 5.1.

Lemma 5.2 (Properties of min-entropy). The min-entropy satisfies the following.

(i) (Positivity for classical systems) Let $\rho_{AB} = \sum_a |a\rangle\langle a| \otimes \rho_B^a$ be classical on $A$. Then $H(A|B)_\rho \geq 0$.

(ii) (Dimension bound) For any $\rho_{AB}$ and $\sigma_B$ with $\sigma_B \leq \rho_B$, we have $H(A|B)_\varrho \leq H_0(A)$, where $H_0(A) = \log \dim \mathcal{H}_A$. In particular, $H(A|B)_\rho \leq H_0(A)$. More generally $H(B|C)_\varrho \geq H(AB|C)_\varrho - H_0(A)$ for any $\rho_{ABC}$ and $\sigma_C$.

(iii) (Subadditivity) $H(A|B)_\varrho \geq H(A|BC)_\varrho$ for any $\rho_{ABC}$ and $\sigma_{BC}$.

(iv) (Recombination-chain-rule) $H(AB|C)_\varrho \geq H(A|BC)_\rho + H(B|C)_\varrho$ for any $\rho_{ABC}$ and $\sigma_C$.

Proof. (i) directly follows from $|a\rangle\langle a| \otimes \rho_B^a \leq \rho_B$ for all $a$.

For the proof of the first part of (ii), we simply take the trace on both sides of the inequality $\rho_{AB} \leq 2^{-H(A|B)_\varrho} \sigma_B$ to get $\text{tr}(\rho_{AB}) \leq 2^{-H_0(A) - H(A|B)_\varrho} \text{tr}(\sigma_B)$, which gives the claim because $\text{tr}(\sigma_B) \leq \text{tr}(\rho_B) = \text{tr}(\rho_{AB})$. For the proof of the second part of (ii), observe that we have $\rho_{ABC} \leq 2^{-H(AB|C)_{\varrho}} \sigma_C$ by definition. Tracing out the system $A$ gives
\[
\rho_{BC} \leq 2^{-H(AB|C)_{\varrho}} \sigma_C.
\]
The claim (ii) then follows from the definition of $H(B|C)_{\varrho}$.

Similarly, (iii) directly follows by tracing out $C$ from the inequality
\[
\rho_{ABC} \leq 2^{-H(A|BC)_{\varrho}} \sigma_{BC}.
\]

For the proof of (iv), observe that
\[
\rho_{ABC} \leq 2^{-H(AB|C)_{\varrho}} \rho_{BC} \leq 2^{-H(AB|C)_{\varrho} - H(B|C)_{\varrho}} \sigma_{BC}.
\]
The claim follows from the definition of $H(AB|C)_{\varrho}$. 

We point out that (ii) and (iii) directly translate into the statements
\[
H_{\min}(B|C)_\rho \geq H_{\min}(AB|C)_\rho - H_0(A)
\]
\[
H_{\min}(A|B)_\rho \geq H_{\min}(A|BC)_\rho
\]
for the min-entropy. An analogous statement cannot be made for the recombination-chain-rule (iv), and we will have to retain the dependence on $\sigma_C$ in our arguments.

Having established subadditivity and a recombination-chain-rule, we will address the problem of finding a converse splitting-chain-rule in the next section. Before doing so, however, we will
mention another property of the min-entropy which will be important for our purposes. This is the fact the entropy of a state $|\Psi'\rangle = Q|\Psi\rangle$ obtained by applying a projection to a state $|\Psi\rangle$ is lower bounded by the entropy of the original state. We will later see that this allows us to retain information about the entropy when going from a state to its split descendants.

Note that this statement is not generally true, but depends crucially on where the projection acts.

**Lemma 5.3 (Monotony under projections).** Let $|\Psi_{ABC}\rangle$ be a pure state, let $Q_C$ be an operator on $C$ and let $|\Psi'_{ABC}\rangle = Q_C|\Psi_{ABC}\rangle$. Let $\rho_{ABC}$ and $\rho'_{ABC}$ be the corresponding density operators. Then

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

Furthermore, if $Q_C$ is a projector, then

$$H(A|B)_{\rho} \geq H(A|B)_{\tilde{\sigma}} \quad \text{and} \quad H(\emptyset|B)_{\rho} \geq H(\emptyset|B)_{\tilde{\sigma}}$$

for arbitrary $\sigma = \sigma_C$.

**Proof.** To prove the first inequality, let $Q_C$ be arbitrary. Applying $Q_C$ from the left and $Q_C^\dagger$ from the right to both sides of the inequality

$$\rho_{ABC} \leq 2^{-H(A|BC)}\rho_{BC}$$

gives

$$\rho'_{ABC} \leq 2^{-H(A|BC)}\rho'_{BC}$$

by definition of $|\Psi'_{ABC}\rangle$ and the properties of the partial trace. This proves the first inequality.

Let now $Q_C$ be a projector, and let $|\varphi_{AB}\rangle \in AB$ be arbitrary. Then

$$\text{tr}(|\varphi_{AB}\rangle\langle\varphi_{AB}| \rho'_{AB}) = \text{tr}((|\varphi_{AB}\rangle\langle\varphi_{AB}| \otimes \text{id}_C) \rho'_{ABC})$$

$$= \text{tr}((|\varphi_{AB}\rangle\langle\varphi_{AB}| \otimes Q_C) \rho_{ABC})$$

by the cyclicity of the trace and the fact that $Q_C$ is a projector. In particular, with $Q_C^\perp = \text{id}_C - Q_C$ denoting the projector onto the orthogonal complement of the image of $Q_C$, we have

$$\text{tr}(|\varphi_{AB}\rangle\langle\varphi_{AB}| (\rho_{AB} - \rho'_{AB})) = \text{tr}((|\varphi_{AB}\rangle\langle\varphi_{AB}| \otimes Q_C^\perp) \rho_{ABC}) \geq 0.$$  

We conclude that $\rho'_{AB} \leq \rho_{AB}$. In particular,

$$\rho'_{AB} \leq \rho_{AB} \leq 2^{-H(A|B)}\tilde{\sigma}_B \quad \text{and} \quad \rho'_B \leq \rho_B \leq 2^{-H(\emptyset|B)}\tilde{\sigma}_B$$

which implies the claim.  

### 5.4 Entropy-splitting: A splitting-chain-rule for min-entropy

To introduce our splitting-chain-rule, we proceed in two steps: In Section 5.4.1 we show a simplified version which does not restrict the number of states the original state is split into. As this is irrelevant for the remainder of our proof, this section can be skipped; however, it nicely illustrates the relevant features. The case of interest, where we split a given state into a fixed number $m$ of states, can be seen as a coarse-graining of the former. It will be the topic of Section 5.4.2.
5.4.1 A warm-up

The chain-rule we will prove in this section concerns a tripartite state $\rho_{ABC}$ with purification $\rho'_{ABCD} = |\Psi_{ABCD}\rangle \langle \Psi_{ABCD}|$ and an operator $\sigma_C$. We will show that we can split $|\Psi_{ABCD}\rangle$ into a sum of states $\{ |\Psi_{\alpha ABCD}\rangle \}_{\alpha}$ as in (16), in a way that

$$H(A|BC)_{\rho^\alpha} + H(B|C)_{\rho^\alpha} \geq H(AB|C)_{\rho^\alpha} \quad \text{for all } \alpha ,$$

(27)

where $\rho^\alpha_{ABCD} = |\Psi^\alpha_{ABCD}\rangle \langle \Psi^\alpha_{ABCD}|$. Note that by taking the supremum over $\sigma_B$, we immediately obtain the inequality

$$H_{\min}(A|BC)_{\rho^\alpha} + H_{\min}(B|C)_{\rho^\alpha} \geq H_{\min}(AB|C)_{\rho}$$

from (27). However, (27) makes a stronger assertion, and we will generally deal with statements of this form.

For the proof of (27), consider the eigendecomposition $\rho_{BC} \sigma_C = \sum_{\alpha} \alpha P^\alpha_{BC}$ of the conditional operator, where $P^\alpha_{BC}$ is the projector onto the eigenspace corresponding to the eigenvalue $\alpha$.

We will use the operators $P^\alpha_{BC}$ to define our split states, which will be labeled by the spectrum of $\rho_{BC} \sigma_C$. Clearly, if we apply $P^\alpha_{BC}$ on both sides of the operator $\rho_{BC} \sigma_C$, we end up with an operator which has a single non-zero eigenvalue $\alpha$. While $P^\alpha_{BC} \rho_{BC} \sigma_C$ thus has a very simple form, it is very different in nature from the original “unconditional” operator $\rho_{BC}$. Intuitively, it therefore makes sense to multiply by $\sigma_C$. The appropriate definition of $|\Psi^\alpha_{ABCD}\rangle$ turns out to be just the result of this, i.e., we can define

$$|\Psi^\alpha_{ABCD}\rangle = \sigma_C^{1/2} P^\alpha_{BC} \rho_{BC} \sigma_C^{-1/2} |\Psi_{ABCD}\rangle .$$

It is easy to check that these states decompose $|\Psi_{ABCD}\rangle$ as in (16). They also satisfy (27), as we will show now. First observe that $\rho^\alpha_{BC} = \sigma_C^{1/2} P^\alpha_{BC} \rho_{BC} \sigma_C^{-1/2} P^\alpha_{BC} \sigma_C^{1/2}$ by their very definition, and thus

$$\rho^\alpha_{BC} = \alpha \cdot \sigma_C^{1/2} P^\alpha_{BC} \sigma_C^{1/2} .$$

In particular, we have

$$\frac{\rho^\alpha_{BC} \sigma_C}{\sigma_C} = \alpha P^\alpha_{BC} ,$$

(28)

which implies that

$$H_{\min}(B|C)_{\rho^\alpha_{BC}} = - \log \alpha .$$

(29)

Combining (28) and (29) gives the statement

$$P^\alpha_{BC} \leq 2^{H_{\min}(B|C)_{\rho^\alpha_{BC}}} \rho^\alpha_{BC} \sigma_C ,$$

(30)

By definition of the quantity $H(AB|C)_{\rho}$, we also have $\frac{\rho_{ABC}}{\sigma_C} \leq 2^{-H(AB|C)_{\rho}}$. Applying the projector $P^\alpha_{BC}$ on both sides of this inequality leads to

$$P^\alpha_{BC} \frac{\rho_{ABC}}{\sigma_C} P^\alpha_{BC} \leq 2^{-H(AB|C)_{\rho}} P^\alpha_{BC} \leq 2^{H_{\min}(B|C)_{\rho^\alpha} - H(AB|C)_{\rho}} \rho^\alpha_{BC} \sigma_C .$$

---

\(^{17}\)As above, we assume that the support of $\sigma_C$ contains the support of $\rho_C$, hence, $\sigma_C$ is invertible on the relevant subspace.
Multiplying this inequality from both sides by $\sigma_C^{1/2}$ immediately gives the desired statement \([27]\), since $\rho^*_{ABC} = \rho^{1/2} P^*_{BC} \rho^{1/2} P^*_{BC} \rho^{1/2}$.

This concludes the proof of our simplified statement, where a state $|\Psi_{ABCD}\rangle$ is split into a family $\{|\Psi^\alpha_{ABCD}\rangle\}$, each of which obeys the splitting-chain-rule inequality \([27]\). The number of states is determined by the number of different eigenvalues of the operator $\rho_{BC}$; indeed, each state $|\Psi^\alpha_{ABCD}\rangle$ corresponds to an eigenvalue $\alpha$.

Before continuing, let us show the following useful properties of the states $\{|\Psi^\alpha_{ABCD}\rangle\}$. They are mutually orthogonal, and each state $|\Psi^\alpha_{ABCD}\rangle = Q^\alpha_{AD} |\Psi_{ABCD}\rangle$ is the result of applying a projection $Q^\alpha_{AD}$ (which acts non-trivially only on $A$ and $D$) to $|\Psi_{ABCD}\rangle$. This statement is the result of using the complementarity property that is inherent in quantum states.

First observe that $\sigma_C^{-1/2} |\Psi_{ABCD}\rangle$ is a purification of the conditional operator $\frac{\rho_{BC}}{\sigma_C}$. Using the Schmidt-decomposition, we can write

$$\sigma_C^{-1/2} |\Psi_{ABCD}\rangle = \sum_\alpha \sqrt{\alpha} |\alpha\rangle_{AD} |\alpha\rangle_{BC},$$

where $\{|\alpha\rangle_{AD}\}$ and $\{|\alpha\rangle_{BC}\}$ are eigenvectors with eigenvalue $\alpha$ of $\frac{\rho_{AD}}{\sigma_C}$ and $\frac{\rho_{BC}}{\sigma_C}$, respectively (slightly abusing notation, we omit multiplicities). We can define $Q^\alpha_{AD}$ as the projector onto the eigenspace of $\frac{\rho_{AD}}{\sigma_C}$ corresponding to the eigenvalue $\alpha$. We then clearly have

$$P^\alpha_{BC} \sigma_C^{-1/2} |\Psi_{ABCD}\rangle = Q^\alpha_{AD} \sigma_C^{-1/2} |\Psi_{ABCD}\rangle$$

for every $\alpha$. Since $\sigma_C^{1/2}$ and $Q^\alpha_{AD}$ act on different systems, they commute, and we obtain

$$|\Psi^\alpha_{ABCD}\rangle = \sigma_C^{1/2} Q^\alpha_{AD} \sigma_C^{-1/2} |\Psi_{ABCD}\rangle = Q^\alpha_{AD} |\Psi_{ABCD}\rangle,$$

as claimed. The orthogonality of these states is now immediate.

### 5.4.2 Splitting into a fixed number of states

In this section, we show that the construction discussed in Section \([5.4.1]\) can be adapted to yield a fixed number $m$ of states. This is quite straightforward: We simply divide the spectrum of the conditional operator $\frac{\rho_{BC}}{\sigma_C}$ into $m$ different intervals $[\mu_{\alpha-1}, \mu_\alpha]$, for $\alpha \in [m]$. Instead of projecting onto the eigenspace corresponding to a single eigenvalue, we use projectors $P^\alpha_{BC}$ onto the direct sum of eigenspaces associated with eigenvalues in the corresponding interval.

**Lemma 5.4** (Entropy splitting). Let $\rho_{ABCD} = |\Psi_{ABCD}\rangle \langle \Psi_{ABCD}|$ be a pure state and let $\sigma_C$ be a nonnegative operator. Let $h_0 \leq h_1 \leq \cdots \leq h_m$ be an $(m + 1)$-tuple of monotonically increasing real values with minimum and maximum given by

$$h_0 := H(B|C)_\rho,$$

$$h_m := H(AB|C)_\rho - H(A|BC)_\rho.$$

Then there are mutually orthogonal projectors $\{Q^\alpha_{AD}\}_{\alpha \in [m]}$ with the property that

$$H(A|BC)_{\rho^\alpha} \geq H(A|BC)_{\rho} - h_\alpha \quad (31)$$

$$H(B|C)_{\rho^\alpha} \geq h_{\alpha-1} \quad (32)$$

where $\rho^\alpha_{ABCD} = |\Psi^\alpha_{ABCD}\rangle \langle \Psi^\alpha_{ABCD}|$ is defined as

$$|\Psi^\alpha_{ABCD}\rangle := Q^\alpha_{AD} |\Psi_{ABCD}\rangle.$$

An alternative expression for these states is

$$|\Psi^\alpha_{ABCD}\rangle := \sigma_C^{1/2} P^\alpha_{BC} \sigma_C^{-1/2} |\Psi_{ABCD}\rangle.$$

(34)
where \( \{ P_{BC}^{\alpha} \}_{\alpha \in [m]} \) are mutually orthogonal projectors. They satisfy

\[
\sum_{\alpha \in [m]} |\Psi^\alpha\rangle = |\Psi\rangle
\]

(35)

Note that, according to the chain rule (Lemma 5.2 (iv)), \( h_0 \leq h_m \), i.e., there always exists a tuple of reals as defined in the lemma.

**Proof.** The proof of this statement is almost identical to the proof given in Section 5.4.1 but given here for completeness. For any \( \alpha \in [m-1] \), define \( \mu_\alpha := 2^{-h_\alpha} \), and let \( \mu_0 := \infty \), \( \mu_m := -\infty \). Note that \( (\mu_\alpha)_{\alpha = 0}^m \) is a monotonically decreasing sequence of values.

Consider the Schmidt-decomposition

\[
\sigma_C^{-1/2} |\Psi_{ABCD}\rangle = \sum_\lambda \sqrt{\lambda} |\lambda\rangle_{BC} |\lambda\rangle_{AD}
\]

(36)

of the “conditional” state \( \sigma_C^{-1/2} |\Psi_{ABCD}\rangle \) (the sum may include multiplicities). For every \( \alpha \in [m] \), we define the projectors \( P_{BC}^{\alpha} \) and \( Q_{AD}^\alpha \) as

\[
P_{BC}^{\alpha} = \sum_{\lambda \in [\mu_\alpha, \mu_{\alpha-1}]} |\lambda\rangle \langle \lambda|_{BC}
\]

\[
Q_{AD}^\alpha = \sum_{\lambda \in [\mu_\alpha, \mu_{\alpha-1}]} |\lambda\rangle \langle \lambda|_{AD} .
\]

By definition, these operators satisfy (33) and (34). Moreover, using the fact that \( Q_{AD}^\alpha \) commutes with \( \sigma_C^{-1/2} \), we conclude that (33) and (34) define the same state \( |\Psi_{ABCD}^\alpha\rangle \).

Since \( P_{BC}^{\alpha} \) is the projector onto the eigenspaces of \( \frac{\rho_{BC}}{\sigma_C} \) which belong to the eigenvalues in \( [\mu_\alpha, \mu_{\alpha-1}] \) for every \( \alpha \in [m] \), we have

\[
\mu_\alpha P_{BC}^{\alpha} \leq P_{BC}^{\alpha} \frac{\rho_{BC}}{\sigma_C} P_{BC}^{\alpha} \leq \mu_{\alpha-1} P_{BC}^{\alpha} .
\]

(37)

We show that

\[
P_{BC}^{\alpha} \frac{\rho_{BC}}{\sigma_C} P_{BC}^{\alpha} \leq 2^{-h_\alpha-1} P_{BC}^{\alpha} \quad \text{for all } \alpha \in [m] .
\]

(38)

This follows directly from (37) for \( \alpha \geq 2 \) since \( 2^{-h_\alpha-1} = \mu_{\alpha-1} \); for \( \alpha = 1 \), it is a consequence of the fact that the eigenvalues of \( \frac{\rho_{BC}}{\sigma_C} \) are upper bounded by \( 2^{-h_0} \) by definition of \( h_0 \).

Claim (32) now directly follows from (38) and the fact that \( \frac{\rho_{BC}}{\sigma_C} = P_{BC}^{\alpha} \frac{\rho_{BC}}{\sigma_C} P_{BC}^{\alpha} \).

Next we show that

\[
P_{BC}^{\alpha} \frac{\rho_{ABC}}{\sigma_C} P_{BC}^{\alpha} \leq 2^{-H(ABC)} + h_\alpha P_{BC}^{\alpha} \frac{\rho_{BC}}{\sigma_C} P_{BC}^{\alpha} \quad \text{for all } \alpha \in [m] .
\]

(39)

We distinguish two cases: For \( \alpha = m \), identity (39) is equivalent to

\[
P_{BC}^{\alpha} \frac{\rho_{ABC}}{\sigma_C} P_{BC}^{\alpha} \leq 2^{-H(A|BC)} + P_{BC}^{\alpha} \frac{\rho_{BC}}{\sigma_C} P_{BC}^{\alpha}
\]

because of the definition of \( h_m \). But this directly follows from \( \rho_{ABC} \leq 2^{-H(A|BC)} \rho_{BC} \) by multiplication from both sides with \( P_{BC}^{\alpha} \sigma_C^{-1/2} \) and its adjoint.

For \( 1 \leq \alpha < m \), we use the fact that the first inequality of (37) is equivalent to

\[
2^{-h_\alpha} P_{BC}^{\alpha} \leq P_{BC}^{\alpha} \frac{\rho_{BC}}{\sigma_C} P_{BC}^{\alpha}
\]

because of the definition of \( h_m \). But this directly follows from \( \rho_{ABC} \leq 2^{-H(A|BC)} \rho_{BC} \) by multiplication from both sides with \( P_{BC}^{\alpha} \sigma_C^{-1/2} \) and its adjoint.

For 1 ≤ α < m, we use the fact that the first inequality of (37) is equivalent to
Figure 1: This figure illustrates the basic building block for our arguments (cf. Corollary 5.5). A state $|\Psi\rangle$ can be decomposed into a sum of $m$ orthogonal states; in the figure, $m = 2$. We illustrate this by a tree; the original state sits at the root, whereas the split states sit at nodes labeled by $\alpha \in [m]$. The state at the root is the sum of its descendants, which are identical to the leaves in this case. Going from a node to its descendants is achieved by applying corresponding projection operators.

since $2^{-h_\alpha} = \mu_\alpha$. Substituting this into the inequality

$$P_{BC}^{\alpha} \frac{\rho_{ABC}}{\sigma_C} P_{BC}^{\alpha} \leq 2^{-H(AB\mid C)} \hat{\rho} P_{BC}^{\alpha}.$$ (which directly follows from $\frac{\Delta_{ABC}}{\sigma_C} \leq 2^{-H(AB\mid C)} \hat{\rho}$) immediately gives (39) for all $1 \leq \alpha < m$. This concludes the proof of the auxiliary statement (39).

The proof of (31) is now straightforward, based on (39). Multiplying the latter inequality by $\sigma_C^{-1/2}$ from both the left and the right yields

$$\rho_{ABC}^{\alpha} \leq 2^{-H(AB\mid C)} \hat{\rho} + h_\alpha \rho_{BC}.$$ which implies (31). \qed

In the previous lemma, we did not specify the intervals $[h_{\alpha-1}, h_\alpha]$ that are used to partition the spectrum of the conditional operator $\frac{\Delta_{ABC}}{\sigma_C}$. A simple choice is to partition the spectrum into $m$ intervals of equal length. This results in the following splitting-chain-rule, which will be our basic tool in what follows.

Corollary 5.5. Let $\rho_{ABCD} = |\Psi_{ABCD}\rangle\langle\Psi_{ABCD}|$ be a pure state and let $\sigma_C$ be a nonnegative operator. Then for any $m \in \mathbb{N}$

$$H(A\mid BC)_{\rho^n} + H(B\mid C)_{\rho^n} \geq H(AB\mid C)_{\hat{\rho}} - \frac{\Delta}{m}$$

where

$$\Delta := H(AB\mid C)_{\hat{\rho}} - H(A\mid BC)_{\rho} - H(B\mid C)_{\hat{\rho}}.$$ and where $\rho_{ABCD}^{\alpha} = |\Psi_{ABCD}^{\alpha}\rangle\langle\Psi_{ABCD}^{\alpha}|$ is defined by (33) or (34) in terms of families of mutually orthogonal projectors $\{Q_{AD}^\alpha\}_{\alpha \in [m]}$ and $\{P_{BC}^\alpha\}_{\alpha \in [m]}$, as in Lemma 5.4.

We are usually able to obtain a bound on $\Delta$; for a comparatively large value of $m$, we therefore get an approximation of (27), which is a converse to the recombination-chain-rule (Item IV of Lemma 5.2).

Proof. Here we choose $h_\alpha = h_0 + \alpha \frac{\Delta}{m}$ for all $\alpha \in \{0, \ldots, m\}$. \qed

We point out that the statement of Corollary 5.5 is also valid with $B$ removed from all expressions. This is because we can always adjoin a trivial system $B$ with Hilbert space $\mathcal{H}_B \equiv \mathbb{C}$.

For later use, we establish a few additional properties of the states $|\Psi_{ABCD}^{\alpha}\rangle$. We first show that the states $|\Psi_{ABCD}^{\alpha}\rangle$ have the same classicality properties as the original state $|\Psi_{ABCD}\rangle$. 23
Remark 5.6 (Preservation of classicality properties). Suppose that $D = D_1 D_2$ is bipartite, and that $\rho_{A_1 B_1 C_1 D_1}$ is classical on $A, B,$ and $D_1$ (relative to some orthonormal bases of these subsystems). Then $\rho_{A_1 B_1 C_1 D_1}$ is classical on $A, B,$ and $D_1$ (relative to the same bases), for any $\alpha \in \{ m \}$.

Proof. According to the discussion at the end of Section 5.1 about classical states and (34), it suffices to show that the operator $\sigma_{C_{1/2}} P_{BC}^{a} \sigma_{C_{1/2}}^{-1}$ has the form

$$
\sigma_{C_{1/2}}^{1/2} P_{BC}^{a} \sigma_{C_{1/2}}^{-1/2} = \sum_{b} |b \rangle \langle b| \otimes O_{C}^{b},
$$

(40)

for some operators $\{ O_{C}^{b} \}_{b}$ on $C$, where $\{ |b \rangle \}_{b}$ is the eigenbasis of $\rho_{B}$.

Because $\sigma_{C_{1/2}}^{-1}$ acts only on $C$, the state $\sigma_{C_{1/2}}^{-1} | \Psi_{ABCD} \rangle$ is classical on $B$ when tracing out $A$ and $D$, i.e.,

$$
\frac{\rho_{BC}}{\sigma_{C}} = \sum_{b} |b \rangle \langle b| B \otimes \theta_{C}^{b},
$$

for some nonnegative operators $\{ \theta_{C}^{b} \}_{b}$ on $C$, because this is true for the original state $| \Psi_{ABCD} \rangle$ by assumption. In particular, the eigenvectors of $\rho_{C}^{BC}$ are of the form $|b \rangle |C \rangle$. Since $P_{BC}^{a}$ is a projector onto an eigenspace of this operator, this proves that $P_{BC}^{a}$ has the form $P_{BC}^{a} = \sum_{b} |b \rangle \langle b| B \otimes T_{C}^{b}$ for some operators $\{ T_{C}^{b} \}_{b}$ on $C$. This immediately gives the claim (40).

As explained in Section 4.4, we will later apply the splitting-chain-rule recursively. In particular, we will further split up split states. Conveniently, orthogonality properties are preserved under such successive splitting operators, as we now explain.

For concreteness, suppose that we split a state $| \Psi_{A_1 B_1 C_1 D_1} \rangle$ into states $\{ |\Psi_{A_1 B_1 C_1 D_1}^{\alpha} \rangle \}_{\alpha}$ satisfying

$$
H(A_1 |B_1 C_{1} \rangle)_{\rho_{A_1}} + H(B_1 |C_{1} \rangle)_{\rho_{A_1}} \geq H(A_1 B_1 |C_{1} \rangle)_{\rho_{A_1}}.
$$

Assume further that $B_1 = A_2 B_2$ is bipartite. We can then split each $| \Psi_{A_1 B_1 C_1 D_1}^{\alpha} \rangle$ further into a family of states $\{ |\Psi_{A_1 B_1 C_1 D_1}^{\alpha_1 \alpha_2} \rangle \}_{\alpha_2}$ such that

$$
H(A_2 |B_2 C_{1} \rangle)_{\rho_{A_1}} + H(B_2 |C_{1} \rangle)_{\rho_{A_1}} \geq H(A_2 B_2 |C_{1} \rangle)_{\rho_{A_1}} = H(B_1 |C_{1} \rangle)_{\rho_{A_1}}
$$

for all $(\alpha_1, \alpha_2)$. Diagrammatically, the grouping/splitting of systems can be drawn as

$$
\begin{array}{cccc}
B_1 & A_2 & A_1 & \{D_2 \} \\
B_2 & C & D_1
\end{array}
$$

Clearly, a desirable property is that these states are orthogonal, such that

$$
| \Psi_{A_1 B_1 C_1 D_1} \rangle = \sum_{(\alpha_1, \alpha_2)} |\Psi_{A_1 B_1 C_1 D_1}^{\alpha_1 \alpha_2} \rangle
$$

is a decomposition of $| \Psi_{A_1 B_1 C_1 D_1} \rangle$ into mutually orthogonal states.

We will prove this statement by considering the corresponding projection operators $\{ Q_{A_1 D_1}^{\alpha_1} \}_{\alpha_1}$ and $\{ Q_{A_2 B_2 D_2}^{\alpha_1 \alpha_2} \}_{(\alpha_1, \alpha_2)}$ (where $D_2 = D_1 A_1$) defined by the splitting-chain-rule; i.e., these are operators satisfying

$$
| \Psi_{A_1 B_1 C_1 D_1}^{\alpha_1} \rangle = Q_{A_1 D_1}^{\alpha_1} | \Psi_{A_1 B_1 C_1 D_1} \rangle
$$

$$
| \Psi_{A_2 B_2 C_2 D_2}^{\alpha_1 \alpha_2} \rangle = Q_{A_2 D_2}^{\alpha_1 \alpha_2} | \Psi_{A_1 B_1 C_1 D_1}^{\alpha_1} \rangle = Q_{A_2 D_2}^{\alpha_1 \alpha_2} Q_{A_1 D_1}^{\alpha_1} | \Psi_{A_1 B_1 C_1 D_1} \rangle
$$

24
By definition, for every \( \alpha_1 \), the operators \( \{ Q_{A_2D_2}^{\alpha_1\alpha_2} \}_{(\alpha_1,\alpha_2)} \) are mutually orthogonal for different \( \alpha_2 \). We will now show that these operators satisfy the inequality

\[
Q_{A_2D_2}^{\alpha_1\alpha_2} \leq Q_{A_1D_1}^{\alpha_1} \, .
\]  

(41)

for all \((\alpha_1,\alpha_2)\). This expresses the fact that the operators \( Q_{A_2D_2}^{\alpha_1\alpha_2} \) are a “refinement” of \( Q_{A_1D_1}^{\alpha_1} \). In particular, their images are orthogonal for different values of \( \alpha_1 \), and we have \( Q_{A_2D_2}^{\alpha_1\alpha_2} Q_{A_1D_1}^{\alpha_1} = Q_{A_2D_2}^{\alpha_1\alpha_2} \) (cf. Lemma 15.2 (ii)). In other words, each of the states \( |\Psi_{A_1B_1C_1D_1}^{\alpha_1\alpha_2} \rangle \) can be obtained by applying a single projection to \( |\Psi_{A_1B_1C_1D_1}^{\alpha_1} \rangle \).

The proof involves the following property of the projection operators.

**Remark 5.7 (Operator inequalities).** Let \( \rho_{ABCD} = |\Psi_{ABCD}^{\alpha_1}\rangle \langle \Psi_{ABCD}^{\alpha_1}| \), \( Q_{AD}^{\alpha} \), and \( \rho_{ABCD}^{\alpha_1\alpha_2} = |\Psi_{ABCD}^{\alpha_1}\rangle \langle \Psi_{ABCD}^{\alpha_1}| \) be defined as in Lemma 5.4. Let \( \text{id}_{\text{supp}(\rho)} \) denote the projector onto the support of the operator \( \rho \). Then\(^\text{18}\)

\[
\text{id}_{\text{supp}(\rho_{\alpha_1A_1D_1F})} \leq Q_{AD}^{\alpha} \leq \text{id}_{\text{supp}(\rho_{AD})}
\]

for any subsystem \( F \subseteq BC \). (By that, we mean that \( \mathcal{H}_{BC} \) is the product \( \mathcal{H}_{BC} \cong \mathcal{H}_F \otimes \mathcal{H}_G \) of two systems \( F \) and \( G \), such that \( BC = FG \).)

Indeed, the second inequality of this remark gives

\[
Q_{A_2D_2}^{\alpha_1\alpha_2} \leq \text{id}_{\text{supp}(\rho_{A_2D_2}^{\alpha_1})} = \text{id}_{\text{supp}(\rho_{A_1A_2D_1}^{\alpha_1})}
\]

because \( D_2 = D_1A_1 \), whereas the first inequality with \( F = A_2 \) (recall that \( B_1 = A_2B_2 \)) leads to

\[
\text{id}_{\text{supp}(\rho_{A_1D_1A_2}^{\alpha_1})} \leq Q_{A_1D_1}^{\alpha_1} \, .
\]

This proves the fundamental property (41).

It remains to give a proof of the statement made in the remark.

**Proof.** According to Lemma 5.2 \((\text{ii})\), it suffices to show that

\[
\text{supp}(\rho_{\alpha_1A_1D_1F}) \subseteq \text{supp}(Q_{AD}^{\alpha} \otimes \text{id}_F)
\]

and

\[
\text{supp}(Q_{AD}^{\alpha}) \subseteq \text{supp}(\rho_{AD}) \, .
\]  

(42)

The first of these inequalities is a direct consequence of the fact that \( \rho_{\alpha_1A_1D_1F} = (\text{id}_F \otimes Q_{AD}^{\alpha}) \rho_{ADF} (\text{id}_F \otimes Q_{AD}^{\alpha}) \). To prove the second inequality, observe that \( Q_{AD}^{\alpha} \) projects onto an eigenspace of the conditional operator \( \text{tr}_{BC}(\sigma_C^{-1/2} |\Psi_{ABCD}^{\alpha} \rangle \langle \Psi_{ABCD}^{\alpha}| \sigma_C^{-1/2}) \), and thus

\[
\text{supp}(Q_{AD}^{\alpha}) \subseteq \text{supp} \left( \text{tr}_{BC}(\sigma_C^{-1/2} |\Psi_{ABCD}^{\alpha} \rangle \langle \Psi_{ABCD}^{\alpha}| \sigma_C^{-1/2}) \right) \, .
\]

The inclusion \((42)\) then follows because the latter set is contained in \( \text{supp}(\rho_{AD}) \). This can be verified for example by using a Schmidt decomposition \( |\Psi_{ABCD}^{\alpha_1}\rangle = \sum_{\mu} \sqrt{\mu} |\mu_{BC} \rangle |\mu_{AD} \rangle \) of \( |\Psi_{ABCD}^{\alpha_1}\rangle \). In terms of this decomposition, we have

\[
\text{tr}_{BC}(\sigma_C^{-1/2} |\Psi_{ABCD}^{\alpha_1}\rangle \langle \Psi_{ABCD}^{\alpha_1}| \sigma_C^{-1/2}) = \sum_{\mu,\mu'} \text{tr}(\sigma_C^{-1/2} |\mu_{BC} \rangle \langle \mu'_{BC}| \sigma_C^{-1/2}) |\mu_{AD} \rangle \langle \mu'_{AD}| \, .
\]

and the support of this operator is clearly contained in \( \text{span} \{|\mu_{AD} \rangle \} = \text{supp}(\rho_{AD}) \). \(\square\)

\(^{18}\)Recall that, according to our convention, the first inequality is an abbreviation for the operator inequality \( \text{id}_{\text{supp}(\rho_{A_1D_1F})} \leq Q_{AD}^{\alpha} \otimes \text{id}_F \) (see Section 5.1 for more details).
5.5 Recombination-rules for split states

As discussed in Section 4.3, we will need a converse to the splitting rule which shows that the entropy of the original state is large if it is large for each split state. Here we show how this works in detail in the most simple case. Again, this section may be omitted, but it is instructive for the slightly more intricate case we will need below (cf. Lemma 6.7).

Remarkably, the statement we will prove is generally true for any system \( F \) which we do not condition on.

**Lemma 5.8.** Let \( |\Psi_{ABCD}\rangle \), \( \{|\Psi_{ABCD}\rangle\}_{\alpha \in [m]} \) and \( \{Q^\alpha_{AD}\}_{\alpha \in [m]} \) be as in Corollary 5.5. Let \( F \subseteq ABD \) be an arbitrary subsystem. Then

\[
\min_{\alpha \in [m]} H(F|C)_{\rho^\alpha} - 2 \log m \leq H(F|C)_{\rho}.
\]

**Proof.** Let \( \lambda := 2^{-\min_{\alpha \in [m]} H(F|C)_{\rho^\alpha}} \). We then have \( \rho^\alpha_{FC} \leq \lambda \sigma_C \) for all \( \alpha \in [m] \), or

\[
\frac{\rho^\alpha_{FC}}{\sigma_C} \leq \lambda \text{id}_{FC}.
\]

Using the commutativity of \( Q^\alpha_{AD} \) and \( \sigma_C \), we can rewrite this as

\[
\text{tr}_{FC}(Q^\alpha_{AD} \frac{\rho_{ABCD}}{\sigma_C} Q^\alpha_{AD}) \leq \lambda \text{id}_{FC} \quad \text{for all } \alpha \in [m].
\]

At this point, we use a statement about operators which we state as Lemma B.1 in the appendix. It tells us that the previous inequalities imply that

\[
\text{tr}_{FC}(Q_{AD} \frac{\rho_{ABCD}}{\sigma_C} Q_{AD}) \leq \lambda m^2 \text{id}_{FC},
\]

where \( Q_{AD} = \sum_{\alpha \in [m]} Q^\alpha_{AD} \). Recall that the operators \( Q^\alpha_{AD} \) are defined in terms of the eigenspaces of \( \text{tr}_{BC}(\rho_{ABCD}) \). Their definition implies that \( Q_{AD} \) restricted to the support of \( \text{tr}_{BC}(\rho_{ABCD}) \) is equal to the identity. Thus the last inequality simply says

\[
\frac{\rho_{FC}}{\sigma_C} \leq \lambda m^2 \text{id}_{FC}.
\]

Multiplying from the left and the right by \( \sigma_C^{1/2} \) gives the claim. \( \square \)

6 Entropy sampling

We now return to our main problem, i.e., the analysis of a state \( \rho_{X^nE} \) with classical part \( X^n = (X_1, \ldots, X_n) \), and the relation of the entropy \( H^\varepsilon_{\min}(X_S|E)_\rho \) of a randomly chosen subset \( S \subset [n] \) to the entropy \( H_{\min}(X^n|E)_\rho \) of all classical parts. We proceed as sketched in Section 4.4. In Section 6.1, we describe the recursive splitting of the joint min-entropy \( H_{\min}(X^n|E)_\rho \) into a sum of individual contributions of each random variable. We then discuss how high-entropy components can be recombined to a state with high min-entropy (Section 6.2). In particular, we relate the smooth min-entropy \( H_{\min}^\varepsilon(X_S|E)_\rho \) to the probability weight \( \omega(\Gamma) \) of a certain set \( \Gamma \) under a given distribution \( \omega \). We then study the behavior of a sampler with respect to this quantity. For this purpose, we introduce the concept of a parallel sampler in Section 6.3. We then show that with high probability over the choice of \( S \), the probability \( \omega(\Gamma) \) of interest is large (Section 6.4).

We finally combine these components in Section 6.5, where we state our main result, i.e., the preservation of (smooth) min-entropy rates under sampling.
As in Figure 1, the immediate descendants of every node give an orthogonal decomposition of the state associated with it, and are obtained by applying corresponding projection operators. In Lemma 6.3, we will show that the states at level $j$ are orthogonal, for every level $j \in [n]$. In particular, this means that the leaves form an orthogonal decomposition of the original state. Observe that we label each vertex by the corresponding sequence of splitting operators; in particular, the leaves carry labels $\alpha^n \in [m]^n$.

6.1 Splitting

We apply the splitting-chain-rule recursively to a state $\rho_{X^n E}$, where $X_1, \ldots, X_n$ are random variables on an alphabet $X$. Let $|\Psi_{X^n E}\rangle$ be a purification of $\rho_{X^n E}$ (for simplicity, we will henceforth often omit subscripts denoting systems, where there is no potential for confusion). Furthermore, let $\sigma_E$ be a nonnegative operator on $E$. In Figure 2, we visualise the set of states introduced in the following definition by a tree.

**Definition 6.1** ("Split states"). Let $\rho_{X^n E} = |\Psi\rangle\langle\Psi|$ and let $\rho_{X^n E}^{\alpha^i} = |\Psi^{\alpha^i}\rangle\langle\Psi^{\alpha^i}|$ be pure states recursively defined as follows. Set $|\Psi^{\alpha^0}\rangle := |\Psi\rangle$. To obtain $|\Psi^{\alpha^j}\rangle$ for $j \in [n]$ and $\alpha^j = (\alpha^0, \alpha^{j-1}) \in [m]^j = [m] \times [m]^{j-1}$, apply Corollary 5.5 to the state $|\Psi^{\alpha^{j-1}}\rangle$ with $A = X_j$, $B = X_{>j}$, $C = E$, $D = X_{\leq j-1}R$. This gives projectors $P_{X_{>j} E}^{\alpha^j}$ and $Q_{X_{\leq j} R}^{\alpha^j}$; we define $|\Psi^{\alpha^j}\rangle$ as in Corollary 5.5 as $|\Psi^{\alpha^j}\rangle = Q_{X_{\leq j} R}^{\alpha^j} Q_{X_{>j} E}^{\alpha^j} |\Psi\rangle$.

Spelling out this recursive definition, we have

$$ |\Psi^{\alpha^j}\rangle = Q_{X_{\leq j} R}^{\alpha^j} \cdots Q_{X_{\leq 1} R}^{\alpha^i} |\Psi\rangle \quad (43) $$

$$ = \tilde{P}_{X_{>j} E}^{\alpha^j} \cdots \tilde{P}_{X_{>1} E}^{\alpha^i} |\Psi\rangle, \quad (44) $$

where $\tilde{P}_{X_{>j} E} = \sigma_{E}^{-1/2} P_{X_{>j} E}^{\alpha^j} \sigma_{E}^{-1/2}$. The following auxiliary result will prove useful. We will apply it to show that the states on each level of the tree in Figure 2 are mutually orthogonal (by level, we mean all vertices at a fixed depth of the tree, i.e., distance from the root). In fact, any two states in different subtrees are mutually orthogonal, but we will not need this statement here. The proof of the following lemma relies on the fact that splitting preserves orthogonality. It is analogous to the derivation of (11) in Section 5.4.

**Lemma 6.2.** For all $j \geq k$ and $\alpha^j \in [m]^j$ we have

$$ Q_{X_{\leq j} R}^{\alpha^j} Q_{X_{\leq k} R}^{\alpha^k} = Q_{X_{\leq k} R}^{\alpha^k} Q_{X_{\leq j} R}^{\alpha^j} = Q_{X_{\leq j} R}^{\alpha^j} . $$

Moreover, the operators $\{Q_{X_{\leq j} R}^{\alpha^j}\}_{\alpha^j \in [m]^j}$ are pairwise orthogonal for a fixed $j \in [n]$.
Proof. Note that the first claim trivially holds for \( j = k \) since the operators are projectors. Observe that for any \( j > 1 \), we have

\[
Q_{X_{\leq j} R}^{\alpha_j} \leq \text{id}_{\text{supp}(\rho_{X_{\leq j} R}^{\alpha_j-1})} = \text{id}_{\text{supp}(\rho_{X_{\leq j-1} R}^{\alpha_j-1})} \leq Q_{X_{\leq j-1} R}^{\alpha_j-1},
\]

where we used Remark \(5.7\) twice (with \( F = X_j \)). Inductively, we obtain

\[
Q_{X_{\leq j} R}^{\alpha_j} \leq Q_{X_{\leq k} R}^{\alpha_k}
\]

for any \( k \leq j \). The first claim therefore follows from Lemma \(5.2\) \(1\).

The orthogonality of the operators \( \{ Q_{X_{\leq j} R}^{\alpha_j} \}_{\alpha_j \in [m]} \) immediately follows from the first claim:

For \( \alpha^j \neq \beta^j \in [m]^j \), let \( k \leq j \) be the minimal index in which they differ, i.e., \( \alpha_k \neq \beta_k \) and \( \alpha^{k-1} = \beta^{k-1} \). We then have by the first claim

\[
Q_{X_{\leq j} R}^{\alpha_j} Q_{X_{\leq j} R}^{\beta_j} = Q_{X_{\leq j} R}^{\alpha_j} Q_{X_{\leq k} R}^{\alpha_k} Q_{X_{\leq k} R}^{\beta_k} Q_{X_{\leq j} R}^{\beta_j} = 0,
\]

since the operators \( Q_{X_{\leq k} R}^{\alpha_k} = Q_{X_{\leq k} R}^{(\alpha_k, \alpha^{k-1})} \) and \( Q_{X_{\leq k} R}^{\beta_k} = Q_{X_{\leq k} R}^{(\beta_k, \alpha^{k-1})} \) are orthogonal for \( \alpha_k \neq \beta_k \). \( \square \)

As promised, we now establish a few properties of the split states such as their orthogonality and the fact that they are partly classical as the original state.

**Lemma 6.3** (Properties of the split states). The states introduced in Definition 6.1 have the following properties.

(i) The states \( \{ | \Psi^{\alpha^j} \rangle \}_{\alpha^j \in [m]^j} \) are pairwise orthogonal for a fixed \( j \in [n] \).

(ii) The states \( \{ | \Psi^{\alpha^n} \rangle \}_{\alpha^n \in [m]^n} \) form an orthogonal resolution of \( | \Psi \rangle \), i.e., \( \sum_{\alpha^n \in [m]^n} | \Psi^{\alpha^n} \rangle = | \Psi \rangle \). In particular, \( \omega(\alpha^n) := \text{tr} | \Psi^{\alpha^n} \rangle \langle \Psi^{\alpha^n} | \) defines a probability distribution on \([m]^n\).

(iii) The state \( | \Psi^{\alpha^j} \rangle \) can be obtained by a single projection on \( X_{\leq j} R \), i.e., \( | \Psi^{\alpha^j} \rangle = Q_{X_{\leq j} R}^{\alpha^j} | \Psi \rangle \).

(iv) For every \( j \in [n] \) and \( \alpha^j \in [m]^j \), the state \( \rho_{X_{\leq j} E}^{\alpha^j} \) is classical on \( X^n \).

(v) For all \( \sigma = \sigma_E \), we have \( H(\emptyset | E)^{\rho_{\alpha^j}} \geq H(\emptyset | E)^{\sigma} \).

The probability distribution \( \omega \) (introduced in \(1\)) on the leaves \( [m]^n \) of the tree in Figure 2 will play an important role in our recombination step. Inequality \(\omega\) can be seen as an expression of the fact that splitting does not affect the part we condition on.

Proof. First observe that \(\omega\) follows inductively from Lemma \(6.2\) and expression \(\omega\). Similarly, the orthogonality \(\omega\) follows from this lemma and \(\omega\). Statement \(\omega\) follows by induction over \( j \) from \(\omega\). Statement \(\omega\) follows inductively from Remark \(6.9\) applied with \(D_1 = X_{\leq j-1}\) and \(D_2 = R\). Finally, the claim \(\omega\) directly follows from \(\omega\) and Lemma \(5.3\) (with \( C = X_{\leq j-1}\) and \( B = E\)). \( \square \)

The main reason for introducing the split states \( \{ | \Psi^{\alpha^j} \rangle \} \) is the fact that they allow us to split the joint entropy \( H(X^n | E)^{\rho} \) into individual contributions according to the splitting-chain-rule (Corollary \(5.5\)). We express this central result as follows.

**Theorem 6.4** (“Splitting”). The split states satisfy

\[
H(\emptyset | E)^{\rho_{\alpha^n}} + \sum_{j=1}^{n} H(X_j | X_{>j} E)^{\rho_{\alpha^j}} \geq H(X_{>0} | E)^{\emptyset} - \frac{n \log |X|}{m}.
\] (45)

for any \( \alpha^n \in [m]^n \).
classical systems to get dimension bound (ii) of Lemma 5.2 and the positivity of the min-entropy (Lemma 5.2 (i)) for this gives the claim follows from this and (47).

Figure 2: For convenience, we use a slightly modified tree $T$ entropic quantities appearing on the lhs of the inequality (45) as attached to the tree given in Figure 3.

Proof. In the following, we sometimes refer to the empty set as $X_{>n}$. By construction and Corollary 5.5, the split states satisfy the inequalities

$$H(X_j|X_{>j}E)_{\rho^j} + H(X_{>j}|E)_{\rho^j} \geq H(X_{>j-1}|E)_{\rho^{j-1}} - \frac{\Delta_j}{m} \quad \text{for all } j \in [n],$$

where $\Delta_j = H(X_{>j-1}|E)_{\rho^{j-1}} - H(X_j|X_{>j}E)_{\rho^j} - H(X_{>j}|E)_{\rho^j}$. Summing these inequalities over all $j \in [n]$, we get

$$\sum_{j \in [n]} H(X_j|X_{>j}E)_{\rho^j} \geq \sum_{j \in [n]} \left( H(X_{>j-1}|E)_{\rho^{j-1}} - H(X_{>j}|E)_{\rho^j} \right) - \frac{1}{m} \sum_{j \in [n]} \Delta_j.$$

Because the rhs is a telescoping sum, i.e.,

$$\sum_{j \in [n]} \left( H(X_{>j-1}|E)_{\rho^{j-1}} - H(X_{>j}|E)_{\rho^j} \right) = H(X_{>0}|E)_{\rho^0} - H(X_{>n}|E)_{\rho^n},$$

this gives

$$H(X_{>n}|E)_{\rho^n} + \sum_{j \in [n]} H(X_j|X_{>j}E)_{\rho^j} \geq H(X_{>0}|E)_{\rho^0} - \frac{1}{m} \sum_{j \in [n]} \Delta_j.$$

Note that $\rho^n_{X^n E}$ is classical on $X^n$, according to Lemma 5.3 [10]. We can therefore use the dimension bound [10] of Lemma 5.2 and the positivity of the min-entropy (Lemma 5.2 [10]) for classical systems to get

$$\log |\mathcal{X}| \geq H(X_j|X_{>j}E)_{\rho^j} - H(X_{>j}|E)_{\rho^j} \geq H(X_{>j-1}|E)_{\rho^{j-1}} - H(X_{>j}|E)_{\rho^{j-1}} - H(X_j|X_{>j}E)_{\rho^j} = \Delta_j \quad \text{for all } j \in [n].$$

The claim follows from this and (47).

To put the statement of Theorem 6.4 into a more concise form, it is useful to think of the entropic quantities appearing on the lhs of the inequality (45) as attached to the tree given in Figure 2. For convenience, we use a slightly modified tree $T_n$ which has spades attached to the leaves of the original tree (see Figure 3).

We can then attach weights to the edges of $T_n$ according the rule $v_\rho$ given in Figure 4. For a path $\alpha^n \in [m]^n$ from the root to a spade (i.e., leaf), we define the weight $v_\rho(\alpha^n)$ of the path $\alpha^n$ as the sum of the values on the edges along this path. In particular, for the weighting $v_\rho$ specified by Figure 3, the weight $v_\rho(\alpha^n)$ coincides with the lhs of (45) in Theorem 6.4.

29
More generally, we slightly abuse notation and define the value $w(T)$ of a tree $T$ with weighting $w$ as the minimal value of a path from the root to a leaf. Theorem 6.4 can then be reformulated as follows.

**Theorem 6.4'.** Let $T_n$ be the tree introduced in Figure 3, and let $v_\rho$ be the weighting specified by Figure 3. Then $v_\rho(T_n) \geq H(X_{>0}|E) \frac{\rho^i}{\sigma - m} \log |X|$.

We will later be interested in different weightings. We will also show a converse to this statement: If the value of a tree is large, then so is the corresponding entropy.

### 6.2 Recombining

To show that the original state $\rho_{X^nE}$ has a large smooth min-entropy $H_{\min}(X_S|E)$ for a randomly selected subset $S \subseteq [n]$, we will now study how the split states can be recombined. More precisely, we are interested in properties of states $|\tilde{\Psi}\rangle$ that are obtained by summing up states $|\Psi^{\alpha^n}\rangle$ corresponding to a subset $\Gamma \subseteq [m]^n$ of leaves of the tree in Figure 2.

In Section 6.2.1, we discuss how such a recombined state can be defined recursively, starting from the bottom of the tree. We then use the corresponding intermediate states in Section 6.2.2 to analyse how a judicious choice of $\Gamma$ yields a recombined state $|\tilde{\Psi}\rangle$ with a large min-entropy $H_{\min}(X_S|E)\tilde{\rho}$.

#### 6.2.1 Partially recombined states and properties

We are interested in properties of the state

$$|\tilde{\Psi}\rangle = \sum_{\alpha^n \in \Gamma} |\Psi^{\alpha^n}\rangle$$

(48)

obtained by summing over a certain subset $\Gamma \subseteq [m]^n$ of paths. To analyse such a “partially recombined” state, we will consider intermediate states attached to a tree. The state $|\tilde{\Psi}\rangle$ will sit at the root of the tree. We will refer to it as $|\tilde{\Psi}\rangle = |\Psi^{\tilde{\alpha}^b}\rangle$ in the following definition, which we illustrate in Figure 5.

**Definition 6.5 (“Recombined states”).** Let $\Gamma \subseteq [m]^n$ be arbitrary, and let $|\Psi^{\alpha^n}\rangle$ for $\alpha^n \in [m]^n$ be the split states introduced in Definition 6.4. We define the recombined states

$$|\tilde{\Psi}^{\tilde{\alpha}^i}\rangle = \sum_{\gamma^n \in \Gamma} |\Psi^{\gamma^n}\rangle$$

and let $\tilde{\rho}_{X^nE}^{\tilde{\alpha}^i} = \langle\tilde{\Psi}^{\tilde{\alpha}^i}|\tilde{\Psi}^{\tilde{\alpha}^i}\rangle$ for all $\tilde{\alpha}^i \in [m]^i$. For simplicity, we omit $\Gamma$ in the notation.

Not surprisingly, the recombined states inherit many properties of the split states. The following lemma summarises these, and is the analog of Lemma 6.3.
Figure 5: Here we illustrate the partially recombined state of Definition 6.5, for $n = 3, m = 2$ and $\Gamma = \{112, 211, 212, 222\}$. We again associate every state $|\tilde{\Psi}^{\alpha_1}\rangle$ with the node carrying the label $\alpha^j \in \{m\}^j$. We start by defining the leaves, i.e., the states $|\tilde{\Psi}^{\alpha^n}\rangle$ for $\alpha^n \in \{m\}^n$: For $\alpha^n \in \Gamma$ (illustrated by triangles), we use the same leaves as in Figure 2, i.e., we set $|\tilde{\Psi}^{\alpha^n}\rangle = |\Psi^{\alpha^n}\rangle$. On the other hand, we set $|\tilde{\Psi}^{\alpha^n}\rangle = 0$ for $\alpha^n \notin \Gamma$. We then work our way up the tree, defining the state in each node as the sum of its immediate descendants. The elements at the dotted nodes are equal to zero, whereas for example $|\tilde{\Psi}^{2}\rangle = |\tilde{\Psi}^{211}\rangle + |\tilde{\Psi}^{222}\rangle = |\tilde{\Psi}^{211}\rangle + |\tilde{\Psi}^{212}\rangle + |\tilde{\Psi}^{222}\rangle$. Clearly, the state at the root is equal to the sum of the leaves in $\Gamma$, i.e., $|\tilde{\Psi}\rangle = \sum_{\alpha^n \in \Gamma} |\tilde{\Psi}^{\alpha^n}\rangle$. We will show in Lemma 6.6 that the states at any given level are orthogonal, and that movement in this diagram is achieved by the same projection operators as in the tree of Figure 2. Moreover, the entropies of interest corresponding to this modified tree are at least as large as those corresponding to Figure 2.

Lemma 6.6. The recombined states have the following properties.

(i) The states $\{|\tilde{\Psi}^{\alpha^j}\rangle\}_{\alpha^j \in \{m\}^j}$ are orthogonal for a fixed $j \in [n]$.

(ii) The states $\{|\tilde{\Psi}^{\alpha^j}\rangle\}_{\alpha^j \in \{m\}^j}$ form a resolution of $|\tilde{\Psi}^{\alpha^{j-1}}\rangle$, i.e., $\sum_{\alpha^j \in \{m\}^j} |\tilde{\Psi}^{\alpha^j}\rangle = |\tilde{\Psi}^{\alpha^{j-1}}\rangle$.

(iii) The states satisfy the recursion relation $|\tilde{\Psi}^{\alpha^j}\rangle = Q^{\alpha^j}_{X_{\leq j} R} |\tilde{\Psi}^{\alpha^{j-1}}\rangle$ for all $j \in [n]$ and $\alpha^j \in \{m\}^j$.

(iv) For every $j = 0, \ldots, n$, there is a projector $T^{\sigma}_{X_{\leq n} R}$ such that $|\tilde{\Psi}^{\alpha^j}\rangle = T^{\sigma}_{X_{\leq n} R} |\tilde{\Psi}\rangle$. In particular, for $\sigma = \sigma_E$ arbitrary, we have $H(\langle 0 | E \rangle)_{\tilde{P}_{\alpha^j}} \geq H(\langle 0 | E \rangle)_{\tilde{P}}$.

(v) We have $H(X_j | X_{> j} E)_{\tilde{P}_{\alpha^j}} \geq H(X_j | X_{> j} E)_{\tilde{P}}$ for all $j \in [n]$ and $\alpha^j \in \{m\}^j$.

(vi) For all $\alpha^n \in \{m\}^n$, we have $H(\langle 0 | E \rangle)_{\tilde{P}_{\alpha^n}} \geq H(\langle 0 | E \rangle)_{\tilde{P}}$.

The recursion relation (iii) will be most important in our analysis. It provides a means of studying properties of the corresponding states in a recursive manner, moving up the tree in Figure 5 to the root.

Proof. The orthogonality (i) of the states $\{|\tilde{\Psi}^{\alpha^j}\rangle\}_{\alpha^j}$ is a direct consequence of the orthogonality of the states $\{|\Psi^{\alpha^n}\rangle\}_{\alpha^n}$ (cf. Lemma 6.3 (i)). Identity (ii) also follows from the definition of $|\tilde{\Psi}^{\alpha^j}\rangle$.

For the proof of (iii), observe that

$$Q^{\alpha^j}_{X_{\leq j} R} |\Psi^{\gamma^j}\rangle = Q^{\alpha^j}_{X_{\leq j} R} Q^{\alpha^n}_{X_{\leq n} R} |\Psi\rangle = Q^{\alpha^j}_{X_{\leq j} R} Q^{\alpha^n}_{X_{\leq n} R} Q^{\alpha^n}_{X_{\leq n} R} |\Psi\rangle = \begin{cases} |\Psi^{\gamma^n}\rangle & \text{if } \gamma^j = \alpha^j \\ 0 & \text{otherwise} \end{cases},$$

by Lemma 6.2. Applying this to compute $Q^{\alpha^j}_{X_{\leq j} R} |\tilde{\Psi}^{\alpha^{j-1}}\rangle$ immediately gives the claim (iii).

31
Lemma 6.7. from one level up to another level closer to the root. tree, it expresses the fact that the entropies of interest do not decrease significantly when we move

where

Definition 6.5 and (44). The claim (v) therefore follows from Lemma 5.3. We next prove an analog of the basic recombination lemma (Lemma 5.8) for the partially recombined states \( \hat{\Psi}^{\alpha_i} \). In terms of the position of the corresponding states in the described tree, it expresses the fact that the entropies of interest do not decrease significantly when we move from one level up to another level closer to the root.

Lemma 6.7. For all \( \mathcal{A} \subseteq [n] \) (possibly empty), \( i \in [n] \) and \( \alpha^{i-1} \in [m]^{i-1} \)

\[
\min_{\alpha_i \in [m]} H(X_{\mathcal{A}}|E)_{\hat{\rho}^{\alpha_i}} - 2 \log m \leq H(X_{\mathcal{A}}|E)_{\hat{\rho}^{\alpha^{i-1}}}.
\]

Proof. Let \( \alpha^{i-1} \in [m]^{i-1} \) be fixed and let \( \lambda := 2^{-\min_{\alpha_i \in [m]} H(X_{\mathcal{A}}|E)_{\hat{\rho}^{\alpha_i}}} \), where \( \alpha_i = (\alpha_i, \alpha^{i-1}) \). By definition

\[
\frac{\hat{\rho}^{X_{\mathcal{A}}|E}}{\sigma_E} \leq \lambda \text{id}_{X_{\mathcal{A}}E} \quad \text{for all } \alpha_i \in [m]. \tag{49}
\]

To relate this to \( \hat{\rho}^{X_{\mathcal{A}}|E} \), we use the recursion relation (iii) of Lemma 6.6 to rewrite (49) as

\[
\text{tr}_{X_{\mathcal{A}}E}(Q_{X_{\mathcal{A}}|E}^{X_{\mathcal{A}}|E} \frac{\hat{\rho}^{X_{\mathcal{A}}|E}}{\sigma_E}) \leq \lambda \text{id}_{X_{\mathcal{A}}E} \quad \text{for all } \alpha_i \in [m].
\]

Lemma 6.1 thus implies

\[
\text{tr}_{X_{\mathcal{A}}E}(Q \frac{\hat{\rho}^{X_{\mathcal{A}}|E} Q}{\sigma_E}) \leq \lambda m^2 \text{id}_{X_{\mathcal{A}}E}, \tag{50}
\]

where \( Q = \sum_{\alpha_i \in [m]} Q_{X_{\mathcal{A}}|E}^{X_{\mathcal{A}}|E} \). But

\[
Q \frac{\hat{\rho}^{X_{\mathcal{A}}|E} Q}{\sigma_E} = \sigma_E^{-1/2} \left( \sum_{\alpha_i} Q_{X_{\mathcal{A}}|E}^{X_{\mathcal{A}}|E} |\hat{\Psi}^{\alpha_i}| \right) \left( \sum_{\alpha_i} Q_{X_{\mathcal{A}}|E}^{X_{\mathcal{A}}|E} |\hat{\Psi}^{\alpha_i}| \right) \sigma_E^{-1/2}
\]

\[
= \sigma_E^{-1/2} \left( \sum_{\alpha_i} |\hat{\Psi}^{\alpha_i}| \right) \left( \sum_{\alpha_i} |\hat{\Psi}^{\alpha_i}| \right) \sigma_E^{-1/2}
\]

\[
= \sigma_E^{-1/2} |\hat{\Psi}^{\alpha^{i-1}}| \langle \hat{\Psi}^{\alpha^{i-1}} | \sigma_E^{-1/2},
\]

where we used (iii) and (ii) of Lemma 6.6. Inserting this into (50) gives

\[
\frac{\hat{\rho}^{X_{\mathcal{A}}|E}}{\sigma_E} \leq \lambda m^2 \text{id}_{X_{\mathcal{A}}E},
\]

which concludes the proof. \( \square \)
6.2.2 Recombining high-entropy components

We now study the entropies associated with recombined states, in the special case where \( \Gamma \subset [m]^n \) is chosen as the set of “high-entropy paths” for a subset \( S \). Our main result of this section is Theorem 6.13, which expresses the fact that the corresponding entropy \( H_{\min}(X_S|E)_b \) is large.

Let us fix a subset \( S \subset [n] \). We will be interested in the entropies of variables \( X_i \) with \( i \in S \). That is, we consider the weighting \( v^S_{\rho} \) defined by Figure 6 of the tree \( T_n \) introduced after Theorem 6.4. A given path \( \alpha \in [m]^n \) then has weight

\[
v^S_{\rho}(T_n, \alpha) = H(\emptyset|E)_{\rho^\alpha} + \sum_{j \in S} H(X_j|X_{>j}E)_{\rho^\alpha}
\]

by definition. We cannot expect this to be large for all \( \alpha \in [m]^n \); in particular, the value \( v^S_{\rho}(T_n) \) will in general be small. We therefore introduce the following sets.

**Definition 6.8 (“\( \lambda \)-good paths”).** For \( \lambda > 0 \) and \( S \subset [n] \), let \( \Gamma(\lambda, S) \subset [m]^n \) be the set of \( n \)-tuples specified by Definition 6.8.

We call \( \Gamma(\lambda, S) \subset [m]^n \) the set of \( \lambda \)-good paths for \( S \).

The choice of the normalisation factor \( |S| \log |\mathcal{X}| \) will become clearer in the sequel when we relate the quantity \( v^S_{\rho}(T_n, \alpha) \) to the entropy-rate \( H(X_S|E)/|S| \log |\mathcal{X}| \).

Let us consider states that arise when recombining only \( \lambda \)-good paths. That is, we fix \( \lambda > 0 \), a subset \( S \subset [n] \) of size \( |S| = r \), and let \( \Gamma = \Gamma(\lambda, S) \) be the set of \( n \)-tuples specified by Definition 6.8. We then define the partially recombined states \( \{ \hat{\rho}^\alpha \} \) as in Definition 6.5.

Contrary to the original weighting \( v^S_{\rho} \), this weighting assigns a large weight to every path. That is, we have the statement

\[
\frac{v^S_{\hat{\rho}}(T_n, \alpha)}{|S| \log |\mathcal{X}|} \geq \lambda.
\]

**Lemma 6.9.** \( v^S_{\hat{\rho}}(T_n) \geq \lambda |S| \log |\mathcal{X}| \).

In other words, when considering the recombined states, all paths are \( \lambda \)-good. This is not the case for the original split states.

**Proof.** Suppose first that \( \alpha \in \Gamma(\lambda, S) \subset [m]^n \). Then

\[
H(X_j|X_{>j}E)_{\hat{\rho}^\alpha} \geq H(X_j|X_{>j}E)_{\rho^\alpha}
\]

for all \( j \in [n] \) by Lemma 6.6 \( \dagger \) and

\[
H(\emptyset|E)_{\hat{\rho}^\alpha} \geq H(\emptyset|E)_{\rho^\alpha}
\]

by Lemma 6.6 \( \dagger \).

---

\( \dagger \): Observe that we now explicitly mention the dependence on the tree \( T_n \) in \( v^S_{\hat{\rho}}(T_n, \alpha) \), as we will be dealing with several different (sub)trees.

33
This directly gives $v_{S}^{S}(T_{n}, \alpha^{n}) \geq \lambda|S| \log |X|$ for $\alpha^{n} \in \Gamma(\lambda, S)$. On the other hand, if $\alpha^{n} \notin \Gamma(\lambda, S)$, then we have $\hat{\rho}^{\alpha^{n}} = 0$ which implies that $H(\emptyset|E)_{\hat{\rho}^{\alpha^{n}}} = \infty$ and thus $v_{S}^{S}(T_{n}, \alpha^{n}) = \infty$.

The claim follows by taking the minimum over $\alpha^{n} \in [m]^{n}$.

Next we apply subadditivity, to go from the weighting $v_{S}^{S}$ defined by Figure 6 to the weighting $w_{S}^{S}$ introduced in Figure 7. This weighting assigns the weight

$$w_{S}^{S}(T_{n}, \alpha^{n}) = H(\emptyset|E)_{\hat{\rho}^{\alpha^{n}}} + \sum_{j \in S} H(X_{j}|X_{j\not\in S}E)_{\hat{\rho}^{\alpha^{j}}}$$

for all $j \in S$ and $\alpha^{j} \in [m]^{j}$. The statement follows immediately.

**Lemma 6.10.** $w_{S}^{S}(T_{n}) \geq v_{S}^{S}(T_{n})$.

**Proof.** With subadditivity (Lemma 5.2 iii), it is straightforward to show that

$$H(X_{j}|X_{j\not\in S}E)_{\hat{\rho}^{\alpha^{j}}} \geq H(X_{j}|X_{j}E)_{\hat{\rho}^{\alpha^{j}}}$$

for all $j \in S$ and $\alpha^{j} \in [m]^{j}$. The statement follows immediately.

Our aim is to show that if every path is $\lambda$-good for some $\lambda$, then the entropy $H(X_{S}|E)_{\hat{\rho}^{\alpha^{0}}}$ is large for the recombined state $\hat{\rho}^{\alpha^{0}}$. This expression can be seen as the value of the tree $T_{0}$ which is defined in Figure 8 i.e., we have

$$v_{S}^{S}(T_{0}) = H(X_{S}|E)_{\hat{\rho}^{\alpha^{0}}}.$$

To obtain an estimate on this quantity, we use a sequence of intermediate trees and show the following:

**Lemma 6.11.** There is a sequence $T_{n-1}, \ldots, T_{1}$ of intermediate trees such that

$$w_{S}^{S}(T_{j-1}) \geq w_{S}^{S}(T_{j}) - 2 \log m \quad \text{for all } j \in [n],$$

(53)
To obtain $T_{j-1}$ from $T_j$, the subtree defined by a vertex $\alpha^{j-1}$ at level $j$ is substituted as shown, for all $\alpha^{j-1} \in [m]^{j-1}$. Note that the vertex $\alpha^{j-1}$ has (in general) $m$ direct descendants; the figure corresponds to $m = 2$.

The tree $T_2$ obtained by applying the substitution rule to the tree $T_3$ of Figure 3.

Figure 9: The substitution rule for obtaining $T_{j-1}$ from $T_j$, $j \in [n]$. The tree $T_j$ has depth $j+1$, with spades sitting on the $j+1$-st level.

where $T_0$ is the tree in Figure 8 and $T_n$ is the original tree (see Figure 3). In particular,

$\mathbf{w}_S^\rho(T_0) \geq \mathbf{w}_S^\rho(T_n) - 2n \log m$.

(54)

Here $X_{>j} \cap S$ denotes $(X_i)_{i \in S, i > j}$ (this is equal to $\emptyset$ if $j > n$). In these expressions, the value of the tree $T_j$ is equal to

$\mathbf{w}_S^\rho(T_j) = \min_{\alpha^j \in [m]^j} \mathbf{w}_S^\rho(T_j, \alpha^j)$ where

$\mathbf{w}_S^\rho(T_n, \alpha^j) = H(X_{>j} \cap S|E)_{\rho^j} + \sum_{i \leq j} H(X_i|X_{>i} \cap S|E)_{\rho^i}$.

(55)

Proof. Note that (54) follows immediately from (53).

We first define the sequence of trees $T_{n-1}, T_{n-2}, \ldots, T_0$. We do this inductively as shown in Figure 8 that is, we obtain $T_{j-1}$ from $T_j$ by substituting subtrees corresponding to vertices $\alpha^{j-1} \in [m]^{j-1}$. Clearly, $T_j$ is a tree characterised as follows: For every $0 \leq k \leq j$, every vertex at level $k$ has $m$ immediate descendants, whereas each vertex at level $j$ has one descendant which is a spade.

The tree $T_0$ defined recursively in this way coincides with the definition given above (Figure 3). Also, it is easy to see that the value of the tree $T_j$ is given by (55). We prove the central inequality (53).

By definition, it suffices to prove that for all $\alpha^{j-1} \in [m]^{j-1}$, there is an $\alpha_j \in [m]$ such that

$\mathbf{w}_S^\rho(T_{j-1}, \alpha^{j-1}) \geq \mathbf{w}_S^\rho(T_j, \alpha^j) - 2 \log m$,

or equivalently

$\delta = \min_{\alpha_j \in [m]} \mathbf{w}_S^\rho(T_j, \alpha^j) - \mathbf{w}_S^\rho(T_{j-1}, \alpha^{j-1}) \leq 2 \log m$.

(56)

Since the two paths to the vertex $\alpha^{j-1}$ are identical in $T_j$ and $T_{j-1}$, the expression on the lhs is equal to

$\delta = \mathbf{w}_S^\rho(A) - \mathbf{w}_S^\rho(B)$.

(57)
where $A$ and $B$ are the subtrees defined by $\alpha^{j-1}$ on the left in Figure 9.

By definition, we have

$$w^S_{\hat{\rho}}(A) = \begin{cases} 
\min_{\alpha_j \in [m]} H(X_{>j} \cap S | E)_{\hat{\rho}^j} & \text{if } j \notin S \\
\min_{\alpha_j \in [m]} (H(X_{>j} \cap S | E)_{\hat{\rho}^j} + H(X_j | X_{>j} \cap S | E)_{\hat{\rho}^j}) & \text{if } j \in S
\end{cases}$$

$$w^S_{\hat{\rho}}(B) = H(X_{>j-1} \cap S | E)_{\hat{\rho}^{j-1}}.$$

We thus have to consider two cases.

(i) If $j \notin S$, then $X_{>j} \cap S = X_{>j-1} \cap S$ and (56) follows with (57) once we show that

$$\min_{\alpha_j \in [m]} H(X_{>j} \cap S | E)_{\hat{\rho}^j} - H(X_{>j-1} \cap S | E)_{\hat{\rho}^{j-1}} \leq 2 \log m.$$

This was shown in Lemma 6.7.

(ii) If $j \in S$, we have

$$\delta = \min_{\alpha_j \in [m]} \left( H(X_{>j} \cap S | E)_{\hat{\rho}^j} + H(X_j | X_{>j} \cap S | E)_{\hat{\rho}^j} - H(X_{>j-1} \cap S | E)_{\hat{\rho}^{j-1}} \right).$$

By the chain-rule (Lemma 5.2 (iv)), we have (observe that $X_{>j-1} \cap S = X_j X_{>j} \cap S$)

$$H(X_{>j} \cap S | E)_{\hat{\rho}^j} + H(X_j | X_{>j} \cap S | E)_{\hat{\rho}^j} \leq H(X_{>j-1} \cap S | E)_{\hat{\rho}^{j-1}}$$

and thus

$$\delta \leq \min_{\alpha_j \in [m]} \left( H(X_{>j-1} \cap S | E)_{\hat{\rho}^{j-1}} - H(X_{>j-1} \cap S | E)_{\hat{\rho}^{j-1}} \right).$$

The claim (56) again follows from Lemma 6.7.

In summary, we have shown the following:

**Lemma 6.12.** Let $S \subset [n]$ be arbitrary and let $\Gamma(\lambda, S) \subset [m]^n$ be the set of $\lambda$-good paths for $S$ as in Definition 6.8. Let $\{\hat{\rho}^{\alpha}\}$ be the corresponding partially recombined states as in Definition 6.5.

Then

$$\frac{H(X_S | E)_{\hat{\rho}^0}}{|S| \log |X|} \geq \frac{\lambda - 2n \log m}{|S| \log |X|}.$$

**Proof.** We have

$$H(X_S | E)_{\hat{\rho}^0} = w^S_{\hat{\rho}}(T_0)$$

(by 52),

$$\geq w^S_{\hat{\rho}}(T_n) - 2n \log m$$

(by 53),

$$\geq v^S_{\hat{\rho}}(T_n) - 2n \log m$$

Lemma 6.10 and

$$\geq \lambda |S| \log |X| - 2n \log m$$

Lemma 6.9.

We have shown that when recombining only $\lambda$-good paths, one ends up with a state with high entropy on the subset $S$ of systems of interest. The recombined state can, however, be far from the original state, if only a few paths are $\lambda$-good (or more precisely, if the share of the $\lambda$-good paths is small). We express this as follows.
Theorem 6.13 (“Recombining”). There is a probability distribution \( \omega \) on \([m]^n\) such that for any subset \( S \subset [n] \), there is a subnormalised state \( \rho_{X^n|ER} \) with

\[
\frac{H(X_S|E)}{|S| \log |X|} \geq \lambda - \frac{2n \log m}{|S| \log |X|}, \\
H(\emptyset|E) \geq H(\emptyset|E) \frac{\lambda}{2}
\]

at distance

\[
\frac{1}{2} \left\| \rho_{X^n|ER} - \rho_{X^n|E} \right\| \leq \sqrt{1 - \omega(\Gamma(\lambda, S))},
\]

from the original state \( \rho_{X^n|E} \), where \( \Gamma(\lambda, S) \subset [m]^n \) is the set of paths \( \alpha^n \in [m]^n \) such that

\[
H(\emptyset|E) \omega^n + \sum_{j \in S} H(X_j|X_{>j}E) \rho_{\alpha^n} \geq \lambda |S| \log |X| \quad \text{for all } \alpha^n = \alpha^n \in \Gamma(\lambda, S).
\]

Proof. Let \( \omega \) be the probability distribution introduced in Lemma 6.3. We set \( \tilde{\rho} = \tilde{\rho}^0 \) equal to the partially recombined state \( \hat{\rho} \).

The first bound was derived in Lemma 6.12. The second bound is identical to the claim (iv) of Lemma 6.13 for \( j = 0 \). For the bound on the distance between \( \tilde{\rho} \) and \( \rho \), we use the fact that \( \tilde{\rho}^0 = Q\rho Q \), where \( Q = \sum_{\alpha^n \in \Gamma} Q_{\alpha^n} \) is a projector (cf. Lemma 6.2). Applying the gentle measurement lemma [Win99, ON02]

\[
\frac{1}{2} \left\| \rho - Q\rho Q \right\| \leq \sqrt{\text{tr}(\rho) - \text{tr}(Q^2\rho)} \quad \text{for all subnormalised } \rho \text{ and } 0 \leq Q \leq \text{id}
\]
gives the claim.

\[\square\]

6.3 Averaging samplers and parallel samplers

To argue that an averaging sampler picks \( \lambda \)-good paths with high probability, it will be necessary to analyse the behavior of a sampler with respect to values attached to a tree. For simplicity, we consider an even simpler situation (which is more general and sufficient for our purposes): We think of values arranged in a matrix, and introduce the concept of a parallel sampler.

Consider a modified sampler situation, where instead of a single vector \( \mathbf{\beta} = (\beta_1, \ldots, \beta_n) \in [0,1]^n \), a family \( \{\beta^n = (\beta^n_1, \ldots, \beta^n_n) \in [0,1]^n\} \alpha \in [M] \) of \( M \) vectors is given. We would like to approximate the values \( \beta^n_\alpha = \frac{1}{n} \sum_{i=1}^n \beta^n_i \) simultaneously by expressions of the form \( \frac{1}{|S|} \sum_{i \in S} \beta^n_i \).

Clearly, a single (small) subset \( S \subset [n] \) will generally not give a good approximation for each one of the \( M \) vectors. However, it is possible to guarantee that it does so for most vectors, in the following sense.

Definition 6.14. Let \( M, n \in \mathbb{N} \). For any subset \( S \subset [n] \), matrix \( \mathbf{\beta} = (\beta^n_\alpha)_{\alpha \in [M], i \in [n]} \in [0,1]^{M \times n} \) and \( \xi \in [0,1] \), let \( B(\mathbf{\beta}, S, \xi) \subset [M] \) be the set of \( \alpha \in [M] \) such that

\[
\frac{1}{|S|} \sum_{i \in S} \beta^n_i \leq \frac{1}{n} \sum_{i=1}^n \beta^n_i - \xi.
\]

A \((M, n, \xi, \delta, \epsilon)\)-parallel sampler is a distribution \( P_S \) over subsets \( S \) of \([n]\) with the property that for every fixed probability distribution \( \omega \) on \([M]\),

\[
\Pr_S [\omega(B(\mathbf{\beta}, S, \xi)) \geq \delta] \leq \epsilon \quad \text{for all } \mathbf{\beta} = (\beta^n_\alpha)_{\alpha \in [M], i \in [n]} \in [0,1]^{M \times n}.
\]

A \((n, \xi, \delta, \epsilon)\)-parallel sampler is a \((M, n, \xi, \delta, \epsilon)\)-parallel sampler for any \( M \in \mathbb{N} \).

37
Clearly, a “standard” sampler corresponds to $M = 1$. In our application, the matrices $\beta \in [0, 1]^{M \times n}$ will not be arbitrary, but have a lot of redundancy. This could perhaps be exploited to find better constructions; however, for our purposes, a parallel sampler is sufficient.

We now use Markov’s inequality to obtain the following generic construction of a parallel sampler; again, more optimal constructions may be possible, but the following one is sufficient for our considerations.

**Lemma 6.15.** A $(n, \xi, \varepsilon)$-sampler is a $(n, \xi, \sqrt{\varepsilon}, \sqrt{\varepsilon})$-parallel sampler.

**Proof.** Let $M \in \mathbb{N}$ be arbitrary. Fix a probability distribution $\omega$ on $[M]$ and let $\beta = (\beta^n_\alpha)_{\alpha \in [M], i \in [n]} \in [0, 1]^{M \times n}$ be arbitrary. Since the probability on the l.h.s of (1) is bounded by $\varepsilon$ for each vector $(\beta^n_1, \ldots, \beta^n_n)$ with $\alpha \in [M]$, it is also bounded if we choose $\alpha$ independently according to $\omega$. That is, we have

$$\varepsilon \geq \Pr_{S, \alpha \in [M]}[\alpha \in B(\beta, S, \xi)] = \frac{\Pr[\alpha \in B(\beta, S, \xi)]}{S} \cdot$$

Markov’s inequality $\Pr[Z \geq c] \leq \mathbb{E}[Z]/c$ with $c = \sqrt{\varepsilon}$ applied to the random variable $Z(S) = \Pr_{\alpha}[\alpha \in B(S)]$ immediately gives the claim. \qed

**6.4 Sampling $\lambda$-good paths**

We now apply the concept of a parallel sampler to the situation of interest. Recall Definition 6.8 of the set $\Gamma(\lambda, S) \subset [m]^n$ of $\lambda$-good paths for every $\lambda > 0$ and $S \subset [n]$. We show that for an appropriate choice of $\lambda$, and a fixed probability distribution $\omega$ on $[m]^n$, the weight of the $\lambda$-good paths for $S$ is large with high probability if $S \subset [n]$ is a random subset which is a parallel sampler.

**Theorem 6.16 (“Sampling”).** Let $\omega$ be an arbitrary probability distribution on $[m]^n$. Let $P_S$ be a probability distribution over subsets of $[n]$ which is a $(n, \xi, \delta, \varepsilon)$-parallel sampler. Then

$$\Pr_S[\omega(\Gamma(\lambda, S)) \geq 1 - \delta] \geq 1 - \varepsilon$$

for

$$\lambda := \frac{H(X_{>0}|E)\#_\phi}{n \log |X|} + \frac{n - |S|}{|S|n \log |X|}H(\emptyset|E)\#_\phi - \left(\frac{1}{m} + \xi\right),$$

where $\Gamma(\lambda, S)$ is the set of $\lambda$-good paths as in Definition 6.8, i.e., the set of $\alpha^n \in [m]^n$ with

$$H(\emptyset|E)\#_{\alpha^n} + \sum_{j \in S} H(X_j|X_{>j}|E)_{\rho_{\alpha^n}} \geq \lambda|S|\log |X|.$$

**Proof.** For every $j \in [n]$ and $\alpha^n \in [m]^n$, we define the quantity

$$\beta^n_\alpha = \frac{H(X_j|X_{>j}|E)_{\rho_{\alpha^n}}}{\log |X|}.$$

(Note that this depends only on the first $j$ entries of $\alpha^n$.) Observe that we have $\beta^n_\alpha \in [0, 1]$ by the dimension bound (Lemma 5.2 (ii)). By Definition 6.14 of a parallel sampler, we therefore get

$$\Pr_S[\omega(B(\beta, S, \xi)) \geq \delta] \leq \varepsilon,$$

where

$$B(\beta, S, \xi) = \{\alpha^n \in [m]^n \mid \frac{1}{|S|} \sum_{i \in S} \beta^n_i \leq \frac{1}{n} \sum_{i \in [n]} \beta^n_i - \xi\}. $$

Inequality (59) can be rewritten as

$$\Pr_S[\omega(B(\beta, S, \xi)) \geq 1 - \delta] \geq 1 - \varepsilon,$$
where we write \( \overline{B}(\beta, S, \xi) = [m^n] \setminus B(\beta, S, \xi) \) for the complement of \( B(\beta, S, \xi) \).

Note that if \( \alpha^n \in \overline{B}(\beta, S, \xi) \), then
\[
\frac{\log |X|}{|S|} \sum_{i \in S} \beta_i^n \geq \frac{\log |X|}{n} \sum_{i \in [n]} \beta_i^n - \xi \log |X| \\
\geq \frac{1}{n} \left( H(X_{\neq 0}|E) - \frac{n \log |X|}{m} - H(\emptyset|E) \right) - \xi \log |X| 
\]
by the definition of \( \beta_i^n \) and Theorem 6.4. This is equivalent to
\[
\mathbf{v}_\rho^S(T_n, \alpha^n) = H(\emptyset|E) \omega^n + \sum_{j \in S} H(X_j|X_{\neq j}E)_{\rho^j} \geq \lambda^n |S| \log |X| \tag{61}
\]
where
\[
\lambda^n := \frac{H(X_{\neq 0}|E)}{n \log |X|} - \frac{1}{m} \frac{n - |S|}{|S| n \log |X|} H(\emptyset|E) - \xi .
\]
We use Lemma 6.3 (v) (with \( j = n \)) to bound the second summand from below, getting \( \lambda^n \geq \lambda \) for all \( \alpha^n \in [m^n] \). With (61), we conclude that
\[
\mathbf{v}_\rho^S(T_n, \alpha^n) \geq \lambda |S| \log |X| \quad \text{for all } \alpha^n \in \overline{B}(\beta, S, \xi) . \tag{62}
\]
In other words, we have \( \overline{B}(\beta, S, \xi) \subset \Gamma(\lambda, S) \), and the claim follows from (60). \( \Box \)

### 6.5 Sampling and recombining: preservation of smooth entropy rate

We will now turn our attention to the smooth min-entropy, as introduced in [Ren05]. We will state and prove our main result in this section; that is, we will show that smooth min-entropy rate is preserved under sampling.

Before discussing our main result, we quickly review an important special case: We will often consider situations where a random variable \( Z = f(X, Y) \) is the result of applying a function to two random variables \( X \) and \( Y \). An example of this is the case where \( Z \) is a randomly chosen substring of \( X \). To show that the uncertainty about \( f(X, Y) \) is large given \( Y \) and a quantum system \( E \), it suffices to show that with high probability over \( Y \), the uncertainty about \( f(X, y) \) is large. This is expressed by the following result.

**Lemma 6.17.** Let \( \rho_{ZYE} \) be such that
\[
\Pr_y \left[ H_{\min}^k(Z|E, Y = y) \geq k \right] \geq 1 - \varepsilon .
\]
Then \( H_{\min}^{k+\varepsilon}(Z|Y E) \geq k \).

The proof of this lemma is deferred to Appendix A

Recall that the smooth min-entropy-rate \( R_{\min}^n(A|B)_\rho \) is defined as in (8) as the smooth min-entropy \( H_{\min}^n(A|B)_\rho \) divided by the size \( H_0(A) \) of \( A \). Our main result is the following

**Theorem 6.18.** Let \( \rho_{X^nE} \) be a quantum state where \( X^n = (X_1, \ldots, X_n) \) on \( X^n \) is classical. Let \( S \) be a random variable over subsets of \([n]\) which is independent of \( X^n E \) and a \( (n, \xi, \delta, \varepsilon) \)-parallel sampler. Assume that \( \kappa = \frac{n}{|S| \log |X^n|} \leq 0.15 \). Then
\[
R_{\min}^{2\sqrt{3}+\varepsilon+20+\tau}(X_\perp S|SE)_\rho \geq R_{\min}^n(X^n|E)_{\rho} - \Delta \quad \text{where}
\]
\[
\Delta = \xi + \frac{2 \log 1/\theta}{n \log |X^n|} + 2 \kappa \log 1/\kappa ,
\]
for all \( \theta, \tau \geq 0 \).
We will give concrete parameters below, which show that $\Delta \to 0$ (in some security parameter), in situations of interest. To put this result into a more convenient form, we choose a certain value of $\theta$, and show how this result applies to general samplers.

**Corollary 6.19.** Let $\rho_{X^nE}$ be a quantum state as in Theorem 6.18 and let $S$ be a $(n, \xi, \varepsilon)$-sampler. Assume that $\kappa = \frac{n |S| \log |X|}{\log |X|} \leq 0.15$. Then

$$R_{\text{min}}^{\varepsilon + \tau} (XS|SE)_{\rho} \geq R_{\text{min}}^{\tau} (X^n|E)_{\rho} - 3\xi - 2\kappa \log 1/\kappa$$

for all $\tau \geq 0$.

**Proof.** We choose $\theta = 2^{-\xi n \log |X|}$. We can then bound $\Delta$ in Theorem 6.18 by

$$\Delta \leq 3\xi + 2\kappa \log 1/\kappa$$

and the claim follows from the fact that a $(n, \xi, \varepsilon)$-sampler is a $(n, \xi, \sqrt{\varepsilon}, \sqrt{\varepsilon})$-parallel sampler (i.e., Lemma 6.15).

In the remainder of this section, we prove Theorem 6.18. We do so in two successive steps. We first show that sampling preserves the entropy rate of a modified (smooth) entropy $h_{\text{min}}^\varepsilon(A|B)$. We then use the fact that this modified entropy $h_{\text{min}}^\varepsilon$ is essentially equivalent to the smooth min-entropy. More precisely, we introduce the quantities

$$h_{\text{min}}(A|B)_{\rho} = \sup_{\sigma_B \geq \rho_B} H(A|B)_{\sigma}$$

$$h_{\text{min}}^\varepsilon(A|B)_{\rho} = \sup_{\bar{\rho}_{AB} : \|\bar{\rho}_{AB} - \rho_{AB}\| \leq \varepsilon} h_{\text{min}}(A|B)_{\bar{\rho}}$$

for any bipartite state $\rho_{AB}$ and $\varepsilon \geq 0$. The only difference to the original definition of the (smooth) min-entropy $H_{\text{min}}^\varepsilon(A|B)_{\rho}$ (Definition (25)) is that the supremum is restricted to states $\sigma_B$ which are bounded from below by $\rho_B$. These quantities give the bounds

$$h_{\text{min}}^{\varepsilon + \delta}(A|B)_{\rho} + 2 \log 1/\varepsilon \geq H_{\text{min}}^{\delta}(A|B)_{\rho} \geq h_{\text{min}}^\delta(A|B)_{\rho}$$

for all $\varepsilon, \delta \geq 0$ (63) on the smooth min-entropy, for all states $\rho_{AB}$. Note that the second inequality follows trivially from the definition; we give a proof of the first inequality in Appendix A (Lemma A.1).

We are ready to combine the recombination theorem (Theorem 6.13) with the sampling theorem (Theorem 6.18). This gives the following main result, which shows that the min-entropy rate (for the modified entropy $h_{\text{min}}^\varepsilon(A|B)_{\rho}$) is preserved under sampling.

**Lemma 6.20.** Consider a quantum state of the form $\rho_{X^nE}$ where $X^n = (X_1, \ldots, X_n)$ on $X^n$ is classical. Let $P_S$ be a probability distribution over subsets of $[n]$ which is a $(n, \xi, \varepsilon)$-parallel sampler. Then

$$\Pr_{S} \left[ \frac{h_{\text{min}}^\varepsilon(XS|E)_{\rho}}{|S| \log |X|} \geq \frac{h_{\text{min}}(X^n|E)_{\rho}}{n \log |X|} - c \right] \geq 1 - \varepsilon$$

where

$$c = \xi + \frac{1}{m} + \frac{2n \log m}{|S| \log |X|}$$

for any $m \in \mathbb{N}$. In particular, inequality (64) is true for the choice

$$c = \xi + 2\kappa \log 1/\kappa$$

if $\kappa = \frac{n |S| \log |X|}{|X|} \leq 0.15$. 



40
Proof. We first show the second part, assuming that the first statement is true. It is obtained by choosing a specific value of \( m \in \mathbb{N} \). Consider the function \( f(m) := \frac{1}{m} + 2\kappa \log m \). We are interested in the minimum value of this function for \( m \in \mathbb{N} \). It is easy to see that the function is minimised for \( m_{\text{min}} = \frac{\ln 2}{2\kappa} \), where \( \ln \) denotes the natural logarithm. However, since this is not necessarily an integer, we use the value \( f\left( \frac{1}{\ln 2} \cdot m_{\text{min}} \right) \). Clearly, for \( \kappa \) small enough (\( \kappa \leq \frac{1}{2}(1 - \ln 2) \approx 0.15 \)), there is an integer \( m_0 \in [m_{\text{min}}, \frac{1}{\ln 2} \cdot m_{\text{min}}] \), and this integer satisfies \( f(m_0) \leq f\left( \frac{m_{\text{min}}}{\ln 2} \right) = 2\kappa \log 1/\kappa \).

The second claim immediately follows from this.

We rephrase the first claim more explicitly: We have to show that the following holds with probability at least \( 1 - \varepsilon \) over the choice of \( S \). There is a subnormalised state \( \bar{\rho}_{X^*\|E} \) (depending on \( S \)) at distance

\[
\frac{1}{2} \| \rho_{X^*\|E} - \bar{\rho}_{X^*\|E} \| \leq \sqrt{\delta}
\]

from the original state \( \rho_{X^*\|E} \) and a state \( \sigma_E \geq \rho_E \) (which happens to be independent of \( S \)) such that

\[
H(X_S|E)_\rho \geq \frac{|S|}{n} h_{\text{min}}(X_{>0}|E)_{\rho} - |S| \log |X| (\frac{1}{m} + \xi) - 2n \log m
\]

\[
H(\emptyset|E)_\rho \geq 0.
\]

Let \( \sigma_E \) be a state which achieves the supremum in the definition of \( h_{\text{min}}(X_{>0}|E)_{\rho} \), i.e., we have \( h_{\text{min}}(X_{>0}|E)_{\rho} = H(X_{>0}|E)_{\rho} \) and \( \sigma_E \geq \rho_E \). Then \( H(\emptyset|E)_{\rho} \geq 0 \) by definition, and we can bound the quantity \( \lambda \) in Theorem 6.16 by

\[
\lambda \geq \frac{H(X_{>0}|E)_{\rho}}{n \log |X|} - \left( \frac{1}{m} + \xi \right).
\]

According to Theorem 6.16, the set \( \Gamma(\lambda, S) \) of \( \lambda \)-good paths has weight at least \( 1 - \delta \) with respect to the distribution \( \omega \) (defined by Theorem 6.13), with probability at least \( 1 - \varepsilon \) over the choice of \( S \). The recombination theorem (Theorem 6.13) therefore guarantees the existence of a state \( \bar{\rho}_{X^*\|E} \) with the required properties, except with probability \( \varepsilon \) over the choice of \( S \).

We can now prove our main result.

Proof of Theorem 6.18. First observe that Lemma 6.20 can be adapted using the triangle inequality to include an additional parameter \( \gamma \geq 0 \), thus replacing the probability in question by

\[
\Pr_S \left[ \frac{h_{\text{min}}^\gamma (X_S|E)_{\rho}}{|S| \log |X|} + \frac{h_{\text{min}}^\gamma (X^n|E)_{\rho}}{|S| \log |X|} - c \right] \geq 1 - \varepsilon.
\]

Setting \( \gamma = 2\theta + \tau \) and \( \Delta = c + \frac{2 \log 1/\theta}{n \log |X|} \), this implies that

\[
\Pr_S \left[ \frac{H_{\text{min}}^{2\gamma + 2\theta + \tau} (X_S|E)_{\rho}}{|S| \log |X|} - \frac{H_{\text{min}}^{2\gamma + 2\theta + \tau} (X^n|E)_{\rho}}{|S| \log |X|} - \Delta \right] \geq 1 - \varepsilon
\]

because of the relations (63) between \( h_{\text{min}} \) and \( H_{\text{min}} \). The claim of the theorem follows from this inequality and Lemma 6.17.
for all $\tau$ a random substring of a given string $Z$.

Similarly, we have $\kappa$ key expansion. This will result in a concrete protocol for the bounded storage model which achieves significant concreteness and simplicity, we restrict our attention to this sampler; in practice, more efficient constructions may be used, and the analysis is analogous.

Figure 10: The subprotocol $\text{Samp}$. We slightly abuse notation by identifying the random bits with the subset we choose at random.

## 7 Recursive sampling and the bounded storage model

We will now consider the random subset sampler and analyse recursive sampling from a string. This will result in a concrete protocol for the bounded storage model which achieves significant key expansion.

The basic building block is the subprotocol $\text{Samp}$ described in Figure 10. This protocol outputs a random substring of a given string $Z$. The effect of protocol $\text{Samp}$ is the following.

**Lemma 7.1.** Let $r$ be fixed, let $\rho_Z E$ be a quantum state where $Z$ is an $L$-bitstring with $L \geq r^4$. Let $S$ be an independent, uniform $\lceil \log (\frac{rL^{1/4}}{r}) \rceil$-bit string. (In particular, these are less than $r \log L$ bits.) Let $Z'$ be the $L^{3/4}$-bitstring $Z' = \text{Samp}(Z, r, S)$. Then

$$R^{\epsilon' + \tau'}_{\text{min}}(Z | ES)_\rho \geq R^{\tau'}_{\text{min}}(Z | ES)_\rho - 5 \log \frac{r}{r^{1/4}},$$

for all $\tau \geq 0$, where $\epsilon' = 5 \cdot 2^{\frac{-1}{T/8}}$.

Note that we could have used any $(n, \xi, \epsilon)$-sampler in place of the subset sampler. However, for concreteness and simplicity, we restrict our attention to this sampler; in practice, more efficient constructions may be used, and the analysis is analogous.

**Proof.** To prove the bound $r \log L$ on the number of bits consumed, we use the inequality $(\binom{p}{q}) \leq \left(\frac{e p}{q}\right)^q$ on the binomial coefficients. It implies that

$$\lceil \log \left(\frac{rL^{1/4}}{r}\right) \rceil \leq \log (L^{1/4} e^{r}) \leq r \left(\frac{\log L}{4} + \log e\right) \leq r \log L.$$

We express everything in terms of the number $r$ of subblocks we sample, and the length $L^{3/4}$ of the final string. That is, we sample $r < n$ subblocks from $n = \frac{L}{r} = L^{1/4} r$ blocks of size $t = \frac{L^{3/4}}{r}$ bits each, obtaining a substring of $rt = L^{3/4}$ bits.

We know from Lemma 2.2 that a randomly chosen subset $S$ of size $r < n$ is a $(n, \xi, e^{-r \xi^2/2})$-sampler, for every $\xi \in [0, 1]$. We choose $\xi = \frac{1}{r^{1/4}} \psi$ such that $e^{-r \xi^2/2} = e^{-\sqrt{r}/2}$. The parameter $\epsilon'$ in Corollary 6.19 then takes the form

$$\epsilon' = 2 \cdot 2^{-n \log |X|} + 3 (e^{-\sqrt{r}/2})^{1/4} = 2 \cdot 2^{-L/r^{1/4}} + 3 e^{-\sqrt{r}/8} \leq 5 \cdot 2^{-\sqrt{r}/8}.$$

Similarly, we have $\kappa = \frac{r}{L^{1/4}}$. Because the function $\kappa \mapsto 2 \kappa \log 1/\kappa$ is monotonically increasing for small enough $\kappa$, its value is maximised for small values of $L$; that is, we can use our lower bound
ReSample($Z, f, r, S$):

**Input:** $L$-bit string $Z$, parameters $r$ and $f$ and independent random bits $S$.

**Output:** $(3/4)^f$-bit substring of $Z$.

**Procedure:** Let $Z^{(0)} = Z$. Iterate the following, for $i = 1, \ldots, f$:

1. Use independent random bits $S^{(i)}$ from $S$ to generate $Z^{(i)} = \text{Samp}(Z^{(i-1)}, r, S^{(i)})$.
2. Output $Z^{(f)}$.

Figure 11: The protocol ReSamp. It calls the subprotocol Samp $f$ times, each time producing a substring of the already generated string. We will determine the amount of randomness this protocol needs below. Note that the output of this recursive protocol can be computed with limited storage. In particular, it is unnecessary to store the intermediate substrings $Z^{(i)}$.

Let $L \geq r^4$ on $L$ get

$$2 \kappa \log 1/\kappa = 2 \frac{r}{L^{1/2}} \log \frac{L^{1/2}}{r} \leq 2 \frac{\log r}{r}.$$  

We conclude that

$$3 \xi + 2 \kappa \log 1/\kappa \leq \frac{3}{r^{1/4}} + \frac{2}{r} \log r \leq \frac{5 \log r}{r^{1/4}}.$$

The claim then follows from Corollary 6.19.

Note that the length $L$ is only reduced to $L^{3/4}$ by the protocol Samp. To reduce the length of the output even further, we use the protocol recursively. That is, we randomly sample substrings $f$ times, each time sampling a substring of the already obtained string. In each step, the original string is partitioned into a certain number of blocks, out of which $r$ are chosen at random (throughout, $r$ will be a fixed parameter). We call the resulting protocol ReSamp; see Figure 11.

To understand the effect of this recursive protocol, observe that the quantities in Lemma 7.1 describing the effect of Samp are all additive in the following sense: repeated application of Samp simply requires addition of the parameters. Moreover, with the chosen parameters, only the number of bits consumed depends on the length of the involved bitstrings. The analysis is therefore particularly simple, and the effect of the procedure ReSamp is described by the following lemma.

**Lemma 7.2.** Let $\rho_{ZE}$ be such that $Z$ is an $L$-bit string. Let $f$ and $L$ be such that $L^{(3/4)} \geq r^4$. Let $S$ be independent random bits and let $Z' = \text{ReSamp}(Z, f, r, S)$. Then $Z'$ is a $(3/4)^f$-bit substring of $Z$, with

$$R_{\min}^{\xi + \gamma}(Z'|ES)_p \geq R_{\min}^{\xi}(Z|ES)_p - 5^f \log \frac{r}{r^{1/4}} \quad \text{(65)}$$

where $\gamma = 5f \cdot 2^{-\sqrt{2}/8}$. The generation of $Z'$ consumes less than $fr \log L$ independent random bits from $S$.

In particular, if $L = 2^r$, then approximately $f \approx \frac{1}{\log 4/\gamma} \log \left( \frac{r}{4 \log r} \right)$ applications of the subprotocol Samp are sufficient to produce a substring $Z'$ of $Z$ of length $\lesssim r^4$, while preserving the min-entropy-rate (for large $r$). This consumes less than $r^3$ independent random bits.

**Proof.** Let $L^{(i)}$ be the length of the string $Z^{(i)}$; by the definition of the protocol Samp, we have $L^{(i)} = (L^{(i-1)})^{3/4}$ and hence $L^{(f)} = L^{(3/4)^f}$, as claimed. Let $S^{(i)}$ be the random bits from $S$ used
to generate $Z^{(i)}$. According to Lemma 7.1, we have for all $i = 1, \ldots, f$

$$R_{\min}^{(i-1)+\delta} (Z^{(i)}|ES^{(1)} \ldots S^{(i)})_p \geq R_{\min}^{(i-1)} (Z^{(i-1)}|ES^{(1)} \ldots S^{(i-1)})_p - 5 \log \frac{r}{r^{1/4}},$$

where $\delta = 4 \cdot 2^{-\sqrt{7}/8}$ and $\varepsilon^{(i-1)}$ is arbitrary. Defining $\varepsilon^{(0)} = \varepsilon$ and $\varepsilon^{(i)} = \varepsilon^{(i-1)} + \delta$, this implies the claim (65).

To compute the number of random bits consumed in this procedure\(^20\), observe that in the $i$-th step, a random subset $S^{(i)}$ of $[n^{(i)}]$ of size $r$ is chosen, where $n^{(i)} = r(L^{(i-1)})^{1/4}$; this consumes $\lceil (n^{(i)}) \rceil$ bits.

The total number of bits we need is bounded by the number of applications times the maximal number of bits $\max_n \lceil \log (\binom{n}{r}) \rceil = \lceil \log (rL^{1/4}) \rceil \leq r \log L$ consumed in a single step. (Here we used the fact that $n^{(i)}$ is a decreasing sequence and the bound on the binomial coefficient shown in Lemma 7.1.) This gives the upper bound $fr \log L$, as claimed. \(\square\)

In summary, we have found a procedure that generates a random substring of a string of $2^r$ bits, with the following parameters:

| Length (bits) | Original $Z$ | Substring $Z'$ | Seed $S$ |
|--------------|--------------|----------------|--------|
| Entropy-rate  | $2^r$        | $r^4$          | $r^3$  |

with error $\text{poly}(\log(r))e^{-\Omega(\sqrt{r})}$. It is computable with $\text{poly}(r)$ bits of storage. Subsequent to this sampling procedure, privacy amplification may be used to extract a secret key from the substring $Z'$. In conclusion, this gives a sample-and-hash procedure for key expansion in the bounded storage model, expanding an initial key of $r^3$ bits to approximately $r^4(R - 1/r^{\Omega(1)})$ bits. ($R \in [0, 1]$ is usually assumed to be constant.)

### A Additional proofs related to entropies

**Proof of Lemma 6.14** Let $\rho_{YZE} = \sum_y P_Y(y)|y\rangle\langle y| \otimes \rho_{ZE|Y=y}$. Let

$$\mathcal{G} := \{ y \in \mathcal{Y} \mid H_{\min}^\delta(Z|E, Y = y) \geq k \} .$$

For every $y \in \mathcal{G}$, there is a subnormalised state $\tilde{\rho}_{ZE}^y$ and a state $\sigma_E^y$ such that

$$\tilde{\rho}_{ZE}^y \leq 2^{-k} \sigma_E^y$$

$$\|\tilde{\rho}_{ZE}^y - \rho_{ZE|Y=y}\| \leq \delta$$

by definition. For every $y \notin \mathcal{G}$, we choose arbitrary states $\tilde{\rho}_{ZE}^y$ and $\sigma_E^y$ satisfying

$$\tilde{\rho}_{ZE}^y \leq 2^{-k} \sigma_E^y .$$

It is easy to verify that the two states

$$\tilde{\rho}_{YZE} = \sum_y P_Y(y)|y\rangle\langle y| \otimes \tilde{\rho}_{ZE}^y$$

$$\sigma_{YE} = \sum_y P_Y(y)|y\rangle\langle y| \otimes \sigma_E^y$$

satisfy

$$\tilde{\rho}_{YZE} \leq 2^{-k} \sigma_{YE}$$

$$\|\tilde{\rho}_{YZE} - \rho_{YZE}\| \leq \delta + \varepsilon .$$

The claim follows from this. \(\square\)

\(^20\)Note that we are only interested in the approximate number of bits we need; see [DHRS04] for a dense encoding of subsets of $[n]$ into bitstrings.
We next prove the non-trivial inequality in \((63)\).

**Lemma A.1.** Let \(\rho_{AB}\) be a subnormalised state. We have

\[
\text{h}_{\min}^{2\varepsilon+\delta}(A|B)_\rho \geq H_{\min}^{\varepsilon}(A|B)_\rho - 2\log 1/\varepsilon .
\]

(66)

**Proof.** Clearly, it suffices to show that the claim of the lemma is true for \(\delta = 0\).

By definition, there is a normalised state \(\sigma_B\) such that

\[
\rho_{AB} \leq 2^{-H_{\min}(A|B)_\rho} \sigma_B .
\]

This implies that

\[
\frac{\rho_{AB}}{\rho_B} \leq 2^{-H_{\min}(A|B)_\rho} \sigma_B .
\]

Let \(P_B\) denote the projector onto the eigenspaces of \(\sigma_B/\rho_B\) corresponding to eigenvalues smaller than or equal to \(1/\varepsilon^2\). Applying this on both sides of the previous inequality gives (with \(P_B \leq \text{id}_B\))

\[
P_B \frac{\rho_{AB}}{\rho_B} P_B \leq \frac{2^{-H_{\min}(A|B)_\rho}}{\varepsilon^2} \text{id}_B .
\]

Multiplying from both sides by \(\rho_B^{1/2}\), we obtain

\[
\tilde{\rho}_{AB} \leq \frac{2^{-H_{\min}(A|B)_\rho}}{\varepsilon^2} \rho_B ,
\]

(67)

where we introduced the operator \(\tilde{\rho}_{AB} = \rho_B^{1/2} P_B \frac{\rho_{AB}}{\rho_B} P_B \rho_B^{1/2}\). We claim that

\[
\tilde{\rho}_B \leq \rho_B ,
\]

(68)

\[
\text{tr}(\tilde{\rho}_{AB}) \leq 1 \quad \text{and}
\]

(69)

\[
\frac{1}{2} \| \tilde{\rho}_{AB} - \rho_{AB} \| \leq \varepsilon .
\]

(70)

Note that \((67)-(70)\) imply the claim \((66)\) (for \(\delta = 0\)).

Inequality \((65)\) directly follows from the fact that \(\text{tr}_A(\tilde{\rho}_{AB}/\rho_B) = \tilde{\rho}_B \leq \text{id}_B\). To prove \((70)\), let \(|\Psi_{ABC}\rangle\) be a purification of \(\rho_{AB}\) and consider the purification

\[
|\tilde{\Psi}_{ABC}\rangle = \rho_B^{1/2} P_B \rho_B^{-1/2} |\Psi_{ABC}\rangle
\]

of \(\tilde{\rho}_{AB}\). Using the Schmidt decomposition \(|\Psi_{ABC}\rangle = \sum_{\lambda} \sqrt{\lambda} |\lambda\rangle_{AC}|\lambda\rangle_B\), it is straightforward to verify that

\[
\langle \Psi_{ABC}|\tilde{\Psi}_{ABC}\rangle = \text{tr}(P_B \rho_B) \quad \text{and} \quad \langle \tilde{\Psi}_{ABC}|\tilde{\Psi}_{ABC}\rangle \leq \text{tr}(P_B \rho_B) \leq \text{tr}(\rho_B) .
\]

(71)

Note that the latter of these inequalities proves \((69)\). The Cauchy-Schwarz inequality \(\sum_{i=1}^d |\lambda_i| \leq \sqrt{d} \sqrt{\sum_{i=1}^d \lambda_i^2}\) implies that for any two pure states \(|\chi\rangle, |\phi\rangle\), we have

\[
\| |\phi\rangle \langle \phi| - |\chi\rangle \langle \chi| \| \leq \sqrt{2} \| |\phi\rangle \langle \phi| - |\chi\rangle \langle \chi| \| = \sqrt{2} \sqrt{|\langle \phi| \langle \phi| - 2|\langle \phi| \chi\rangle|^2 + |\langle \chi\rangle|^2 ,}
\]

where \(\|A\|_2 = \sqrt{\text{tr}(A^\dagger A)}\), since the difference \(|\phi\rangle \langle \phi| - |\chi\rangle \langle \chi|\) has rank at most 2. Applying this to \(|\Psi_{ABC}\rangle\) and \(|\tilde{\Psi}_{ABC}\rangle\) and using \((71)\) gives

\[
\| |\tilde{\Psi}_{ABC}\rangle \langle \tilde{\Psi}_{ABC}| - |\Psi_{ABC}\rangle \langle \Psi_{ABC}| \| \leq \sqrt{2} \sqrt{\text{tr}(\rho_B)^2 - \text{tr}(P_B \rho_B)^2}
\]

\[
\leq \sqrt{2} \sqrt{\text{tr}(\rho_B) - \text{tr}(P_B \rho_B)) (\text{tr}(\rho_B) + \text{tr}(P_B \rho_B))}
\]

\[
\leq 2 \sqrt{\text{tr}(P_B \rho_B)} .
\]
Here $P_B^\perp = \text{id}_B - P_B$ projects onto the orthogonal complement of the image of $P_B$. In the last inequality, we have used the assumption that $\rho_{AB}$ is subnormalised and thus $\text{tr}(\rho_B) \leq 1$. Since the trace distance is non-increasing under partial traces, the claim \[ (70) \] follows once we show that
\[
\text{tr}(P_B^\perp \rho_B) \leq \epsilon^2.
\]
Note that we have $\frac{1}{\epsilon^2} P_B^\perp \leq P_B^\perp \rho_B P_B^\perp$ by definition. Inserting this into the expression of interest gives
\[
\text{tr}(P_B^\perp \rho_B) \leq \epsilon^2 \text{tr} \left( P_B^\perp \rho_B \right) \leq \epsilon^2 \text{tr} \left( \frac{\sigma_B}{\rho_B} \rho_B \right) \leq \epsilon^2 \text{tr}(\sigma_B) = \epsilon^2,
\]
as claimed. \qed

\section{Additional lemmas}

\textbf{Lemma B.1.} Let $\{Q^\alpha\}_{\alpha \in [m]}$ be a family of Hermitian operators on a Hilbert space $A \otimes B$ and suppose that $\text{tr}_B(Q^\alpha \rho_{AB} Q^\alpha) \leq \sigma_A$, for any $\alpha \in [m]$. Then
\[
\text{tr}_B(Q \rho_{AB} Q) \leq m^2 \sigma_A
\]
for $Q := \sum_\alpha Q^\alpha$. In particular, if $\{Q^\alpha\}_{\alpha \in [m]}$ resolves the identity on $\text{supp}(\rho_{AB})$ then $\rho_A \leq m^2 \sigma_A$.

\textit{Proof.} Let $|\varphi_A\rangle \in A$ be arbitrary. By definition, we have
\[
\text{tr} \left( \text{tr}_B(Q \rho_{AB} Q) |\varphi_A\rangle \langle \varphi_A| \right) = \text{tr} \left( Q \rho_{AB} Q |\varphi_A\rangle \langle \varphi_A| \otimes \text{id}_B \right) = \sum_{\alpha, \beta} \text{tr} \left( Q^\alpha \rho_{AB} Q^\beta |\varphi_A\rangle \langle \varphi_A| \otimes \text{id}_B \right) \tag{72}
\]
By the cyclicity of the trace and the fact that $|\varphi_A\rangle \langle \varphi_A| \otimes \text{id}_B$ is a projector, we have
\[
\text{tr} \left( Q^\alpha \rho_{AB} Q^\beta |\varphi_A\rangle \langle \varphi_A| \otimes \text{id}_B \right) = \text{tr} \left( Z^\alpha (Z^\beta)^\dagger \right),
\]
with $Z^\alpha = (|\varphi_A\rangle \langle \varphi_A| \otimes \text{id}_B) Q^\alpha \rho_{AB}$. The operator-Cauchy-Schwarz-inequality
\[
\text{tr}(EF) \leq \sqrt{\text{tr}(E) \text{tr}(F)}
\]
applied to $E = Z^\alpha$ and $F = (Z^\beta)^\dagger$ therefore gives (with the cyclicity of the trace)
\[
\text{tr} \left( Q^\alpha \rho_{AB} Q^\beta |\varphi_A\rangle \langle \varphi_A| \otimes \text{id}_B \right) \leq \sqrt{\text{tr} \left( (Z^\alpha)^\dagger Z^\alpha \right) \text{tr} \left( (Z^\beta)^\dagger Z^\beta \right) } \tag{73}
\]
It is straightforward to verify that
\[
\text{tr} \left( (Z^\alpha)^\dagger Z^\alpha \right) = \text{tr} \left( \text{tr}_B(Q^\alpha \rho_{AB} Q^\alpha) |\varphi_A\rangle \langle \varphi_A| \right) \leq \text{tr}(\sigma_A |\varphi_A\rangle \langle \varphi_A|) \quad \text{for all } \alpha \in [m], \tag{74}
\]
where we used the assumption in the last inequality. Combining \[ (72) \] with \[ (73) \] and \[ (74) \] gives
\[
\text{tr} \left( \text{tr}_B(Q \rho_{AB} Q) |\varphi_A\rangle \langle \varphi_A| \right) \leq m^2 \sigma_A |\varphi_A\rangle \langle \varphi_A| \right),
\]
Since $|\varphi_A\rangle \in A$ was arbitrary, the claim follows. \qed

\textbf{Lemma B.2.} Let $P$ and $P'$ be two projectors on a Hilbert space $\mathcal{H}$. Then
\begin{enumerate}
  \item[(i)] If $\text{supp} P \subseteq \text{supp} P'$, then $P \leq P'$.
  \item[(ii)] If $P \leq P'$, then $PP' = P'P = P$.
\end{enumerate}

\textit{Proof.} Both statements follow immediately from the fact that
\[
\text{supp} P' = \text{supp} P \oplus (\text{supp} P)^\perp,
\]
where $(\text{supp} P)^\perp$ is the orthogonal complement of $\text{supp} P$ in $\text{supp} P'$. This identity implies $P' = P + \text{id}_{(\text{supp} P)^\perp}$, where $\text{id}_{(\text{supp} P)^\perp}$ is the projector onto $(\text{supp} P)^\perp$. Thus $PP' = P'P = P$. \qed

46
References

[ANTSV99] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the 31th ACM Symposium on Theory of Computing*, 1999. arXiv:quant-ph/9804043.

[BARd07] A. Ben-Aroya, O. Regev, and R. de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing, 2007. arXiv.org:0705.3806.

[BBCM95] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer. Generalized privacy amplification. *IEEE Transaction on Information Theory, 41*(6):1915–1923, 1995.

[BBR88] C. H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM Journal on Computing, 17*(2):210–229, 1988.

[BH05] L. Babai and T. P. Hayes. Near-independence of permutations and an almost sure polynomial bound on the diameter of the symmetric group. In *SODA ’05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1057–1066, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[DHRS04] Y. Z. Ding, D. Harnik, A. Rosen, and R. Shaltiel. Constant-round oblivious transfer in the bounded storage model. In *Second Theory of Cryptography Conference, TCC 2004*, pages 446–472, 2004.

[DM02] S. Dziembowski and U. Maurer. Tight security proofs for the bounded-storage model. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 341–350. ACM, May 2002.

[DM04] S. Dziembowski and U. Maurer. Optimal randomizer efficiency in the bounded-storage model. *Journal of Cryptology, 17*(1):5–26, January 2004. (Conference version appeared in Proc. of STOC’02.).

[FS07] S. Fehr and C. Schaffner. Randomness extraction via delta-biased masking in the presence of a quantum attacker, 2007. arXiv:0706.2606.

[GKK+07] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and Ronald de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *to appear in STOC’07*, pages 369–376, 2007. arXiv:quant-ph/0611209.

[Gol97] O. Goldreich. A sampler of samplers: A computational perspective on sampling, 1997. [http://www.eccc.uni-trier.de/eccc/](http://www.eccc.uni-trier.de/eccc/)

[Hol73] A. S. Holevo. Statistical problems in quantum physics. In *Proceedings of the Second Japan-USSR Symposium on Probability Theory*, Lecture Notes in Mathematics, pages 104–119. Springer, 1973.

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *STOC ’89: Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 12–24, New York, NY, USA, 1989. ACM.

[KMR05] R. König, U. Maurer, and R. Renner. On the power of quantum memory. *IEEE Transaction on Information Theory, 51*(7):2391–2401, July 2005. arXiv:quant-ph/0305154.

[KSR07] R. König, C. Schaffner, and R. Renner. The operational meaning of conditional min-entropy. to appear, 2007.

[KT07] R. König and B. M. Terhal. The bounded storage model in the presence of a quantum adversary. *IEEE Transactions on Information Theory, 2007*. to appear; arXiv:quant-ph/0608101
[Lei07] M. S. Leifer. Conditional density operators and the subjectivity of quantum operations. In G. Adenier, C. A. Fuchs, and A. Yu. Khrennikov, editors, *Foundations of Probability and Physics-4, AIP Conference Proceeding*, volume 889, pages 172–186, 2007. arXiv:quant-ph/0611233.

[Lu02] C. Lu. Hyper-encryption against space-bounded adversaries from on-line strong extractors. In *Advances in Cryptology — CRYPTO 2003*, Lecture Notes in Computer Science, pages 18–22. Springer, 2002.

[Mau92] U. Maurer. Conditionally-perfect secrecy and a provably-secure randomized cipher. *J. of Cryptology*, 5(1):53–66, 1992.

[Nay99] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of 40th IEEE FOCS*, pages 369–376, 1999. arXiv:quant-ph/9904093.

[NZ96] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52:43–52, 1996. A preliminary version appeared at STOC ’93.

[ON02] T. Ogawa and H. Nagaoka. New proof of the channel coding theorem via hypothesis testing in quantum information theory, 2002. arXiv:quant-ph/0208139.

[Ren05] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich, 2005. arXiv:quant-ph/0512258.

[RK05] R. Renner and R. König. Universally composable privacy amplification against quantum adversaries. In *Second Theory of Cryptography Conference TCC*, volume 3378 of *Lecture Notes in Computer Science*, pages 407–425. Springer, 2005. arXiv:quant-ph/0403133.

[Sha02] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002. http://dblp.uni-trier.de.

[Vad03] S. Vadhan. On constructing locally computable extractors and cryptosystems in the bounded storage model. In *Advances in Cryptology — CRYPTO 2003*, pages 61–77, 2003.

[Win99] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45, 1999.

[Wul07] J. Wullschleger. Oblivious-transfer amplification. In *EUROCRYPT 2007*, pages 555–572, 2007.