One-Shot Non-Catalytic Distributed Purity Distillation*

Sayantan Chakraborty[1] Rahul Jain[1,3,4] Pranab Sen[1,2]

[1] Centre for Quantum Technologies, National University of Singapore
[2] Tata Institute of Fundamental Research, Mumbai
[3]Department of Computer Science, National University of Singapore
[4] MajuLab, UMI 3654, Singapore

Abstract

Pure states are an important resource in many quantum information processing protocols. However, even making a fixed pure state, say $|0\rangle$, in the laboratory requires a considerable amount of effort. Often one ends up with a mixed state ρ whose classical description is nevertheless known. Hence it is important to develop protocols that extract a fixed pure state from a known mixed state. In this work, we study the problem of extracting a fixed pure state $|0\rangle^{A'}|0\rangle^{B'}$ from a known pure state ρ^{AB} distributed between two parties A and B. Here, A', B' are subspaces of A, B and the total amount of purity extracted is $\log|A'| + \log|B'|$. The parties can borrow local pure ancilla, apply local unitary operations and send a message from A to B through a dephasing channel. If local pure ancilla is borrowed, it must be subtracted in order to properly account for the purity extracted. We obtain the most efficient achievable bounds on one shot distributed purity extraction, in terms of the rate of local ancilla borrowed by the protocol, while distilling pure qubits at the best known rate. Our protocols borrow little to no local pure ancilla. Our bounds improve upon the existing bounds for this problem in both one shot as well as asymptotic iid settings. In particular they subsume all the asymptotic iid results of Devetak and Krovi-Devetak. In addition, we derive upper bounds for the rate of distillation in the one shot setting, which nearly match our achievable bounds.

1 Introduction

Pure states are an important and ubiquitous resource in most quantum information processing protocols. Often, while implementing a quantum algorithm, one assumes the availability of pure states in the form of ancilla qubits which can be used as workspace for some computational operation. A specific example of this is the implementation of isometric operators as quantum gates in a circuit. Due to their widespread use, pure states are often assumed to be a freely available resource in most quantum information processing protocols. However, the question remains as to the cost one has to incur to prepare such pure states in the lab. Indeed, that this is a nontrivial operation was realised by Landauer [14], who showed that to initialise an arbitrary classical bit to some preset value, an operation called erasure, one has to do work. Along a similar vein, the works of Bennett et al. and Szilard [1, 18] prove that one can extract work from a thermal bath if the system is initialised to a pure state.

The above works underscore the importance of characterising the resources that are necessary to produce pure states in the lab. To that end, we consider the problem of *purity distillation* and give an informal introduction below.

An Informal Description of Purity Distillation

The problem of purity distillation is concerned with characterising the rate at which pure qubit states can be obtained from a given known input state, using certain admissible quantum operations. To that end, we first require a *measure of purity* of a given known state, and a list of allowable operations under which our chosen measure of purity does

*A preliminary version of this work appeared at the 2023 59th Annual Allerton Conference on Communication, Control, and Computing [4]

not increase. For a given state ρ^A, a good choice for the measure of purity found in the literature [15, 8, 17] is the following:

$$\log|A| - \widetilde{H}(A),$$

where $\widetilde{H}(\cdot)$ is a placeholder for a well defined notion of entropy of a state which is suitable for our purposes. A list of allowable operations then takes the following form:

1. Tracing out a subsystem.
2. Appending a maximally mixed state in some register A_{mix}, i.e. we allow access to private randomness.
3. A special class of quantum operations $\{\mathcal{N}\}$.

The special class of quantum operations depends on the generality in which one wishes to treat the theory of purity distillation, and different authors have used increasing larger sets of operations in their treatment of the topic (see [17]). What is important is that the measure of purity should be non-increasing under these sets of operations.

Looking ahead, one can add a further allowable operation to the above list, that of borrowing pure ancilla qubits locally in some register C_{pure}. However, since this clearly increases the measure of purity, one must account for this by modifying the formula of the measure of purity by the term $-\log\left|C_{\text{pure}}\right|$. We refer to this as operation as 'borrowing pure ancilla qubits in a catalytic manner'.

Given the above setting, we can now define the tasks of *local* and *distributed* purity distillation:

Local Purity Distillation : Given a quantum state ρ^A, a party Alice can use any finite sequence of operations from the list of allowed operations to produce a state σ^{A_p}, such that:

$$\left\|\sigma^{A_p} - |0\rangle\langle 0|^{A_p}\right\|_1 \leq \varepsilon,$$

for some error parameter ε. The goal is to maximise $\log|A_p|$.

Distributed Purity Distillation : Given a quantum state ρ^{AB}, where the party Alice has access to the system A and the party Bob has access to the system B, and a completely dephasing channel $\mathcal{P}^{X_A\to X_B}$ from Alice to Bob, the parties are can use any finite sequence of local allowable operations, together with one classical message from Alice to Bob via the completely dephasing channel, to produce a joint state $\sigma^{A_pB_p}$, such that:

$$\left\|\sigma^{A_pB_p} - |0\rangle\langle 0|^{A_p} \otimes |0\rangle\langle 0|^{B_p}\right\|_1 \leq \varepsilon,$$

for some error parameter ε. The goal is to maximise $\log|A_p| + \log|B_p|$.

A further generalisation of the task of distributed purity distillation (DPD) is when we also charge for the amount of classical communication from Alice to Bob. To that end, we consider (informally) the following task:

DPD with Bounded Classical Communication : Given the setting of DPD, the goal is to maximise the quantity $\log|A_p| + \log|B_p|$ with the additional constraint that the number of classical bits that Alice is allowed to send to Bob is at most $C_{\text{classical}}$.

We will now make some remarks regarding the choice of special quantum operations in the set of allowable operations. From an informal perspective, it would seem logical that one should only allow local unitary operations in the set of special quantum operations $\{\mathcal{N}\}$. This is because any other quantum operation would require additional ancilla qubits to be implemented in the lab using quantum circuits. Nevertheless, one may ask whether the set of special quantum operations may be enlarged from the set of unitary operators, to include maps which do not increase the measure of purity of the input state. Note that if one is allowed to include any such non-unitary map in the set $\{\mathcal{N}\}$, its use will be *free*, in the sense we will not charge for the number of ancilla qubits required to implement this map as a quantum circuit. Indeed, this topic has been studied in the works [8, 17], where the authors show that their choice of purity measure does not increase under the action of unital CPTP maps.

However, the choice of $\{\mathcal{N}\}$ in this paper is more restrictive and informed by a distinction between the tasks of local and distributed purity distillation. To see this distinction, note that one must allow some communication channel from Alice to Bob in the distributed setting. If not, then the best the parties can do is two locally optimal protocols on their systems A and B. The quantum operator used as the communication channel must be a member of the set $\{\mathcal{N}\}$. However, the choice of this channel cannot be an arbitrary operator from $\{\mathcal{N}\}$. For example, if one allows the identity superoperator on the system A, which is a unital CPTP map, to be used as a channel, Alice can then send her entire system A to Bob This trivially reduces the distributed distillation problem to the local distillation problem. Thus, we must *fix* a choice of channel in the distributed case.

A judicious choice of channel is a *classical* communication channel from Alice to Bob, modelled by the completely dephasing map $\mathcal{P}^{X_A \to X_B}$. Informally, we do not allow Alice to send any entangled bits to Bob, but allow classical communication. Note that this map is also a unital CPTP. This choice can be justified from a practical perspective as well, given that robust quantum channels across large distances have not yet been realised.

Given the above discussion, throughout this paper we will fix our set of special operations $\{\mathcal{N}\}$ to include only unitary operators and completely dephasing maps. In addition, as pointed out above, we allow partial trace, completely mixed ancilla and borrowing local pure ancilla catalytically. We will show that our choice of purity measure is non-increasing under the set of allowable operations which we choose to work with. In fact the same set of operations were allowed by the earlier works of [6, 13]. The reader is referred to Sections 3 and 4 for the rigorous definitions and lemmas pertaining to the discussion above.

Remark 1.1. We remark that our choice of purity measure is non-increasing even under the action of unital CPTP maps. This is easily seen from the proofs presented in Section 4. However, we do not comment on this further to focus on the main contribution of this paper.

History and Previous Works

The problems of local and distributed purity distillation first appeared in the works [15, 10, 11]. Specifically, the distributed distillation problem was first introduced in the asymptotic iid setting [15], and some preliminary bounds for the case when both 1-way and 2-way communication is allowed between Alice and Bob was given in [10] in the CLOCC (closed local operations and classical communication) setting in the asymptotic iid regime. In this setting the parties are not allowed to borrow any ancilla qubits catalytically, nor do they have access to private randomness. This implies that for an input state ρ^A, the set of operations $\{\mathcal{N}\}$ are allowed to be *only* unitary operators on A along with completely dephasing maps.

A further generalisation of this setting where the parties have access to private randomness, abbreviated as NLOCC (noisy local operations and classical communication) was also considered in [10]. In this case one allows local unitaries to act on both the input register as well as the system which holds the completely mixed state. The set of quantum operations $\{\mathcal{N}\}$ is clearly larger in this case than CLOCC, since one can construct operations on the input register which are convex combinations of unitary operators. However, this model still does not allow the parties to borrow ancilla qubits catalytically. A tight lower bound for the local distillation problem was provided in [12], in the asymptotic iid setting.

Aside from the preliminary works mentioned above, the first detailed treatment for the purity distillation problem appeared in the work of Devetak [6]. Devetak was the first to introduce the idea of borrowing pure ancilla in a catalytic manner, formalised as the CLOCC$'$ paradigm. In this paradigm one is allowed to borrow pure local ancilla qubits, but has to discount them from the final expression for number of pure qubits distilled. This relaxation allowed Devetak to characterise the rate of distributed purity distillation, when unbounded one-way classical communication is allowed. In particular, Devetak showed that, given n iid copies of a bipartite state ρ^{AB}, where A and B are shared between two parties, and unbounded one-way classical communication, it is possible to distil pure qubits at (roughly) a rate:

$$\log|A| - H(A) + \log|B| - H(B) + \frac{1}{n}\max_{\Lambda_n} I(X^n : B^n) \tag{1}$$

for a large enough n, where Λ_n is a rank-1 POVM that acts on the system A^n to produce a classical register X_n. It was also shown in the same paper that in the case of unbounded classical communication, this bound is tight in the iid limit.

The reader may have guessed that the additive mutual information term appearing in the expression above contributes a surplus of pure distillable qubits, more so than what a naïve application of two local protocols on the A and B systems would have allowed. As we shall see shortly, these surplus pure qubits are distilled by using the *classical-quantum* correlations between the two systems A and B (see [7] for more details). These correlations are extracted during the protocol execution by using the POVM Λ_n. Since the mutual information quantity above is maximised by rank-1 POVM, Devetak only considers these and indeed his protocol and proof techniques are heavily reliant on this fact.

The problem of DPD with bounded classical communication in the asymptotic iid setting was first considered by Krovi and Devetak in [13], where the authors not only provided tight upper and lower bounds for this problem but also significantly simplified the original proof given in [6]. In fact, the authors of that paper showed that under the constraint that Alice is allowed to send at most $nC_{\text{classical}}$ number of bits to Bob, it is possible to recover pure states from ρ^{AB} at a rate similar in expression to the formula in Equation 1, with the important distinction that maximisation is now over the set of all POVMs Λ_n such that $I(X_n : B^n R^n) \leq nC_{\text{classical}}$. Here the mutual information is computed with respect to the post measurement state $\rho^{X_n B^n R^n} := \left(I^{R^n B^n} \otimes \Lambda_n\right) \left((|\rho\rangle \langle\rho|^{ARB})^{\otimes n} \right)$ and $|\rho\rangle^{ABR}$ is some purification of the shared state ρ^{AB}. Note that in this case the POVM Λ_n will in general no longer be rank-1.

All the works mentioned above tackle the problem of purity distillation in the asymptotic iid setting, that is, when one assumes that many independent copies of the resources are available to the parties in the protocol. Recently, Chakraborty, Nema and Buscemi [3] presented one-shot versions of the local and distributed purity distillation protocols in [3], where the authors assumed that only *one* copy of the underlying state is available to the parties taking part in the protocol. Although the techniques presented in that paper generalise Devetak's [6] original techniques to the one-shot setting, it is not immediately clear how one can adapt them to the case of DPD with bounded communication. In particular it is not clear how one can extend the asymptotic iid results of Krovi Devetak on DPD with bounded communication to the one shot setting.

Our Contribution

All the protocols for distributed purity distillation mentioned so far work in the paradigm where one is allowed to borrow some ancilla qubits at the beginning of the protocol but must account for them in the final rate. In fact, all the existing protocols which achieve the best known rate for this problem, whether in the asymptotic iid setting ([6] and [13]) or the one shot setting ([3]) crucially require ancilla qubits which they use in this catalytic manner. Furthermore, the rate at which these protocols borrow ancilla is typically quite high, roughly $\frac{1}{n} I(X_n : R^n B^n)$ for the asymptotic iid protocols and $I^{\varepsilon}_{\max}(X : RB)$ for one shot protocols, where the mutual information quantities are always computed with respect to the post measurement state ρ^{XRB} obtained after the action of the POVM. This is clearly undesirable from a practical standpoint, since one would hope that protocols used to distil pure qubits would themselves require only a few initial pure qubits to function.

Note that there are ad-hoc techniques, called bootstrapping, to reduce the rate of pure qubits which the protocol consumes in the asymptotic iid setting. Indeed, given n iid copies of the underlying state, one can divide these states into blocks of size $\sqrt{n}$. One can then use some ancilla to run the Krovi-Devetak protocol on the first block, and recover this ancilla at the end of the protocol. The recovered ancilla can then be used catalytically on subsequent runs of the Krovi-Devetak protocol on the other $\sqrt{n}$ sized blocks. There may be other similar strategies which use the idea of dividing the iid states into smaller blocks to reduce the number of ancilla qubits that the parties have to borrow (see Devetak [6]) However, these strategies are ad-hoc and depend upon assumptions regarding the number of pure qubit states that each party can distil. Further, these techniques completely fail in the one-shot setting where only *one* copy of the underlying state is available, and one cannot do any bootstrapping.

In this paper we present a uniform approach towards distilling the maximum number of pure qubits in the distributed setting for a given amount of classical communication, while at the same time reducing the number of initial pure catalytic ancilla qubits borrowed, which works both in the one-shot and the asymptotic iid setting. We call the proposed protocol as `FewQubits` (see Theorem 7.1 in Section 7). In comparison with existing protocols, `FewQubits` has several key improvements with regard to the number of ancilla qubits it requires, while maintaining the same rate of distillation as existing protocols. To highlight these improvements, we present a comparison of `FewQubits` with

the currently existing protocols, both in the asymptotic iid and one shot settings. We show that FewQubits offers advantages over existing protocols in both paradigms:

1. In the asymptotic iid limit, the rate at which FewQubits requires input ancilla qubits to function is 0, independent of the input mixed state ρ^{AB} or the input POVM Λ_n.

2. In the one shot setting, when unbounded 1-way classical communication is allowed (i.e. the setting of Devetak's original paper), FewQubits requires at most $O(\log \frac{1}{\varepsilon})$ pure ancilla qubits in the worst case. In comparison, the one shot protocol of [3] requires roughly $I^{\varepsilon}_{\max}(X : RB)$ ancilla qubits to work under similar assumptions. See Corollary 8.2 in Section 8 for details.

3. To facilitate a fair comparison in the more general scenario when the rate of classical 1-way communication is bounded in the one shot setting, we first present an appropriate one shot generalisation of the original Krovi-Devetak protocol [13], which we call KD_OneShot(Section 6.4). We prove that under mild conditions, FewQubits requires provably fewer ancilla qubits to function than KD_OneShot. See Corollary 8.1 in Section 8 for details.

The main technical ingredient in the construction of FewQubits is an embedding technique which we use to simulate the action of the POVM Λ on the A space without requiring too many extra ancilla qubits. Note that since the parties are allowed *only* local unitary operations, any POVM must be implemented coherently. To implement the POVM Λ one would then require an extra register to store the classical outcomes, i.e., given any POVM Λ its coherent counterpart can be expressed as the isometry $\sum_x |x\rangle^X \sqrt{\Lambda_x}^A$, where the log dimension of the system X is precisely the number of intial pure ancilla qubits required.

Note that since Λ is an arbitrary POVM, one cannot hope to bound the number of possible outcomes that this POVM has. The first step therefore is to replace Λ with another POVM $\widetilde{\Lambda}$ which has far fewer outcomes (typically $2^{I^{\varepsilon}_{\max}(X:RB)}$ many) but which nevertheless preserves the correlations between the the classical output register and the system B. This step is made possible by the measurement compression theorem of Winter [23] and has been used by both Devetak [6] and Krovi and Devetak [13]. The problem is harder in the one shot setting but one can use a recent one shot measurement compression theorem of [5] to get around it. This theorem is the key to results presented in [3].

Note however, that even after bounding the number of outcomes of the POVM, one still needs to borrow some $I^{\varepsilon}_{\max}(X : RB)$ initial pure ancilla qubits to store the outcomes. Our main contribution goes towards reducing this rate as much as possible. The main idea is that we design a unitary operator $U_{\widetilde{\Lambda}}$ which simulates the action of measuring the A register coherently with $\widetilde{\Lambda}$ *in place*, i.e., in the A register itself, requiring very little, and in many cases zero, additional pure ancilla in order to store the measurement outcomes. The details of this are technical and can be found in Section 7.

Aside from our main result, we also present upper bounds on the rates of purity distillation that *any* local and distributed distillation algorithm can hope to achieve in the one shot setting. Prior to our work such bounds were not known in the one shot regime. These upper bounds nearly match the rate of distillation given by KD_OneShotand FewQubits .

Organisation of the paper

The paper is organised as follows: in Section 2 we present the definitions of the one-shot quantities that we use throughout the paper. We also state several known properties of these quantities as facts and prove some other relevant properties as lemmas in this section. In Section 3 we formally present the definitions of ε-purity and the tasks of local and distributed purity distillation. In Sections 4 and 5 we present upper bounds pertaining to the tasks of local and distributed purity distillation. We should mention that Section 4 also contains details regarding an optimal achievable protocol for local purity distillation. We would also like to highlight Section 5.1 in which we focus on upper bounds for distributed purity distillation in the case when unbounded classical communication is allowed. In Section 6 we derive lower bounds for distributed purity distillation in the case of bounded communication, with the KD_OneShotprotocol presented in Section 6.4. We present our main result, the existence of FewQubits , in Section 7. Section 8 contains a comparison between FewQubits and KD_OneShot.

2 Preliminaries: Relevant Quantities

2.1 Notation and Some Basics

Definition 2.1 (**Quantum State).** *A quantum state ρ on some register A is a positive semi-definite matrix with trace* 1.

Remark 2.2. The notation $\sigma \geq 0$, for a matrix σ is used to denote the fact that σ is positive semi-definite. More generally, $\rho \geq \sigma$ implies that the matrix $\rho - \sigma \geq 0$. This partial order of the positive semi-definite matrices is referred to as the Loewner order.

Definition 2.3 (**Fidelity and Generalised Fidelity).** *Given two quantum states ρ and σ, the fidelity between the two states is defined as:*

$$F(\rho, \sigma) \coloneqq \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1 .$$

The fidelity can be extended in a meaningful way to matrices which are sub-states, i.e. matrices ρ and σ such that $0 \leq \rho, \sigma, \leq I$, in the following way:

$$\overline{F}(\rho, \sigma) \coloneqq F(\rho, \sigma) + \sqrt{(1 - \mathrm{Tr}[\rho])(1 - \mathrm{Tr}[\sigma])}.$$

$\overline{F}(\cdot, \cdot)$ is referred to as the generalised fidelity.

Remark 2.4. The generalised fidelity was defined in [20].

Definition 2.5. $\cdot$ **Operation** *Given an operator $M^{A \to B}$ and the operator N^A, we define the $\cdot$ operation as follows:*

$$M \cdot N \coloneqq MNM^\dagger.$$

2.2 Definitions: One-Shot Entropic Quantities

In this section we introduce the one-shot entropic quantities which we will be using in the subsequent sections to describe our protocols.

Definition 2.6 (**Smoothed Support Max Entropy**). *Given a quantum state ρ^A, let us denote its eigenvalues by $\lambda_1, \ldots \lambda_{|\mathrm{supp}(\rho)|}$ (in ascending order) corresponding to the eigenvectors $v_1, \ldots, v_{|\mathrm{supp}(\rho)|}$. Let $\lambda_1, \ldots \lambda_k$ denote the smallest eigenvalues such that $\sum_i \lambda_i \leq \varepsilon$. We define the* **smoothed support max entropy** *of ρ^A as*

$$\widetilde{H}^{\varepsilon}_{\max}(A)_\rho \coloneqq \log\left(|\mathrm{supp}(\rho)| - k\right).$$

Definition 2.7 (**Smoothed Norm Max Entropy**). *Given the setup of in Definition 2.6, we define the* **smoothed norm max entropy** *of the state ρ^A as*

$$H'_{\max}{}^{\varepsilon}(A)_\rho \coloneqq \log \frac{1}{\lambda_{k+1}}$$

Definition 2.8 (**Conditional Smooth Hypothesis Testing Entropy**). *Given a quantum state ρ^{AB} we define the* **Smooth Hypothesis Testing Entropy** *as*

$$H^{\varepsilon}_H(A|B)_\rho \coloneqq -D^{\varepsilon}_H(\rho^{AB} \,||\, \mathbb{I}^A \otimes \rho^B)$$

where D^{ε}_H, the hypothesis testing relative entropy, is defined as:

$$2^{-D^{\varepsilon}_H(\rho||\sigma)} \coloneqq \min_{\substack{0 \leq \Pi \leq \mathbb{I} \\ \mathrm{Tr}[\Pi\rho] \geq 1-\varepsilon}} \mathrm{Tr}\left[\Pi\sigma\right].$$

For the couple of definitions that follow we will require the notion of an ε-ball around a state ρ. The following definition can be found in [20, Definition 10]:

Definition 2.9. *Given a quantum state ρ^A, we define the ε-ball $\mathcal{B}^\varepsilon(\rho)$ as:*

$$\mathcal{B}^\varepsilon(\rho) \coloneqq \{\tau \ : \ 0 \leq \tau, \mathrm{Tr}\left[\tau\right] \leq 1, P(\tau, \rho) \leq \varepsilon\},$$

where $P(\cdot,\cdot)$ is the purified distance on the space of sub-normalised states (see [20] for details).

Definition 2.10. (Conditional Smooth Max Entropy) *Given a bipartite quantum state ρ^{AB}, we define the* **smooth max entropy** *as*

$$H_{\max}^{\varepsilon}(A|B)_\rho \coloneqq \min_{\rho' \in \mathcal{B}^\varepsilon(\rho)} \max_{\substack{\sigma^B \geq 0 \\ \mathrm{Tr}[\sigma]=1}} 2\log F\left(\rho'^{AB}, \mathbb{1}^A \otimes \sigma^B\right)$$

Definition 2.11. (Conditional Smooth Min Entropy) *Given a bipartite quantum state ρ^{AB}, the* **conditional smooth min entropy** *is defined as:*

$$H_{\min}^{\varepsilon}(A|B)_\rho \coloneqq \max_{\rho' \in \mathcal{B}^\varepsilon(\rho)} -\log \min\left\{\mathrm{Tr}\left[\sigma^B\right] \ | \ \sigma^B \geq 0, \rho'^{AB} \leq \mathbb{I}^A \otimes \sigma^B\right\}.$$

2.3 Properties: One-Shot Entropic Quantities

Fact 2.12. (Data Processing Inequality for the Smooth Min and Max entropies, [20]) *Given a state ρ^{AB} and $\varepsilon > 0$, a CPTP map $\mathcal{E}^{B\to D}$, we define $\sigma^{AD} \coloneqq (\mathbb{I}^A \otimes \mathcal{E}^B)(\rho^{AB})$. Then it holds that:*

$$H_{\min}^{\varepsilon}(A|B)_\rho \leq H_{\min}^{\varepsilon}(A|D)_\sigma$$
$$H_{\max}^{\varepsilon}(A|B)_\rho \leq H_{\max}^{\varepsilon}(A|D)_\sigma.$$

We will also require the following data processing type inequality, presented in [19]. We present a simplified version of the original result, which is much more pertinent for our purposes:

Fact 2.13. ([19] *Given a state ρ^{AB} and $\varepsilon > 0$, a unital CPTP map $\mathcal{E}^{A\to C}$, we define $\sigma^{CB} \coloneqq (\mathcal{E}^A \otimes \mathbb{I}^B)(\rho^{AB})$. Then it holds that:*

$$H_{\min}^{\varepsilon}(A|B)_\rho \leq H_{\min}^{\varepsilon}(C|B)_\sigma$$
$$H_{\max}^{\varepsilon}(A|B)_\rho \leq H_{\max}^{\varepsilon}(C|B)_\sigma.$$

We refer below to a subset of the chain rules for the smooth min and max entropies, presented in [21], that will prove useful for our purposes. We present the chain rules in a simplified form which is most pertinent for us.

Fact 2.14. (Chain Rules for Smooth Min and Max Entropy, [21]) *Given a state ρ^{AB}, it holds that:*

$$H_{\max}^{\varepsilon}(AB)_\rho \geq H_{\min}^{O(\varepsilon)}(A|B)_\rho + H_{\max}^{O(\varepsilon)}(B)_\rho - O(\log\frac{1}{\varepsilon})$$
$$H_{\max}^{\varepsilon}(AB)_\rho \geq H_{\max}^{O(\varepsilon)}(A|B)_\rho + H_{\min}^{O(\varepsilon)}(B)_\rho - O(\log\frac{1}{\varepsilon}).$$

Fact 2.15 ([3])**.** *For any quantum state ρ^A it holds that*

$$H_{\max}^{\varepsilon}(A)_\rho \leq \widetilde{H}_{\max}^{\varepsilon}(A)_\rho \leq {H'_{\max}}^{\varepsilon}(A)_\rho \leq \log\frac{|A|}{\varepsilon}$$

Lemma 2.16. *Given a state ρ^A and an arbitrary purification $|\rho\rangle^{RA}$, it holds that*

$$H_H^{\varepsilon}(A)_\rho = H_H^{\varepsilon}(R)_\rho.$$

Proof. We will first show that we can assume that the optimising operator in the definition of $H_H^\varepsilon(A)_\rho$ commutes with ρ. Let this operator be Π. To begin, consider the Schmidt decomposition of $|\rho\rangle^{AR}$:

$$|\rho^{AR}\rangle = \sum_a \sqrt{P_A(a)}\, |a\rangle^A |\zeta_a\rangle^R,$$

which implies that

$$\rho^A = \sum_a P_A(a)\, |a\rangle \langle a|^A.$$

Then,

$$\mathrm{Tr}\left[\Pi\rho\right] = \sum_a P_A(a) \langle a|\Pi|a\rangle$$
$$\mathrm{Tr}\left[\Pi\right] = \sum_a \langle a|\Pi|a\rangle$$

Without loss of generality we can assume that Π has non-negative eigenvalues only on a subspace of the support of ρ. Now, consider the operator:

$$\widetilde{\Pi} \coloneqq \sum_{|a\rangle\langle a| \in \mathrm{supp}(\rho)} \langle a|\Pi|a\rangle\, |a\rangle \langle a|$$

It is easy to see that $\widetilde{\Pi}$ has all the properties of Π that we require. Thus we can assume that the optimising operator commutes with ρ. Next, we wish to compute the quantity $H_H^\varepsilon(R)_\rho$. As before, we can assume that the optimising operator commutes with ρ^R, i.e., it diagonalises in the basis $\left\{|\zeta_a\rangle^A\right\}$ and has non-negative eigenvalues only on a subspace of the support of ρ^R. This implies that, the optimising operator , say Σ, can be written as:

$$\Sigma^R = \sum_{\zeta_a \in \mathrm{supp}(\rho^R)} \lambda_a\, |\zeta_a\rangle \langle \zeta_a|^R$$

Finally, we see that the definition of $H_H^\varepsilon(R)_\rho$ reduces to solving the following LP:

$$\begin{aligned} \min & \sum_a \lambda_a \\ & \sum_a P_A(a)\lambda_a \geq 1-\varepsilon \\ & 0 \leq \lambda_a \leq 1 \end{aligned}$$

It is not hard to see that this same LP that defines $H_H^\varepsilon(A)_\rho$, if only we replace λ_a with $\mu_a \coloneqq \langle a|\Pi|a\rangle$. Thus, it holds that

$$H_H^\varepsilon(R)_\rho = H_H^\varepsilon(A)_\rho.$$

This concludes the proof. □

Lemma 2.17. *For a state ρ^A and any pure state $|\phi\rangle^B$, it holds that:*

$$H_H^\varepsilon(AB)_{\rho\otimes\phi} = H_H^\varepsilon(A)_\rho.$$

Proof. Note that the following holds since ϕ^B is pure:

$$\rho \otimes |\phi\rangle \langle\phi| = \sum_a P_A(a)\, |a\rangle \langle a|^A \otimes |\phi\rangle \langle\phi|^B.$$

Therefore, from the proof of Lemma 2.16 we can see that $H_H^\varepsilon(AB)_{\rho\otimes\phi}$ is given by the log of the solution of the LP:

$$\begin{aligned} \min & \sum_a \lambda(a) \\ & \sum_a P_A(a)\lambda(a) \geq 1-\varepsilon. \end{aligned}$$

However, this is the same LP that gives the expression for $H_H^\varepsilon(A)_\rho$. This concludes the proof. □

Lemma 2.18 (Equivalence of the Smoothed Norm Max and Smooth Hypothesis Testing Entropies). *For any quantum state ρ^A it holds that*

$$\widetilde{H}^{\varepsilon}_{\max}(A)_\rho - 1 \leq H^{\varepsilon}_H(A)_\rho \leq \widetilde{H}^{\varepsilon}_{\max}(A)_\rho$$

Proof. To prove this lemma, we first observe that without loss of generality we can assume that the optimising operator for $H^{\varepsilon}_H(A)_\rho$ diagonalises in the same basis as ρ^A. To see this, we argue via contradiction. Suppose the assumption isn't true. Let $\rho^A = \sum_a p(a) \left|v_a\right\rangle \left\langle v_a\right|^A$. Then, by definition:

$$\begin{aligned} &\operatorname{Tr}\left[\Pi^A \rho\right] \\ =&\sum_a p(a) \left\langle v_a|\Pi|v_a\right\rangle \\ \geq& 1-\varepsilon. \end{aligned}$$

Since $0 \leq \Pi \leq \mathbb{I}$, it holds that for all a, $0 \leq \left\langle v_a|\Pi|v_a\right\rangle \leq 1$. We can then define a new operator Π_{OPT} whose eigenbasis contains the vectors $\{|v_a\rangle\}$ (it can have more eigenvectors since the rank of Π may be larger than the rank of ρ), and which has the same eigenvalues as Π. Clearly, Π_{OPT} satisfies the criteria that $\operatorname{Tr}\left[\Pi_{\text{OPT}}\rho\right] \geq 1-\varepsilon$, and also $\operatorname{Tr}\left[\Pi_{\text{OPT}}\right] = \operatorname{Tr}\left[\Pi\right]$. Therefore, we can always assume that the optimiser for $H^{\varepsilon}_H(A)_\rho$ commutes with ρ. This immediately implies the upper bound since we obtain $\widetilde{H}^{\varepsilon}_{\max}(A)_\rho$ by projecting onto all but those eigenvectors of ρ^A whose eigenvalues are the smallest and add up to at most ε.

Now, since we know that the optimising operator for $H^{\varepsilon}_H(A)_\rho$ diagonalises in the same basis as ρ^A, once can assume that the following holds for all such candidate operators Π^A:

$$\begin{aligned} \rho^A &= \sum_a P_A(a) \left|a\right\rangle \left\langle a\right|^A \\ \Pi^A &= \sum_a \lambda(a) \left|a\right\rangle \left\langle a\right|^A \end{aligned}$$

Then, it holds that the problem of finding $H^{\varepsilon}_H(A)_\rho$ can be reduced to solving the following LP:

$$\begin{aligned} \min &\sum_a \lambda(a) \\ &\sum_a P_A(a)\lambda(a) \geq 1-\varepsilon \end{aligned}$$

We know from [16] that the log of the solution to this LP is at least $\widetilde{H}^{\varepsilon}_{\max}(A)_\rho - 1$. The lower bound follows. This concludes the proof. □

Lemma 2.19. (Subadditivity of the Smooth Hypothesis Testing Entropy) *Given a bipartite quantum state ρ^{AB}, it holds that*

$$H^{3\sqrt{\varepsilon}}_H(AB)_\rho \leq H^{\varepsilon}_H(A)_\rho + H^{\varepsilon}_H(B)_\rho.$$

Proof. To see that this holds, let Π^A and Π^B be the optimising operators for $H^{\varepsilon}_H(A)_\rho$ and $H^{\varepsilon}_H(B)_\rho$ respectively. Then,

$$\begin{aligned} \operatorname{Tr}\left[\Pi^A \otimes \Pi^B \rho^{AB}\right] &= \operatorname{Tr}\left[\left(\sqrt{\Pi^A} \otimes \sqrt{\Pi^B}\right) \cdot \rho^{AB}\right] \\ &= \operatorname{Tr}\left[\left(\mathbb{I}^A \otimes \sqrt{\Pi}^B\right) \cdot \left(\sqrt{\Pi}^A \otimes \mathbb{I}^B \cdot \rho^{AB} - \rho^{AB}\right)\right] + \operatorname{Tr}\left[\left(\mathbb{I}^A \otimes \sqrt{\Pi}^B\right) \cdot \rho^{AB}\right] \\ &\geq 1 - \varepsilon - \left\| \sqrt{\Pi}^A \cdot \rho^{AB} - \rho^{AB} \right\|_1 \end{aligned}$$

We know that

$$\operatorname{Tr}\left[\Pi^A \otimes \mathbb{I}^B \rho^{AB}\right] \geq 1-\varepsilon$$

By the Gentle Measurement Lemma, we can then see that

$$\left\| \sqrt{\Pi}^A \cdot \rho^{AB} - \rho^{AB} \right\|_1 \leq 2\sqrt{\varepsilon}$$

Therefore we can conclude that,

$$\mathrm{Tr}\left[\Pi^A \otimes \Pi^B \rho^{AB}\right] \geq 1 - 3\sqrt{\varepsilon}$$

Therefore, $\Pi^A \otimes \Pi^B$ is a candidate optimiser for $H_H^{3\sqrt{\varepsilon}}(AB)$, which implies that

$$H_H^{3\sqrt{\varepsilon}}(AB)_\rho \leq H_H^\varepsilon(A)_\rho + H_H^\varepsilon(B)_\rho.$$

This concludes the proof. □

Lemma 2.20. *Given the quantum state $\rho^A \otimes \pi^B$, where π^B is the maximally mixed state on the system B, it holds that:*

$$H_H^\varepsilon(AB)_{\rho\otimes\pi} = H_H^\varepsilon(A)_\rho + \log|B|.$$

Proof. By the arguments used in the proof of Lemma 2.18, we know that $H_H^\varepsilon(AB)_{\rho\otimes\pi}$ is obtained by solving the following LP, which we call $LP1$:

$$\begin{aligned} &\min \sum_{a,b} \lambda(a,b) \\ &\sum_{a,b} \frac{P_A(a)}{|B|} \lambda(a,b) \geq 1-\varepsilon, \end{aligned}$$

where $\rho^A = \sum_a P_A(a) |a\rangle \langle a|^A$ and $\pi^B = \sum_b \frac{1}{|B|} |b\rangle \langle b|^B$. Consider also the following LP, which we call $LP2$:

$$\begin{aligned} &\min |B| \cdot \left(\sum_a \lambda(a) \right) \\ &\sum_a P_A(a)\lambda(a) \geq 1-\varepsilon. \end{aligned}$$

Now note that an optimising $\{\lambda(a)\}$ for $LP2$ can be turned into a *feasible* $\{\lambda(a,b)\}$ for $LP1$ by simply declaring $\lambda(a,b) = \lambda(a), \forall b$. Similarly, an optimising $\{\lambda(a,b)\}$ for $LP1$ can be turned into a *feasible* $\{\lambda(a)\}$ for $LP2$ by defining:

$$\lambda(a) \coloneqq \sum_b \frac{1}{|B|} \lambda(a,b).$$

This argument immediately implies that the minima of $LP1$ and $LP2$ are equal. However, note that the minima of $LP2$ is precisely:

$$|B| \cdot 2^{H_H^\varepsilon(A)_\rho}.$$

Collating all the arguments above implies that:

$$H_H^\varepsilon(AB)_{\rho\otimes\pi} = H_H^\varepsilon(A)_\rho + \log|B|.$$

This concludes the proof. □

Corollary 2.21. *Given a state ρ^{AB}, it holds that :*

$$H_H^\varepsilon(AB)_\rho \leq H_H^\varepsilon(A)_\rho + \log|B|.$$

Proof. From the theory of unitary 1-designs we know that there exists a set of unitaries U_i^B on the system B and probability distribution $\{p_i\}$ such that, for any matrix M^B, it holds that:

$$\sum_i p_i U_i M U_i^\dagger = \mathrm{Tr}\left[M\right] \frac{I^B}{|B|}.$$

Let us denote the operation $\sum_i p_i U_i(\cdot)U_i^\dagger$ as $\mathcal{T}^{B\to B}$. Note that $\mathcal{T}$ is a unital CPTP map. Also, note that it is not hard to show that for any state ρ^{AB}, it holds that:

$$\left(\mathbb{I}^A \otimes \mathcal{T}^B\right)\left(\rho^{AB}\right) = \rho^A \otimes \pi^B,$$

where π^B is the maximally mixed state on the system B. Therefore, by Lemma cite, we see that:

$$\begin{aligned} H_H^\varepsilon(AB)_{\rho^{AB}} &\leq H_H^\varepsilon(AB)_{(\mathbb{I}^A\otimes\mathcal{T}^B)(\rho^{AB})} \\ &= H_H^\varepsilon(AB)_{\rho^A\otimes\pi^B} \\ &= H_H^\varepsilon(A)_\rho + \log|B|. \end{aligned}$$

This concludes the proof. □

Lemma 2.22. *Let σ^A be a state such that*

$$\left\|\sigma^A - |0\rangle\langle 0|^A\right\|_1 \leq \varepsilon$$

Then

$$H_H^\varepsilon(A)_\sigma \leq 0$$

Proof. The condition in the statement of the lemma implies that

$$\langle 0|\sigma|0\rangle \geq 1-\varepsilon$$

This implies that $|0\rangle\langle 0|^A$ is a valid candidate for the optimising operator for $H_H^\varepsilon(A)_\sigma$. Since $|0\rangle\langle 0|$ has trace 1, the result follows. This concludes the proof. □

Lemma 2.23. *Given a quantum cq state $\rho^{XB} = \sum_x P_X(x)|x\rangle\langle x|^X \otimes \rho_x^B$ where $x \in \mathcal{X}$, it holds that there exists a subset $\mathcal{S} \subseteq \mathcal{X}$ such that*

$$\begin{aligned} &\Pr_{P_X}\left[\mathcal{S}\right] \geq 1 - 2\sqrt{\varepsilon} \\ &H_H^{\sqrt{\varepsilon}}(\rho_x^B) \leq H_H^\varepsilon(B|X)_\rho - \log\varepsilon, \quad \forall x \in \mathcal{S}. \end{aligned}$$

Proof. Without loss of generality we can assume that the optimising operator Π^{XB} in the definition of $H_H^\varepsilon(B|X)_\rho$ is of the form:

$$\Pi^{XB} = \sum_x |x\rangle\langle x| \otimes \Pi_x^B.$$

By definition, this operator has the property that:

$$\sum_x P_X(x)\,\mathrm{Tr}\left[\Pi_x^B \rho_x^B\right] \geq 1-\varepsilon.$$

By Markov's inequality, we can then see that there exists a set $\mathcal{T}_1$ such that $\Pr_{P_X}\left[\mathcal{S}\right] \geq 1-\sqrt{\varepsilon}$ and for all $x \in \mathcal{T}_1$,

$$\mathrm{Tr}\left[\Pi_x^B \rho_x^B\right] \geq 1-\sqrt{\varepsilon}.$$

Again, by definition, it holds that:

$$2^{H_H^\varepsilon(B|X)_\rho} = \sum_x P_X(x) \operatorname{Tr}\left[\Pi_x^B\right].$$

Again, Markov's inequality tells us that there exists a set $\mathcal{T}_2 \subseteq \mathcal{X}$ of probability (under P_X) of at least $1-\varepsilon$ such that for all $x \in \mathcal{T}_2$, it holds that:

$$\operatorname{Tr}\left[\Pi_x^B\right] \leq \frac{2^{H_H^\varepsilon(B|X)_\rho}}{\varepsilon}.$$

Therefore, for all $x \in \mathcal{T}_1 \bigcap \mathcal{T}_2$ (which has probability at least $1-2\sqrt{\varepsilon}$ under P_X), it holds that Π_x^B is a candidate for the optimiser in the definition of $H_H^{\sqrt{\varepsilon}}(\rho_x^B)$. Thus defining $\mathcal{S} \coloneqq \mathcal{T}_1 \bigcap \mathcal{T}_2$ we see that the result follows. This concludes the proof. □

Lemma 2.24. *Given a cq state*

$$\rho^{XB} = \sum_x P_X(x) \left|x\right\rangle\left\langle x\right|^X \otimes \left|v_x\right\rangle\left\langle v_x\right|^B,$$

it holds that $H_H^\varepsilon(B \mid X)_\rho \leq 0$.

Lemma 2.25. *Given a cq state of the form*

$$\rho^{XAB} = \sum_x P_X(x) \left|x\right\rangle\left\langle x\right|^X \otimes \left|v_x\right\rangle\left\langle v_x\right|^{AB},$$

it holds that $H_H^\varepsilon(B \mid X)_\rho = H_H^\varepsilon(A \mid X)_\rho$.

The proofs of Lemma 2.24 and 2.25 can be found in Appendix B. Finally, we will require a data processing inequality for H_H^ε:

Lemma 2.26. *Given a state* ρ^{AB}*, a CPTP map* $\mathcal{E}^{B\to D}$*, and a unital CPTP map* $\mathcal{F}^{A\to C}$*, it holds that:*

$$H_H^\varepsilon(A|B)_\rho \leq H_H^\varepsilon(A|D)_{(\mathbb{I}^A\otimes\mathcal{E}^B)(\rho^{AB})},$$
$$H_H^\varepsilon(A|B)_\rho \leq H_H^\varepsilon(C|B)_{(\mathcal{F}^A\otimes\mathbb{I}^B)(\rho^{AB})}.$$

Proof. The proof of the first inequality follows directly from the data-processing inequality for $D_H^\varepsilon(\cdot\|\cdot)$ and the definition of $H_H^\varepsilon(A|B)_\rho$. For the second inequality, note that by definition, we know that:

$$\exp\left(H_H^\varepsilon(C|B)_{(\mathcal{F}^A\otimes\mathbb{I}^B)(\rho)}\right) = \min_{\substack{\Pi^{CB}\,:\,0\leq\Pi^{CB}\leq I^{CB}\\ \operatorname{Tr}\left[\Pi\left(\mathcal{F}^A\otimes\mathbb{I}^B\right)(\rho^{AB})\right]\geq 1-\varepsilon}} \operatorname{Tr}\left[\Pi^{CB}\left(I^C\otimes\rho^B\right)\right].$$

Let Π^* be the optimising operator in the expression of $H_H^\varepsilon(C|B)_{(\mathcal{F}^A\otimes\mathbb{I}^B)(\rho)}$. We will show that the operator $(\mathcal{F}^\dagger)^{C\to A}\otimes\mathbb{I}^B(\Pi^*)$ is a candidate optimiser for $H_H^\varepsilon(A|B)_\rho$, where $\mathcal{F}^\dagger$ is the adjoint of $\mathcal{F}$. Firstly, note that since $\mathcal{F}$ is unital and completely positive, $\mathcal{F}^\dagger$ is trace preserving and completely positive, i.e., CPTP. Also, since $\mathcal{F}$ is trace preserving, $\mathcal{F}^\dagger$ is unital. Note that since CPTP maps preserve operator inequalities, it holds that:

$$\begin{aligned}
0 &\leq \left(\mathcal{F}^{\dagger C}\otimes\mathbb{I}^B\right)(\Pi^*)\\
&\leq \left(\mathcal{F}^{\dagger C}\otimes\mathbb{I}^B\right)(I^C\otimes I^B)\\
&\overset{(a)}{=} I^A\otimes I^B.
\end{aligned}$$

In equality (a) we have used the fact that $\mathcal{F}^\dagger$ is unital, which implies that $\mathcal{F}^\dagger(I^C) = I^A$. With these observations in hand, note that the following holds:

$$\begin{aligned}
\operatorname{Tr}\left[\Pi^*\left(\mathcal{F}^A\otimes\mathbb{I}^B\right)(\rho^{AB})\right] &= \left\langle\Pi^*, \left(\mathcal{F}^A\otimes\mathbb{I}^B\right)(\rho^{AB})\right\rangle\\
&= \left\langle\left(\mathcal{F}^{\dagger C}\otimes\mathbb{I}^B\right)(\Pi^*), \rho^{AB}\right\rangle\\
&= \operatorname{Tr}\left[\left(\mathcal{F}^{\dagger C}\otimes\mathbb{I}^B\right)(\Pi^*)\rho^{AB}\right]\\
&\geq 1-\varepsilon.
\end{aligned}$$

This implies that $\left(\mathcal{F}^{\dagger C} \otimes \mathbb{I}^B\right)(\Pi^*)$ is a candidate optimiser for $H_H^\varepsilon(A|B)_\rho$, which in turn implies that:

$$\begin{aligned}
\exp(H_H^\varepsilon(A|B)_\rho) &\leq \mathrm{Tr}\left[\left(\mathcal{F}^{\dagger C} \otimes \mathbb{I}^B\right)(\Pi^*)\left(I^A \otimes \rho^B\right)\right] \\
&= \left\langle \left(\mathcal{F}^{\dagger C} \otimes \mathbb{I}^B\right)(\Pi^*), I^A \otimes \rho^B \right\rangle \\
&= \left\langle \Pi^*, \left(\mathcal{F}^A \otimes \mathbb{I}^B\right)\left(I^A \otimes \rho^B\right)\right\rangle \\
&\overset{(b)}{=} \left\langle \Pi^*, I^C \otimes \rho^B \right\rangle \\
&= \mathrm{Tr}\left[\Pi^*\left(I^C \otimes \rho^B\right)\right] \\
&= \exp\left(H_H^\varepsilon(C|B)_{(\mathcal{F}^A \otimes \mathbb{I}^B)(\rho)}\right),
\end{aligned}$$

where in equality (b) we have used the fact that $\mathcal{F}$ is unital. □

3 Definitions: ε-Purity, Local and Distributed Purity Distillation

In this section we present the formal definitions of the ε-purity of a state and the tasks of local and distributed purity distillation.

3.1 ε-Purity and Allowable Local Operations

We will first define the ε-purity of a state:

Definition 3.1. **ε-Purity** *Given a state ρ^A, the ε-purity of ρ^A is defined to as the number* $\log|A| - H_H^\varepsilon(A)_\rho$.

As mentioned in the introduction, the notion of ε-purity puts a bound on the number of single qubit pure states $|0\rangle$ that may be extracted from a given state ρ. To make this connection precise, we have to list the kinds of local operations that a party is allowed to perform on ρ to extract pure states from it. It is crucial that these operations do not increase the ε-purity of the state. To that end, we consider below a list of allowed local operations. We later show in Lemma 4.1 that indeed the operations listed below cannot increase the ε-purity of a given state.

Definition 3.2. Allowable Local Operations *Given a state ρ^A, we allow the following operations to be performed on the system A:*

1. *Appending a register A_{mix} to the system A, where the state on A_{mix} is maximally mixed.*
2. *Unitary operations.*
3. *Local completely dephasing maps $\mathcal{P}$.*
4. *Tracing out a subsystem.*

Along with the above operations, we will also allow appending pure states $|0\rangle\langle 0|$ to the system A in a register C_{pure}. To account for this, we require that the formula for the ε-purity of the state on AC_{pure} be modified as follows:

$$\log\left|AC_{\text{pure}}\right| - H_H^\varepsilon(AC_{\text{pure}})_{\rho\otimes|0\rangle\langle 0|} - \log\left|C_{\text{pure}}\right|.$$

3.2 Local Purity Distillation

We will now give an operational interpretation to the ε-purity, by building protocols out of the allowable operations which extract pure states from the given input state. To do this, we first define the notion of a local purity distillation code:

Definition 3.3. (Local Purity Distillation Code) *Given a quantum state ρ^A in the register A, we define a ε local purity distillation code as a sequence of allowable operations which produce a state σ^{A_p}, with the property that:*

$$\left\| \sigma^{A_p} - |0\rangle\langle 0|^{A_p} \right\|_1 \leq \varepsilon.$$

The rate of the code is given by

$$R^{\varepsilon}_{\text{local}} := \log |A_p| - \log \left| C_{\text{pure}} \right|.$$

A rate R is said to be ε-achievable for local purity distillation with respect to the state ρ^A if there exists an ε purity distillation code such that

$$R^{\varepsilon}_{\text{local}} = R - O(\log \frac{1}{\varepsilon})$$

Definition 3.4. (ε-Local Distillable Purity) *Given a state ρ^A, the* ε-local distillable purity $\kappa_\varepsilon(\rho^A)$ *is defined as the supremum over all ε-achievable rates R for local purity distillation.*

3.3 Distributed Purity Distillation

As mentioned in the introduction, the main topic of this paper is the task of distributed purity distillation. In this task we envision two parties, Alice and Bob, each of whom possess a share of a bipartite quantum state ρ^{AB}. The goal is for them to coordinate and extract pure states from this shared state. Under the supposition that Alice and Bob are allowed only local allowable operations on their systems A and B, they can each perform an optimal local purity distribution protocol, and recover pure states roughly at the rate $\log |AB| - H^{\varepsilon^2}_H(A)_\rho - H^{\varepsilon^2}_H(B)_\rho$. However, note that in this setup since we did not allow any communication between Alice and Bob, this is the best that they can do. The question then is that whether given the ability to communicate, can they do better?

We must keep in mind that whatever communication channel we introduce must be implementable by composing some allowable operations. This naturally leads us to the following definition of a distributed purity distillation protocol:

Definition 3.5. (Distributed Purity Distillation (DPD)) *Given a bipartite quantum state ρ^{AB} to two parties Alice and Bob, where Alice has access to the register A and Bob has access to the register B. A distributed purity distillation protocol with error ε is then defined as a protocol consisting of :*

1. *Local allowable operations on the system A.*
2. *A completely dephasing channel $\mathcal{P}^{X_A \to X_B}$, where the system X_A is generated at Alice's end and X_B is a classical register belonging to Bob.*
3. *Local allowable operations on the system B.*

Suppose that the state generated at the end of the protocol is $\sigma^{A_p B_p}$, where the system A_p belongs to Alice and B_p belongs to Bob. We require that:

$$\left\| \sigma^{A_p B_p} - |0\rangle\langle 0|^{A_p} \otimes |0\rangle\langle 0|^{B_p} \right\|_1 \leq \varepsilon.$$

The rate of the protocol is defined as

$$R^{\varepsilon}_{\text{dist}} := \log |A_p| + \log |B_p| - \log |C|$$

where the system $C \cong C_{\text{alice}} \otimes C_{\text{bob}}$ accounts for the local pure ancilla qubits borrowed by both Alice and Bob in the registers C_{alice} and C_{bob} respectively.

In this paper, as in [13], we will be concerned with DPD protocols with bounded classical communication from Alice to Bob. To that end, we introduce the following definition:

Definition 3.6. (DPD with Bounded Classical Communication) *Given a bipartite state ρ^{AB}, we define a $(R^{\varepsilon}_{\text{dist}}, C^{\varepsilon}_{\text{com}}, \varepsilon)$ protocol as a distributed purity distillation protocol with error ε, as defined in Definition 3.5, where it holds that*

$$\log |X_B| \leq C^{\varepsilon}_{\text{com}}.$$

where X_B is the classical output register of the perfectly dephasing channel $\mathcal{P}^{X_A \to X_B}$.

A rate pair $(R_{\text{pure}}, C_{\text{classical}})$ is said to be ε-achievable for DPD with bounded classical communication if there exists a $(R^{\varepsilon}_{\text{dist}}, C^{\varepsilon}_{\text{com}}, \varepsilon)$ protocol such that:

$$R^{\varepsilon}_{\text{dist}} = R_{\text{pure}} - O(\log \frac{1}{\varepsilon})$$
$$C^{\varepsilon}_{\text{com}} \leq C_{\text{classical}} + O(\log \frac{1}{\varepsilon}).$$

Definition 3.7. (ε 1-way Distillable Purity) *Given a state ρ^{AB} and $C_{\text{classical}} \geq 0$, the ε 1-way distillable purity $\kappa^{\rightarrow}_{\varepsilon}(\rho^{AB}, C_{\text{classical}})$ is defined as the supremum of R_{pure} over all ε-achievable rates $(R_{\text{pure}}, C_{\text{classical}})$ for distributed purity distillation.*

Remark 3.8. We will use the notation $\kappa^{\rightarrow}_{\varepsilon}(\rho^{AB}, \infty)$ to indicate the 1-way distillable purity in the case when we allow unbounded but finite classical communication.

4 Optimal Protocols for Local Purity Distillation

In this section we will show that given a state ρ^A, any finite sequence of allowable operations cannot increase the ε-purity of this state. We will then give an operational interpretation of the ε-purity, by constructing a local purity distillation code which extracts pure states from the given state at a rate which is almost equal to the ε-purity. We will also show that the ε-purity is the best rate of pure state production which any local purity distillation code can hope to achieve.

Lemma 4.1. *The ε-purity of a state ρ^A is non-increasing under allowable local operations.*

Proof. Recall that, given a state ρ^A, the following local operations are allowed:

1. Introducing a maximally mixed state in a register A_{mix}.
2. Introducing pure states in a register C_{pure} which must be accounted for.
3. Unitary operations.
4. A local completely dephasing channel $\mathcal{P}$.
5. Discard (trace out) a subsystem.

We will show that each of the above operations do not increase the ε-purity of ρ^A i.e., $\log |A| - H^{\varepsilon}_H(A)_{\rho}$.

Appending $\pi^{A_{\text{mix}}}$:
The state under consideration is now $\rho^A \otimes \pi^{A_{\text{mix}}}$. Then, the following holds:

$$\begin{aligned} & \log |AA_{\text{mix}}| - H^{\varepsilon}_H(AA_{\text{mix}})_{\rho\otimes\pi} \\ \overset{(a)}{=} & \log |AA_{\text{mix}}| - H^{\varepsilon}_H(A)_{\rho} - \log |A_{\text{mix}}| \\ = & \log |A| - H^{\varepsilon}_H(A)_{\rho}, \end{aligned}$$

where equality (a) follows from Lemma 2.20.

Appending C_{pure}:
In this case, recall that Definition 3.2 requires the formula for the ε-purity to be adjusted with a correction term $-\log\left|C_{\text{pure}}\right|$. With this correction and the fact that $H_H^\varepsilon(AC_{\text{pure}})_{\rho\otimes|0\rangle\langle 0|} = H_H^\varepsilon(A)_\rho$ (see Lemma 2.17) it is trivial to see that the ε-purity does not change.

Unitary Operations:
In this case, suppose that a unitary operator U^A acts on ρ^A to give σ^A. The unitary invariance of $H_H^\varepsilon(\cdot)$ implies that $H_H^\varepsilon(A)_\rho = H_H^\varepsilon(A)_\sigma$. This directly implies that unitary operations keep the ε-purity invariant.

Completely Dephasing Maps:
In this case, suppose that the system A is comprised of the registers $A'X_1$, and there exists a completely dephasing map $\mathcal{P}^{X_1\to X_2}$, where $|X_1| = |X_2|$. Since there exists a natural isomorphism between A and $A'X_1$, we can write the following:

$$\begin{aligned} H_H^\varepsilon(A)_\rho &= H_H^\varepsilon(A'X_1)_\rho \\ &\overset{(b)}{\leq} H_H^\varepsilon(A'X_2)_{\mathbb{I}^{A'}\otimes\mathcal{P}^{X_1}(\rho)}, \end{aligned}$$

where we have used the fact that the map $\mathbb{I}^{A'}\otimes\mathcal{P}^{X_1}$ is a unital CPTP map and Lemma 2.26 in step (b). This directly implies that the ε-purity is non-increasing under these maps, since $\log|A| = \log|A'X_1|$.

Discarding Subsystems:
Again, suppose that A is comprised of the systems $A''G$, where the system G is to be discarded. Suppose that the state after discarding G, on the system A'' is $\sigma^{A''}$. Then, the following holds:

$$\begin{aligned} H_H^\varepsilon(A)_\rho &= H_H^\varepsilon(A''G)_\rho \\ &\overset{(c)}{\leq} H_H^\varepsilon(A'')_\sigma + \log|G|\,. \end{aligned}$$

Step (c) follows from Corollary 2.21. This implies that:

$$\begin{aligned} &\log|A| - H_H^\varepsilon(A)_\rho \\ =\ &\log\left|A''G\right| - H_H^\varepsilon(A''G)_\rho \\ \geq\ &\log\left|A''G\right| - H_H^\varepsilon(A'')_\sigma - \log|G| \\ =\ &\log\left|A''\right| - H_H^\varepsilon(A'')_\sigma. \end{aligned}$$

This concludes the proof. □

We will now provide an operational interpretation of the ε-purity, by exhibiting an ε purity distillation code which recovers pure states from the given input state at a rate almost equal to the ε-purity of the input state. Such a protocol was shown to exist in [3, Theorem 1.7]. Further using Fact 2.18 along with the results of [3], we get the following fact:

Fact 4.2. Lower Bound for $\kappa_\varepsilon(\rho^A)$ *Given a quantum state ρ^A, there exists an ε purity distillation code with rate*

$$R_{\text{local}}^\varepsilon = \log|A| - H_H^{\varepsilon^2/9}(A)_\rho + O(\log\varepsilon) - 1$$

This also implies that:

$$\kappa_\varepsilon(\rho^A) \geq \log|A| - H_H^{\varepsilon^2/9}(A)_\rho + O(\log\varepsilon) - 1.$$

In fact, the ε purity distillation code which achieves the above lower bound consists only of a unitary operator U^A acting on the system A, and does not require any other allowable operations.

The following theorem encapsulates out discussion so far and connects the local distillable purity with the ε-purity by showing that the latter is an upper bound for the former:

Theorem 4.3. *Given a quantum state ρ^A and $\varepsilon > 0$, the ε local distillable purity of the state $\kappa_\varepsilon(\rho^A)$ satisfies the following bounds:*

$$\log|A| - H_H^{\varepsilon^2/9}(A)_\rho + O(\log\varepsilon) - 1 \leq \kappa_\varepsilon(\rho^A) \leq \log|A| - H_H^\varepsilon(A)_\rho$$

Proof. The lower bound follows directly from Fact 4.2. To get the upper bound, note that any ε purity distillation code is a sequence of allowable operations, which finally output a state σ^{A_p}, such that:

$$\left\| \sigma^{A_p} - |0\rangle\langle 0|^{A_p} \right\|_1 \leq \varepsilon.$$

Using the allowed operations listed above, we will now characterise the form of any finite sequence of operations. To do this, we adopt the notation that any subsystem with the name G_i (for some $i \in \mathbb{N}$) will be discarded at the end of the protocol. We also denote the final output state as σ^{A_pG}, where A_p is to be retained and G discarded. Note that without loss of generality we can assume that any registers which contain maximally mixed states or pure states can be introduced at the very beginning of the protocol, and any systems that are to be traced out can be discarded at the very end. Then, any general local purity distillation protocol takes the form in Table 1.

Alice
State ρ^A
Append $\pi^{A_{\text{mix}}} \otimes |0\rangle\langle 0|^{C_{\text{pure}}}$.
State $\sigma^{AA_{\text{mix}}C_{\text{pure}}} := \rho^A \otimes \pi^{A_{\text{mix}}} \otimes |0\rangle\langle 0|^{C_{\text{pure}}}$
Unitary $U_1 : AA_{\text{mix}}C_{\text{pure}} \to A_1X_1G_1$ Channel $\mathcal{P} : X_1 \to X_2G_2$ Unitary $U_2 : A_1X_2 \to A_2X_3G_3$ ...
State σ^{A_pG}
Discard the subsystem G
State σ^{A_p}

Table 1: **General Schema of a Local Protocol**

Then, one can use Lemma 4.1 at every step of the protocol iteratively, to see that:

$$\log|A| - H_H^\varepsilon(A)_\rho \geq \log|A_p| - H_H^\varepsilon(A_p)_\sigma - \log\left|C_{\text{pure}}\right|.$$

However, using the requirement that σ^{A_p} has to be close to the pure state $|0\rangle\langle 0|^{A_p}$ and invoking Lemma 2.22, we see that:

$$\log|A_p| - H_H^\varepsilon(A_p)_\sigma \geq \log|A_p|.$$

Collating these arguments, we see that:

$$\begin{aligned}\kappa_\varepsilon(\rho^A) &\leq \log|A_p| - \log\left|C_{\text{pure}}\right| \\ &\leq \log|A_p| - H_H^\varepsilon(A_p)_\sigma - \log\left|C_{\text{pure}}\right| \\ &\leq \log|A| - H_H^\varepsilon(A)_\rho.\end{aligned}$$

This concludes the proof. □

In the following section, we will thus refer to the *locally optimal* protocol, in reference to Theorem 4.3. Note that this locally optimal protocol consists only of a a unitary operator U^A acting on the system A, as given in Fact 4.2.

5 Distributed Protocols with Ancilla: Upper Bounds

In this section we prove a one-shot upper bound on the number of qubit states that Alice and Bob can hope to distil, given the setting of the distributed purity distillation problem with classical communication bounded by the rate $C_{\text{classical}}$. Throughout the rest of this section, to impose the bound on classical communication, we make the following assumption:

Assumption 5.1. The completely dephasing channel $\mathcal{P}^{X_A \to X_B}$ is such that

$$\log |X_B| \leq C_{\text{classical}},$$

where $C_{\text{classical}}$ is the maximum allowable rate of classical communication.

Notation

In our proofs of the upper bounds for distributed purity distillation, we will have to deal with several entropic quantities related to states which exist at different times during the protocol. For example, we may use a relation of the form $H_H^\varepsilon(A_p A_g X_A) \geq H_H^\varepsilon(A C_{\text{alice}} A_{\text{mix}})$. These two entropic quantities correspond to two different states, related by Alice's application of her local operations. In the interest of brevity, we will not explicitly spell out the state corresponding to which these registers are defined. However, in all cases the state and the point in the protocol when that state exists will be clear from the context provided by Table 2.

We will also require the following lemma:

Lemma 5.2. *Given a quantum state ρ^{AB} with the A register belonging to Alice and the B register belonging to Bob, any distributed purity distillation protocol making error at most ε can achieve a rate at most*

$$R_{\text{dist}}^\varepsilon \leq \log |A| + \log |B| - H_{\max}^{g(\varepsilon)}(A) - H_{\min}^{f(\varepsilon)}(B \mid X_B) + 2 \log \varepsilon,$$

where the entropic quantities are computed with respect to states as defined in Table 2.

Proof. Before we start the proof, we will first characterise what any general distributed purity distillation protocol looks like:

Alice	**Bob**
State ρ^{AB}	
Append state $\pi^{A_{\text{mix}}} \otimes \left\|0\right\rangle \left\langle 0\right\|^{C_{\text{alice}}}$	Append state $\pi^{B_{\text{mix}}} \otimes \left\|0\right\rangle \left\langle 0\right\|^{C_{\text{bob}}}$
Allowable local operations	
Create state $\sigma^{A_1 X_A}$	
State $\sigma^{A_1 X_A B}$	
$\xrightarrow{X_A \to X_B}$	
Allowable local operations on A_1	Allowable local operations $BB_{\text{mix}}C_{\text{bob}}X_B$.
State $\sigma^{A_p A_g B_p B_g}$	
Discard system A_g	Discard system B_g
Final state $\sigma^{A_p B_p}$	

Table 2: **General Schema of a Distributed Protocol**

Note that in the general protocol, although we can roll all of Bob's actions together, we must treat Alice's actions before and after she sends the classical messages to Bob separately. As in the proof of Theorem 4.3, we can assume without loss of generality that all systems that contain either maximally mixed states or pure states can be appended at the very beginning of the protocol and all systems to be traced out can be discarded at the end of the protocol. To that end, we make the convention that the actual state before both Alice and Bob discard some sub-systems is given by $\sigma^{A_p A_g B_p B_g}$, where A_g and B_g contain all systems that are to be discarded. Note that this means that the expression 'Allowable local operations' during the protocol execution refers only to some finite sequence of local unitary operators and local completely dephasing maps.

Before we move on with the main proof, we will state a useful claim:

Claim 5.3. *In reference to the protocol in Table 2, it holds, for any $\delta > 0$, that:*

$$\begin{aligned} H_H^\delta(B_p B_g) &\geq H_H^\delta(BB_{\text{mix}}C_{\text{bob}}X_B) \\ H_{\max}^\delta(A_p A_g) &\geq H_{\max}^\delta(A_1) \\ H_{\max}^\delta(A_1 X_A) &\geq H_{\max}^\delta(AA_{\text{mix}}C_{\text{alice}}). \end{aligned}$$

Proof. Note that in going from $\sigma^{BB_{\text{mix}}C_{\text{bob}}X_B}$ to $\sigma^{B_p B_g}$, Bob uses either local unitary operators or completely dephasing maps. Since $H_H^\delta(\cdot)$ for any state is non-decreasing under these operations, the claim first inequality follows. Similar observations holds for Alice's actions in going from σ^{A_1} to $\sigma^{A_p A_g}$, and from $\rho^{AA_{\text{mix}}C_{\text{alice}}}$ to $\sigma^{A_1 X_A}$. Since the smooth max entropy is invariant under the action of isometries and non-increasing under the action of unital CPTP maps (Fact 2.13), the other two inequalities follow. □

Next, note that:

$$\log|A_p B_p| - \log|C_{\text{alice}}C_{\text{bob}}| = \log|ABA_{\text{mix}}B_{\text{mix}}| - \log|A_g B_g|$$

We will lower bound $\log|A_gB_g|$ which will in turn allow us to upper bound $\log|A_pB_p| - \log|C_{\text{alice}}C_{\text{bob}}|$. Before we begin, we would like to point out that the systems X_A and X_B are isomorphic, however, they differ in the fact that the system X_B holds a classical state (diagonalisable with respect to the basis $\{|x\rangle\}$ of the completely dephasing channel) which is the output of the completely dephasing channel upon acting on the contents of the system X_A. Thus the state on the registers X_BBB_{mix} after Alice sends the contents of the register X_A through the channel is a cq state (with pure qubits $|0\rangle^{C_{\text{bob}}}$ in the register C_{bob} in tensor with the rest of the systems), while the state on the systems $A_pA_gX_A$ are *not* cq in general.

We will now lower bound $\log|A_gB_g|$:

$$\begin{aligned}\log|A_gB_g| &\geq H_H^{\varepsilon}(A_g) + H_H^{\varepsilon}(B_g) + 2\log\varepsilon \\ &\geq H_H^{3\sqrt{\varepsilon}}(A_pA_g) + H_H^{3\sqrt{\varepsilon}}(B_pB_g) + 2\log\varepsilon\end{aligned}$$

The above inequality uses the subadditivity of the smooth hypothesis testing entropy twice, along with the fact that both $H_H^{\varepsilon}(A_p)$ and $H_H^{\varepsilon}(B_p)$ are 0. Thus, LHS is

$$\begin{aligned}&\overset{(a)}{\geq} H_H^{3\sqrt{\varepsilon}}(A_pA_g) + H_H^{3\sqrt{\varepsilon}}(BX_BB_{\text{mix}}C_{\text{bob}}) + 2\log\varepsilon \\ &\overset{(b)}{=} H_H^{3\sqrt{\varepsilon}}(A_pA_g) + H_H^{3\sqrt{\varepsilon}}(BX_B) + \log|B_{\text{mix}}| + 2\log\varepsilon \\ &\overset{(c)}{\geq} H_{\max}^{3\sqrt{\varepsilon}}(A_pA_g) + H_{\max}^{3\sqrt{\varepsilon}}(X_B) + H_{\max}^{3\sqrt{\varepsilon}}(BX_B) - H_{\max}^{3\sqrt{\varepsilon}}(X_B) + \log|B_{\text{mix}}| + O(\log\varepsilon) \\ &\overset{(d)}{\geq} H_{\max}^{3\sqrt{\varepsilon}}(A_pA_g) + H_{\max}^{3\sqrt{\varepsilon}}(X_A) + H_{\min}^{f(\varepsilon)}(B \mid X_B) + \log|B_{\text{mix}}| + O(\log\varepsilon) \\ &\overset{(e)}{\geq} H_{\max}^{3\sqrt{\varepsilon}}(A_1) + H_{\max}^{3\sqrt{\varepsilon}}(X_A) + H_{\min}^{f(\varepsilon)}(B \mid X_B) + \log|B_{\text{mix}}| + O(\log\varepsilon)\end{aligned}$$

In inequality (a) we have used Claim 5.3. In equality (b) we used Lemma 2.20 and also the fact that the register C_{bob} contains a pure state in tensor with all the other systems. In inequality (c) we have used Fact 2.15 and Lemma 2.18 to lower bound both $H_H^{3\sqrt{\varepsilon}}$ terms by $H_{\max}^{3\sqrt{\varepsilon}}$, and we have absorbed the constant -1 arising from Lemma 2.18 into the $O(\log\varepsilon)$ term, assuming small enough ε. In inequality (d) above we have used the fact that the completely dephasing channel is a unital CPTP and the smooth max entropy cannot be decreased by the action of such a map [19]. We have also used the chain rules from Fact 2.14. Inequality (e) follows from Claim 5.3. Next, we will use the subadditivity of the max entropy to see that the LHS is:

$$\begin{aligned}&\geq H_{\max}^{h(\varepsilon)}(A_1X_A) + H_{\min}^{f(\varepsilon)}(B \mid X_B) + \log|B_{\text{mix}}| + O(\log\varepsilon) \\ &\overset{(f)}{\geq} H_{\max}^{h(\varepsilon)}(AC_{\text{alice}}A_{\text{mix}}) + H_{\min}^{f(\varepsilon)}(B \mid X_B) + \log|B_{\text{mix}}| + O(\log\varepsilon) \\ &\overset{(g)}{\geq} H_{\max}^{g(\varepsilon)}(A) + H_{\min}^{g(\varepsilon)}(C_{\text{alice}}A_{\text{mix}}|A) + H_{\min}^{f(\varepsilon)}(B \mid X_B) + \log|A_{\text{mix}}| + \log|B_{\text{mix}}| + O(\log\varepsilon) \\ &\overset{(h)}{=} H_{\max}^{g(\varepsilon)}(A) + H_{\min}^{f(\varepsilon)}(B \mid X_B) + \log|A_{\text{mix}}| + \log|B_{\text{mix}}| + O(\log\varepsilon).\end{aligned}$$

We have used Claim 5.3 in inequality (f). In inequality (g) we have used the chain rule from Fact 2.14. Finally, for equality (h), we use the following observation, which holds for any $\delta > 0$:

$$\begin{aligned}H_{\min}^{\delta}(C_{\text{alice}}A_{\text{mix}}|A) &\geq H_{\min}(C_{\text{alice}}A_{\text{mix}}|A) \\ &\geq \max_{\sigma^A} \sup\log\left\{\lambda \mid 2^{-\lambda}\mathbb{I}^{C_{\text{alice}}A_{\text{mix}}} \otimes \sigma^A \geq \pi^{A_{\text{mix}}} \otimes |0\rangle\langle 0|^{C_{\text{alice}}} \otimes \rho^A\right\} \\ &\geq \sup\log\left\{\lambda \mid 2^{-\lambda}\mathbb{I}^{C_{\text{alice}}A_{\text{mix}}} \otimes \rho^A \geq \pi^{A_{\text{mix}}} \otimes |0\rangle\langle 0|^{C_{\text{alice}}} \otimes \rho^A\right\} \\ &\geq \log|A_{\text{mix}}|.\end{aligned}$$

This shows that, for any distributed purity distillation protocol with error at most ε, it holds that

$$R_{\text{dist}}^{\varepsilon} \leq \log|A| + \log|B| - H_{\max}^{g(\varepsilon)}(A) - H_{\min}^{f(\varepsilon)}(B \mid X_B) + O(\log\varepsilon)$$

This concludes the proof. □

We are now ready to state and prove a theorem about the upper bound of the distributed purity of any quantum state:

Theorem 5.4. (Upper Bound for Distributed Purity of a State) *Given a quantum state ρ^{AB}, the 1-way distillable purity $\kappa_{\varepsilon}^{\rightarrow}(\rho^{AB}, C_{\text{classical}})$ is at most*

$$\kappa_{\varepsilon}^{\rightarrow}(\rho^{AB}, C_{\text{classical}}) \leq \log|A| + \log|B| - H_{\max}^{g(\varepsilon)}(A) - \min_{\Lambda \in \mathcal{S}} H_{\min}^{f(\varepsilon)}(B \mid X)_{\Lambda^A \otimes \mathbb{I}^B(\rho^{AB})} + O(\log \varepsilon)$$

where the set $\mathcal{S}$ is a subset of the set of all POVMs on the system A and is defined as follows:

$$\mathcal{S} \coloneqq \left\{ \Lambda^{A \to X} \mid I_{\max}^{\varepsilon}(X : RB)_{\mathbb{I}^{RB} \otimes \Lambda(|\rho\rangle\langle\rho|^{ABR})} + O(\log \varepsilon) \leq C_{\text{classical}} \right\}$$

where $|\rho\rangle^{ABR}$ is an arbitrary purification of ρ^{AB}.

Proof. From Lemma 5.2, we know that any distributed purity distillation protocol for ρ^{AB} and which makes an error at most ε, can extract a purity of at most

$$R_{\text{dist}}^{\varepsilon} \leq \log|A| + \log|B| - H_{\max}^{g(\varepsilon)}(A) - H_{\min}^{f(\varepsilon)}(B \mid X_B) + O(\log \varepsilon) \tag{2}$$

Recall that we obtained the system X_B by:

1. Using local allowed operations to obtain $\sigma^{A_1 X_A}$.
2. Sending X_A through the completely dephasing channel $\mathcal{P}^{X_A \to X_B}$.
3. Using local allowed operations on A_1 to obtain the systems $A_p A_g$.

Suppose that $V_1^{A A_{\text{mix}} C_{\text{pure}} \to A_1 X_A E_1}$ and $V_2^{A_1 \to A_p A_g E_2}$ are the Stinespring dilations of the maps that Alice enacts in Steps 1 and 3 above. Now consider the isometry:

$$\begin{aligned} V_3 : A A_{\text{mix}} C_{\text{pure}} &\to A_p A_g X_A E_1 E_2 \\ &\coloneqq \left(V_2 \otimes \mathbb{I}^{X_A}\right) \circ V_1. \end{aligned}$$

Define the unitary extension of V_3 in the usual way by appending an ancilla system W to the domain of W_3. We will call this unitary U_{ALICE}. Then note that if we measure the X_A system in the computation basis, the probability of getting the outcome x is given by the following expression:

$$\text{Tr}\left[\left(\mathbb{I}^{A_p A_g E_1 E_2} \otimes |x\rangle\langle x|^{X_A}\right) \cdot \left(U_{\text{ALICE}}(\rho^A \otimes \pi^{A_{\text{mix}}} \otimes |0\rangle\langle 0|^{C_{\text{alice}}} \otimes |0\rangle\langle 0|^W) U_{\text{ALICE}}^{\dagger}\right)\right]$$

By defining $\tau^{A_{\text{mix}} C_{\text{alice}} W} \coloneqq \pi^{A_{\text{mix}}} \otimes |0\rangle\langle 0|^{C_{\text{alice}}} \otimes |0\rangle\langle 0|^W$, and using the cyclicity of trace, we see that the above expression simplifies to:

$$\text{Tr}_A\left[\text{Tr}_{A_{\text{mix}} C_{\text{alice}} W}\left(\left(I^A \otimes \sqrt{\tau}^{A_{\text{mix}} C_{\text{alice}} W}\right) \cdot U_{\text{ALICE}}^{\dagger}\left(\mathbb{I}^{A_p A_g E_1 E_2} \otimes |x\rangle\langle x|^{X_A}\right) U_{\text{ALICE}}\right)(\rho^A)\right]$$

We define:

$$\Lambda_x^A \coloneqq \text{Tr}_{A_{\text{mix}} C_{\text{alice}} W}\left[\left(\left(I^A \otimes \sqrt{\tau}^{A_{\text{mix}} C_{\text{alice}} W}\right) \cdot U_{\text{ALICE}}^{\dagger}\left(\mathbb{I}^{A_p A_g E_1 E_2} \otimes |x\rangle\langle x|^{X_A}\right) U_{\text{ALICE}}\right)\right].$$

Clearly $\Lambda_x \geq 0$. Additionally, it is easy to see that:

$$\sum_x \Lambda_x^A = I^A.$$

Therefore, $\Lambda \coloneqq \left\{\Lambda_x^A\right\}$ is a POVM. Note that the following holds, with respect to the state $(\mathbb{I}^{RB} \otimes \Lambda)(|\rho\rangle\langle\rho|^{ABR})$:

$$\begin{aligned} C_{\text{classical}} &\geq \log |X_B| \\ &\geq H_{\max}^{O(\varepsilon^2)}(X_B)_{(\mathbb{I}^{RB}\otimes\Lambda)(|\rho\rangle\langle\rho|^{ABR}} + O(\log \varepsilon) \\ &\geq I_{\max}^{\varepsilon}(X_B : RB)_{(\mathbb{I}^{RB}\otimes\Lambda)(|\rho\rangle\langle\rho|^{ABR}} + O(\log \varepsilon). \end{aligned}$$

Thus, renaming X_B to X, it is clear that the supremum of the rate R_{pure} over all ε achievable rates pairs $(R_{\text{pure}}, C_{\text{classical}}, \varepsilon)$ is bounded above by the supremum of the upper bound on $R_{\text{dist}}^{\varepsilon}$ obtained in Equation 2, over the set $\mathcal{S}$ of all POVMS $\Lambda^{A\to X}$ such that $I_{\max}^{\varepsilon}(X : RB)_{\mathbb{I}^{RB}\otimes\Lambda(|\rho\rangle\langle\rho|^{ABR})} + O(\log\varepsilon) \leq C_{\text{classical}}$. This immediately implies that:

$$\kappa_{\varepsilon}^{\rightarrow}(\rho^{AB}, C_{\text{classical}}) \leq \log|A| + \log|B| - H_{\max}^{g(\varepsilon)}(A) - \inf_{\Lambda\in\mathcal{S}} H_{\min}^{f(\varepsilon)}(B \mid X)_{\Lambda^A\otimes\mathbb{I}^B(\rho^{AB})} + O(\log\varepsilon)$$

This concludes the proof. □

5.1 The Special Case of Unbounded Classical Communication

As mentioned earlier in the introduction, the original version of the distributed purity distillation problem was considered by Devetak in [6], in the regime when unbounded communication is allowed. In that spirit, we will show in this section that $\kappa_{\varepsilon}^{\rightarrow}(\rho^{AB}, \infty)$ can be bounded above by the expression in Theorem 5.4, with the important distinction that the infimum over all POVMs in the set $\mathcal{S}$ can be replaced by an infimum over all rank-1 POVMs.

Theorem 5.5. *Given a quantum state ρ^{AB}, the 1-way distillable purity in the case of unbounded communication, $\kappa_{\varepsilon}^{\rightarrow}(\rho^{AB}, \infty)$, can be bounded above by:*

$$\kappa_{\varepsilon}^{\rightarrow}(\rho^{AB}, \infty) \leq \log|A| + \log|B| - H_{\max}^{g(\varepsilon)}(A) - \inf_{\Lambda:\text{rank-1}} H_{\min}^{f(\varepsilon)}(B \mid X)_{\Lambda^A\otimes\mathbb{I}^B(\rho^{AB})} + O(\log\varepsilon).$$

Proof. The proof follows easily from Theorem 5.4, by noticing that:

$$\inf_{\Lambda\in\mathcal{S}} H_{\min}^{f(\varepsilon)}(B \mid X)_{\Lambda^A\otimes\mathbb{I}^B(\rho^{AB})} \geq \inf_{\Lambda:\text{rank-1}} H_{\min}^{f(\varepsilon)}(B \mid X)_{\Lambda^A\otimes\mathbb{I}^B(\rho^{AB})},$$

by the data processing inequality for the smooth min entropy. Specifically, for any POVM Λ, one can always create a new POVM Λ' by decomposing each POVM element in Λ into rank one operators $|\varphi\rangle\langle\varphi|$ such that $0 \leq \text{Tr}\left[|\varphi\rangle\langle\varphi|\right] \leq 1$, and then assigning a new label to the outcome corresponding to each of these operators. This concludes the proof. □

6 Distributed Protocols with Ancilla: Lower Bounds

In this section we will present a DPD protocol with bounded classical communication, which uses additional ancilla qubits in a catalytic manner, with an almost optimal rate of pure state distillation. We call this protocol `KD_OneShot`(see Section 6.4). This protocol can be viewed as a one-shot version of the protocol presented in [13].

The main theorem in this section quantifies the rate of pure state distillation for the protocol `KD_OneShot`. The theorem will take the following form: we will first fix the state ρ^{AB} and a rate of classical communication $C_{\text{classical}}$. We will then fix a POVM $\Lambda^{A\to X}$ such that $I_{\max}^{\varepsilon}(X : RB)_{\mathbb{I}^{RB}\otimes\Lambda(|\rho\rangle\langle\rho|^{ABR})} + O(\log\varepsilon) \leq C_{\text{classical}}$. Finally, we will show the existence of a protocol which distils pure states at the rate (roughly) $\log|A| - H_H^{\varepsilon}(A|X) + \log|B| - H_H^{\varepsilon}(B|X) - I_{\max}^{\varepsilon}(RB : X)$, with communication $I_{\max}^{\varepsilon}(RB : X) + O(\log\frac{1}{\varepsilon})$ (which is at most $C_{\text{classical}} + O(\log\frac{1}{\varepsilon})$). This protocol is `KD_OneShot`. The final lower bound on $\kappa_{\varepsilon}^{\rightarrow}(\rho^{AB}, C_{\text{classical}})$ is given by taking the supremum over all POVMs Λ such that $I_{\max}^{\varepsilon}(X : RB)_{\mathbb{I}^{RB}\otimes\Lambda(|\rho\rangle\langle\rho|^{ABR})} + O(\log\varepsilon) \leq C_{\text{classical}}$.

To be precise, we prove the following theorem:

Theorem 6.1. *Given the bipartite quantum state ρ^{AB}, it holds that*

$$\kappa^{\rightarrow}_{\varepsilon^{1/32}}(\rho^{AB}, C_{\text{classical}}) \geq \sup_{\Lambda \in \mathcal{S}} R_{\text{pure}} + O(\log \varepsilon),$$

where

$$R_{\text{pure}} = \log |A| - H^{\varepsilon}_{H}(A \mid X) + \log |B| - H^{\varepsilon}_{H}(B \mid X) - I^{\varepsilon^4}_{\max}(X : RB)$$

and the set $\mathcal{S}$ is as follows:

$$\mathcal{S} := \left\{ \Lambda^{A \to X} \mid I^{\varepsilon}_{\max}(X : RB)_{\mathbb{I}^{RB} \otimes \Lambda(|\rho\rangle\langle\rho|^{ABR})} + O(\log \varepsilon) \leq C_{\text{classical}} \right\}.$$

The rate of communication of the protocol is $I^{\varepsilon^4}_{\max}(X : RB) + O(\log \frac{1}{\varepsilon})$. All the entropic quantities above are computed with respect to the state $(\mathbb{I}^{RB} \otimes \Lambda)(|\rho\rangle\langle\rho|^{ARB})$.

Proof. First, fix $\Lambda \in \mathcal{S}$. Then, using this POVM in conjunction with Proposition 6.10 and Lemma 6.11, shows that there exists a protocol (`KD_OneShot`) which for which the rate R_{pure} is achievable with error at most $\varepsilon^{1/32}$, and classical communication $I^{\varepsilon^4}_{\max}(X : RB) + O(\log \frac{1}{\varepsilon})$. Then, taking the supremum of R_{pure} over $\mathcal{S}$ concludes the proof. □

In the following sections, we will prove Proposition 6.10 and Lemma 6.11. To state and prove these claims, we will assume that a POVM Λ is already provided, and no further mention of the set $\mathcal{S}$ will be made. Before we present the actual protocol, we will first start with a bad protocol, which distils a small number of qubits and needs unbounded communication. This protocol, although bad, will serve towards building intuition. We present this in Section 6.1. The full description of `KD_OneShot`can be found in the Sections 6.3 and 6.4.

Remark 6.2. An important feature of our 'achievable' protocols will be that we will not require the use of local randomness in the form of $\pi^{A_{\text{mix}}}$ and $\pi^{B_{\text{mix}}}$, neither will we need Bob to borrow ancilla qubits. We will also not need Alice or Bob to use *local* completely dephasing channels, although they will of course need access to the dephasing channel which sends messages from Alice to Bob. Nevertheless, we will show that our achievable protocols will be almost optimal, in the sense that they will be able to recover pure qubit states *almost* at the optimal rate given in Theorem 5.4.

Thus, in all that follows, the registers $A_{\text{mix}}, B_{\text{mix}}$ and C_{bob} will be omitted. In the interest of brevity we thus abbreviate the register C_{alice} to just C.

6.1 The Need for Measurement Compression

In this section we will introduce a 'bad' protocol for distributed purity distillation which is not optimal with respect to the number of pure qubit states that it distils, but nevertheless helps in understanding some of the key ideas that lead to the other optimal protocol construction that follow in later sections. For the purposes of this demonstration we will not put a bound on classical communication.

We remind the reader of Fact 4.2, restated here for convenience, which we shall use throughout the rest of the paper:

Fact 6.3. *Given a quantum state ρ^A, there exists an ε purity distillation code, which takes the form a unitary operator U^A, which is almost optimal.*

To setup the protocol, we recall the setup of the distributed purity distillation problem, modified suitably according to the statements made in Remark 6.2:

1. Alice and Bob share the state ρ^{AB} at the beginning of the protocol, where Alice has access to the system A and Bob has access to the system B. Alice is also given the POVM $\left\{\Lambda^{A \to X}_x\right\}$ which has outcomes x from the set of symbols $\mathcal{X}$.

2. Alice can borrow any number of qubits $|0\rangle$ as ancilla, but has to account for them at the end of the protocol. For example, Alice can choose to act the POVM Λ on the system A, but she has to do this *coherently* by borrowing $\log|\mathcal{X}|$ number of qubits.

3. Suppose Alice borrows the ancilla qubits in the system C. Then she is allowed to perform any local unitary of the following form:
$$U_{\text{ALICE}} : AC \to A_p A_g X_A.$$
The system A_p is meant to hold the pure states that Alice distils on her end. Note that as per Remark 6.2, our protocol will not require Alice to have access to private randomness or local completely dephasing maps.

4. Alice and Bob share a completely dephasing channel, i.e. a CPTP map $\mathcal{P} : X_A \to X_B$ where the systems X_A and X_B are isomorphic. The action of the map is described with respect to a fixed basis $\left\{|x\rangle^{X_A}\right\}$:
$$\mathcal{P}\left(\rho^{X_A}\right) = \sum_x \langle x|\rho|x\rangle\, |x\rangle\langle x|^{X_B}$$
The choice of basis can be fixed by Alice and Bob before the protocol starts.

5. Bob is allowed to use local unitaries on the systems in his possession, i.e., he is allowed to use unitaries of the following sort:
$$U_{\text{BOB}} : BX_B \to B_p B_g$$
where the system B_p is meant to hold the pure states that he distils. Note that as per Remark 6.2, our protocol will not require Bob to have access to local pure states, private randomness or local completely dephasing maps.

We require that at the end of the protocol, the state $\sigma^{A_pB_p}$ should satisfy the following constraint:
$$\left\|\sigma^{A_pB_p} - |0\rangle\langle 0|_p^A \otimes |0\rangle\langle 0|^{B_p}\right\|_1 \leq \varepsilon$$

For the purposes of this section we will assume that $X_A \cong X_B \cong X$. The protocol itself is given in Table 3.

**Here (see Table 3) the unitary operators U_x and V_x are given by Fact 6.3. Let us analyse Protocol A. We claim the following proposition:

Proposition 6.4. *Protocol A produces*
$$\log|A| - H_H^{\varepsilon^2}(A|X) + \log|B| - H_H^{\varepsilon^2}(B|X) - \log|\mathcal{X}| - O\left(\log\frac{1}{\varepsilon}\right)$$
number of pure qubit states.

Proof. First, fix an $x \in \mathcal{X}$, and consider the Schmidt decomposition of the state $|\rho_x\rangle^{ABR}$:
$$|\rho_x\rangle^{ABR} = \sum_s \lambda_s |s\rangle^A |s\rangle^{BR}.$$
Consider the set of the smallest λ_s whose squares add up to at most ε. Let us call this set BAD. The action of U_x is to relabel those $|s\rangle^A$ which have a corresponding λ_s which is not in BAD:
$$U_x : |s\rangle^A \to |s\rangle^{A_g} |0\rangle^{A_p} \quad \forall\, |s\rangle \quad \text{such that } \lambda_s \notin \text{BAD}.$$
The vector $|s\rangle^{A_g}$ is simply a low dimensional embedding of $|s\rangle^A$ into the system A_g, which has dimension at least $2^{\widetilde{H}_H^{\varepsilon}(A)-1}$. This embedding preserves the pairwise inner products between the vectors, i.e., for all s, s' such that $\lambda_s, \lambda_{s'} \notin$ BAD:
$$\langle s|s'\rangle^{A_g} = \langle s|s'\rangle^A.$$

Alice	**Bob**
Borrow $\log \lvert\mathcal{X}\rvert$ qubits as pure ancilla.	
Act the POVM $\Lambda^{A\to X}$ coherently on $\lvert\rho\rangle^{ABR}$	
State : $\lvert\rho\rangle^{X_A ABR} \coloneqq \sum_{x\in\mathcal{X}} \lvert x\rangle^{X_A} \sqrt{\Lambda_x^A}\, \lvert\rho\rangle^{ABR}$	
Define: $\lvert\rho_x\rangle^{ABR} \coloneqq \frac{1}{\sqrt{\mathrm{Tr}[\Lambda_x^A \rho^A]}} \sqrt{\Lambda_x}\, \lvert\rho\rangle^{ABR}$	
**Define: $U_x^{A\to A_p A_g}$ be the locally optimal local purity distillation protocol for ρ_x^A	
Act unitary $\sum_{x\in\mathcal{X}} \lvert x\rangle\langle x\rvert^{X_A} \otimes U_x^{A\to A_p A_g}$ on state $\lvert\rho\rangle^{X_A ABR}$	
$\xrightarrow{X_A\to X_B}$	
End of Alice's Actions	
	State on $X_B B$: $\rho^{X_B B} \coloneqq \sum_x P_X(x) \lvert x\rangle\langle x\rvert^{X_B} \otimes \rho_x^B$
	**Define: $V_x^{B\to B_p B_g}$ be the locally optimal local purity distillation protocol for ρ_x^B
	Act unitary $\sum_{x\in\mathcal{X}} \lvert x\rangle\langle x\rvert^{X_B} \otimes V_x^{B\to B_p B_g}$ on state $\rho^{X_B B}$

Table 3: **Protocol A**

We can then write:

$$U_x^{A\to A_p A_g} \lvert\rho_x\rangle^{ARB} = \sum_{s:\lambda_s\notin \mathrm{BAD}} \lambda_s \lvert 0\rangle^{A_p} \lvert s\rangle^{A_g} \lvert s\rangle^{RB} + \lvert \mathrm{JUNK}\rangle^{A_p A_g RB}.$$

It is then not hard to see that

$$\left\| U_x \cdot \rho_x^{ARB} - \lvert 0\rangle\langle 0\rvert^{A_p} \otimes \sum_{\substack{s,s' \\ \lambda_s,\lambda_{s'}\notin \mathrm{BAD}}} \lambda_s \lambda_{s'} \lvert s\rangle\langle s'\rvert^{A_g} \otimes \lvert s\rangle\langle s'\rvert^{RB} \right\|_1 \leq O(\sqrt{\varepsilon}).$$

Tracing out the system A_g and noting that the substate $\lvert 0\rangle\langle 0\rvert^{A_p} \otimes \sum_{s\lambda_s\notin \mathrm{BAD}} \lambda_s \lvert s\rangle\langle s\rvert^{RB}$ is ε close to ρ_x^{RB}, one can see that:

$$\left\| \mathrm{Tr}_{A_g}\left[U_x \cdot \rho_x^{ARB}\right] - \lvert 0\rangle\langle 0\rvert^{A_p} \otimes \rho_x^{RB} \right\|_1 \leq O(\sqrt{\varepsilon}).$$

Now, consider the cq state:

$$\sum_x P_X(x) \lvert x\rangle\langle x\rvert^{X_A} \otimes \rho_x^A,$$

where $P_X(x)$ is the probability of the outcome x when ρ^A is measured with the POVM Λ. Then, using Lemma 2.23 we see that there exists a subset of x's, which we call $\mathcal{S}_{\text{ALICE}}$ such that

$$\begin{aligned}&\Pr_{P_X}\left[\mathcal{S}_{\text{ALICE}}\right] \geq 1 - 2\sqrt{\varepsilon} \\ &H_H^{\varepsilon}(\rho_x^A) \leq H_H^{\varepsilon^2}(A|X) - \log\varepsilon \;\; \forall x \in \mathcal{S}_{\text{ALICE}}.\end{aligned}$$

Collating the arguments above, one can then see that the state on the system $X_B RB$ after Alice sends the system X_A through the dephasing channel satisfies the following property:

$$\left\| \sum_x P_X(x) \left|x\right\rangle \left\langle x\right|^{X_B} \otimes \text{Tr}_{A_g}\left[U_x \cdot \rho_x^{ARB}\right] - \left|0\right\rangle \left\langle 0\right|^{A_p} \otimes \left(\sum_x P_X(x) \left|x\right\rangle \left\langle x\right|^{X_B} \otimes \rho_x^{RB} \right) \right\|_1 \leq O(\sqrt{\varepsilon}),$$

where the system A_p is constituted by $H_H^{\varepsilon^2}(A|X) - \log\varepsilon$ qubits. Another applpication of Lemma 2.23 shows us that there exists a set $\mathcal{S}_{\text{BOB}}$ such that

$$\begin{aligned}&\Pr_{P_X}\left[\mathcal{S}_{\text{BOB}}\right] \geq 1 - 2\sqrt{\varepsilon} \\ &H_H^{\varepsilon}(\rho_x^B) \leq H_H^{\varepsilon^2}(B|X) - \log\varepsilon \;\; \forall x \in \mathcal{S}_{\text{BOB}}.\end{aligned}$$

where the entropic quantities in the expression above are computed with respect to the cq state $\sum_x P_X(x) \left|x\right\rangle \left\langle x\right|^{X_B} \otimes \rho_x^B$. Therefore, using arguments that are similar to those we used in the case of Alice, we see that after Bob's actions and discarding the system B_g, the global state is $O(\sqrt{\varepsilon})$ close to pure states on the A_p and B_p system, where:

$$\log\left|A_p B_p\right| \geq \log\left|AB\right| - H_H^{\varepsilon^2}(A|X) - H_H^{\varepsilon^2}(B|X) + O(\log\varepsilon).$$

Recall however that we now have to adjust for the fact that Alice had borrowed $\log|\mathcal{X}|$ qubits. Therefore, the net number of pure qubits distills is:

$$\log\left|A_p B_p\right| - \log|C| \geq \log\left|AB\right| - H_H^{\varepsilon^2}(A|X) - H_H^{\varepsilon^2}(B|X) - \log|\mathcal{X}| + O(\log\varepsilon).$$

This concludes the proof. □

As mentioned earlier, the number of pure qubit states that Protocol A distils is nowhere near optimal. The is of course due to the $-\log|\mathcal{X}|$ term over which we have no control. To fix this issue, we need to replace the POVM Λ with some other POVM Λ' which has far fewer number of outcomes, yet still allows Alice and Bob to distil $\log|A| - H_H^{\varepsilon}(A|X)$ and $\log|B| - H_H^{\varepsilon}(B|X)$ pure qubit states. This is exactly what the measurement compression theorem allows us to do, as we explain in the next section.

6.2 Measurement Compression

As mentioned in the last section, the we need to replace the POVM Λ with a POVM Λ' which has a much smaller number of outcome, in order to increase the number of pure qubit states that Protocol A distils. However, an issue with this strategy is that this new POVM may not allow Alice and Bob to individually distil $\log|A| - H_H^{\varepsilon}(A|X)$ and $\log|B| - H_H^{\varepsilon}(B|X)$ pure qubits. The measurement compression theorem comes to our aid here. Thus in this section we take a small detour from our exposition to state the measurement compression theorem. Suppose we are given a bipartite quantum state ρ^{AB} and a POVM $\Lambda^{A\to X}$. To understand the action of this POVM on the state ρ^{AB}, consider a purification $\left|\rho\right\rangle^{ABR}$. It can be shown (see [22]) that the global state, after the action of the POVM on the system A, looks like

$$\sum_x P_X(x) \left|x\right\rangle \left\langle x\right|^X \otimes \rho_x^{BR}, \tag{1}$$

where

$$\rho_x^{BR} \coloneqq \frac{1}{\text{Tr}\left[\Lambda_x \left|\rho\right\rangle \left\langle\rho\right|^{ABR}\right]} \text{Tr}_A\left[\Lambda_x \left|\rho\right\rangle \left\langle\rho\right|^{ABR}\right].$$

and $P_X(x)$ is the probability of the outcome x when ρ^A is measured using the POVM Λ. The goal of the measurement compression theorem is to replace Λ by some other POVM $\Lambda'^{A\to Y}$ such that the support size of the distribution P_Y induced by Λ' is much smaller than that of P_X, yet the post measurement state $\sum_y P_Y(y) |y\rangle \langle y|^Y \otimes \rho_y^{BR}$ is close to the ideal post measurement state in Equation 1. This of course may not be possible with a single POVM Λ' (the distribution P_X may not be compressible). However, the measurement compression theorem gets around this issue by using multiple POVMs, indexed by $k \in [K]$, each with a small number of outcomes. Which of these POVMs one chooses to actually do the measurement is decided by picking k randomly. Let us refer to these 'smaller' POVMs as $\Theta^A(k) = \left\{\Theta_1^A(k), \Theta_2^A(k), \ldots, \Theta_L^A(k)\right\}$, where L is the number of outcomes. The new measurement process can then be encapsulated as follows:

1. Pick $k \xleftarrow{R} [K]$.
2. Measure the register A of the state ρ^{AB} using the smaller POVM $\Theta^A(k)$.
   Suppose this measurement produces an outcome $\ell \in [L]$.
3. Map the symbol (k, ℓ) appropriately to an $x \in \mathcal{X}$ to recover the correct measurement outcome.

Roughly, the measurement compression theorem says that, as long as K and L are large enough, the procedure above produces a post measurement state that is close to the ideal state in Equation 1. We give the precise statement of the theorem below [5]:

Fact 6.5. *Given the bipartite quantum state ρ^{AB} and the POVM $\{\Lambda_x\}_x$ where $x \in \mathcal{X}$, let $|\rho\rangle^{ABR}$ be some purification of ρ^{AB} and the ideal post measurement state, when the A register of ρ^{AB} is measured using Λ is given by:*

$$\sum_x P_X(x) |x\rangle \langle x|^X \otimes \rho_x^{BR}.$$

Here P_X is the distribution induced by the measurement on the set of symbols $\mathcal{X}$. Suppose we are given integers K and L. Then, as long as

$$\log K + \log L \geq H_{\max}^{'\varepsilon^4}(X) + O(\log \frac{1}{\varepsilon})$$
$$\log L \geq I_{\max}^{\varepsilon^4}(X : RB) + O(\log \frac{1}{\varepsilon}).$$

there exist POVMs $\Theta^A(1), \Theta^A(2), \ldots \Theta^A(K)$, where each POVM Θ_k^A has outcomes in the set $[L] \bigcup \{\perp\}$ ($\perp$ signifying the outcome corresponding o failure), and a function

$$f : [K] \times [L] \to \mathcal{X}$$

such that

$$\left\| \rho^{XBR} - \sum_x \sum_{k,\ell} Q_{KL}(k,\ell) \cdot 1_{f(k,\ell)=x} |x\rangle \langle x|^X \otimes \sigma_{f(k,\ell)}^{RB} \right\|_1 \leq O(\varepsilon).$$

where $Q_K \overset{\varepsilon^{1/4}}{\approx} \mathbf{Unif}[K]$, $Q_{L|k}$ is the distribution induced on the set $[L] \bigcup \{\perp\}$ by the POVM $\Theta^A(k)$ and

$$\sigma_{f(k,\ell)}^{RB} \coloneqq \frac{1}{\mathrm{Tr}\left[\Theta_\ell^A(k) |\rho\rangle \langle\rho|^{ARB}\right]} \mathrm{Tr}_A \left[\Theta_\ell^A(k) |\rho\rangle \langle\rho|^{ARB}\right].$$

We will make us of a couple of other useful facts about measurement compression, specifically that the distribution Q_{KL} is close to the uniform distribution on $[K] \times [L]$ and that for all (k, ℓ) in the support of Q_{KL}, the state $\sigma_{k,\ell}^{RB}$ is close to $\rho_{f(k,\ell)}^{RB}$. We state these facts below:

Fact 6.6. *For all k in the support of the distribution Q_K, and for all ℓ in the support of the distribution $Q_{L \mid k}$ (aside from the outcome $\perp$), it holds that*

$$\left\| \sigma^{RB}_{f(k,\ell)} - \rho^{RB}_{f(k,\ell)} \right\|_1 \leq O(\varepsilon).$$

Fact 6.7. *Given the setup of Fact 6.5, it holds that*

$$\|Q_{KL} - \mathbf{Unif}[K] \times \mathbf{Unif}[L]\|_1 \leq O(\varepsilon^{1/2}).$$

For a proof of these facts, see the proof of Proposition 4.1 in [5].

6.3 An Improvement in Protocol A: Protocol B

We can now use the measurement compression theorem to design a better protocol for purity distillation than Protocol A. The idea is to make use of the two indices k and ℓ that are implicit in the measurement compression theorem. Recall that the index of the POVM to be used in the measurement process is given by k. Naturally, Alice and Bob can use this as shared randomness. Although shared randomness is not one of the resources that Alice and Bob are allowed to have for purity distillation, we will soon get rid of it by derandomising. Next, Alice can measure her register A using the POVM $\Theta^A(k)$ indicated by the shared randomness. Since this POVM has only L outcomes, Alice needs to borrow only $L \geq I^{\varepsilon}_{\max}(X : RB) + O(\log \frac{1}{\varepsilon})$ qubits, which is much smaller than $\log |\mathcal{X}|$. By the measurement compression theorem this measurement process produces a state that is close to the ideal post measurement state if Alice had measured with Λ, after the (k, ℓ) indeces have been mapped to appropriate values of x. Thus, one would expect, via similar reasoning as that which we used to prove Lemma 2.23, that for *most* setting of (k, ℓ), it would hold that:

$$H^{\varepsilon}_H(\rho^A_{k,\ell}) \leq H^{\varepsilon^2}_H(A|X) - \log \varepsilon.$$

Alice can then send her L register to Bob via the dephasing channel. Via the same reasoning as above, we expect that for most values of (k, ℓ) the following should hold:

$$H^{\varepsilon}_H(\rho^B_{k,\ell}) \leq H^{\varepsilon^2}_H(B|X) - \log \varepsilon.$$

Modulo the two assumptions above, this would complete the description of the protocol. Note that the number of qubit states produced by thus protocol would be roughly:

$$\log |A| - H^{\varepsilon^2}_H(A|X) + \log |B| - H^{\varepsilon^2}_H(B|X) - I^{\varepsilon}_{\max}(X : RB),$$

where we have suppressed the additive $\log \varepsilon$ terms. One can show that indeed our intuition is correct, as is shown by the following lemma:

Lemma 6.8. *Given the setup of the measurement compression theorem, there exists a subset $\mathcal{S}$ of $[K] \times [L]$ such that*

$$|\mathcal{S}| \geq (1 - \varepsilon^{1/8})KL$$

and for all $(k, \ell) \in \mathcal{S}$ it holds that it holds that

$$H^{O(\varepsilon^{1/8})}_H(RB \mid k, \ell) \leq H^{O(\varepsilon)}_H(RB \mid X) + O(\log \frac{1}{\varepsilon})$$

and

$$H^{O(\varepsilon^{1/8})}_H(B \mid k, \ell) \leq H^{O(\varepsilon)}_H(B \mid X) + O(\log \frac{1}{\varepsilon}).$$

The proof of this lemma is long but does not offer much further insight into the protocol. The reader can find it in Appendix A. To describe our new protocol, we first list the necessary assumptions as required by Fact 6.5:

Assumption 6.9. Assumptions for Protocol B

1. Alice and Bob are given a bipartite state ρ^{AB} with purification $|\rho\rangle^{ABR}$, and also a POVM Λ. The ideal post measurement state when this POVM acts on the register A is given by:

$$\sum_x P_X(x) |x\rangle \langle x| \otimes \rho_x^{BR}.$$

2. There exist integers K and L such that

$$\log K + \log L \geq H_{\max}^{\prime \varepsilon^4}(X) + O(\log \frac{1}{\varepsilon})$$
$$\log L \geq I_{\max}^{\varepsilon^4}(X : RB) + O(\log \frac{1}{\varepsilon}).$$

3. Alice possesses the POVMs $\Theta^A(1), \Theta^A(2), \ldots, \Theta^A(K)$ whose existence is implied by Fact 6.5. Each of these POVMs produces outputs in the set $[L] \bigcup \{\perp\}$.

4. We will use the notation $K_A K_B$ to denote a public coin register that is available to both Alice and Bob. We set $|K_A| = |K_B| = K$. The register in which Alice will store the outcome of the measurement will be referred to as L_A.

5. There is a completely dephasing channel from Alice to Bob given by $\mathcal{P}^{L_A \to L_B}$ where $L_A \cong L_B$.

6. The distribution on the public coin register is given by Q_K, as defined in Fact 6.5.

7. Given a POVM element $\Theta_\ell^A(k)$, we define

$$|\rho_{k,\ell}\rangle^{ABR} := \frac{1}{\sqrt{\mathrm{Tr}\left[\Theta_\ell^A(k) \rho^{ABR}\right]}} \sqrt{\Theta_\ell^A(k)} |\rho\rangle^{ABR}.$$

   and its associated marginals of interest accordingly.

We now describe Protocol B in Table 4.

Proposition 6.10. *Protocol B distils*

$$\log |A| - H_H^\varepsilon(A \mid X) + \log |B| - H_H^\varepsilon(B \mid X) - I_{\max}^{\varepsilon^4}(X : RB) + O(\log \varepsilon)$$

number of pure qubits with error $O(\varepsilon^{1/16})$. The protocol also uses $I_{\max}^{\varepsilon^4}(RB : X) + O(\log \frac{1}{\varepsilon})$ amount of classical communication. The entropic quantities above are all computed with respect to the state

$$\sum_x |x\rangle \langle x|^X \otimes \mathbb{I}^{RB} \otimes \Lambda_x \left(|\rho\rangle \langle \rho|^{ABR} \right)$$

where $|\rho\rangle^{ABR}$ is a purification of ρ^{AB}.

Proof. We will first invoke Lemma 6.8 to note that, for at least $(1 - O(\varepsilon^{1/8}))$ fraction of indices KL, it holds that

$$H_H^{O(\varepsilon^{1/8})}(\rho_{k,\ell}^{RB}) \leq H_H^{O(\varepsilon)}(RB|X) + O(\log \frac{1}{\varepsilon}).$$

Note that, for fixed (k, ℓ) we recover at least

$$\log |A| - H_H^{O(\varepsilon^{1/8})}(\rho_{k,\ell}^A) - 1$$

amount of purity. Next, we use the fact that for pure states, such as $|\rho_{k,\ell}\rangle^{ARB}$, Lemma 2.16 implies that:

$$H_H^{\varepsilon^{1/8}}(\rho_{k,\ell}^A) = H_H^{\varepsilon^{1/8}}(\rho_{k,\ell}^{RB})$$

Alice	**Bob**
Shared public coin in $K_A K_B$	
Borrow $I_{\max}^{\varepsilon^4}(X:RB) + O(\log \frac{1}{\varepsilon})$ pure ancilla qubits in system L_A.	
Apply the isometry $\sum_k \|k\rangle\langle k\|^{K_A} \otimes \sum_\ell \|\ell\rangle^{L_A} \sqrt{\Theta_\ell^A(k)}$ on the system A.	
Define $U_{k,\ell}$ as local optimal distillation code for $\rho_{k,\ell}^A$.	
Apply $\sum_\ell \|\ell\rangle\langle\ell\|^{L_A} \otimes U_{k,\ell}^{A\to A_p A_g}$ on AL_A.	
$\xrightarrow{L_A \to L_B}$	
End of Alice's Actions	
	Do locally optimal protocol on B conditioned on the contents of $K_B L_B$.

Table 4: **Protocol B**

Thus, using the same arguments as we saw in the proof of Proposition 6.4, we can show that for all the pairs (k,ℓ) which satisfy the above conditions, it holds that:

$$\left\| \mathrm{Tr}_{A_g}\left[U_{k,\ell} \cdot \rho_{k,\ell}^{ARB}\right] - |0\rangle\langle 0|^{A_p} \otimes \rho_{k,\ell}^{RB} \right\|_1 \leq O(\sqrt{\varepsilon}).$$

where the system A_p consists of $\log|A| - H_H^{\varepsilon^{1/4}}(RB|X) + O(\log\varepsilon) - O(1)$ qubits. We can then conclude that the global state after Alice sends the system L_A through the dephasing channel satisfies the following condition:

$$\left\| \sum_{k,\ell} Q_{KL}(k,\ell) |k\rangle\langle k|^K \otimes |\ell\rangle\langle\ell|^{L_A} \otimes \mathrm{Tr}_{A_g}\left[U_{k,\ell} \cdot \rho_{k,\ell}^{ARB}\right] \right.$$
$$\left. - |0\rangle\langle 0|^{A_p} \otimes \sum_{k,\ell} Q_{KL}(k,\ell) |k\rangle\langle k|^K \otimes |\ell\rangle\langle\ell|^{L_A} \otimes \rho_{k,\ell}^{RB} \right\|_1 \leq O(\varepsilon^{1/8})$$

where we have used the fact that the distribution Q_{KL} is $O(\varepsilon^{1/2})$ to the uniform distribution on $[K]\times[L]$ (see Fact 6.7). Thus, on her side, Alice distils at least

$$|A_p| \geq \log|A| - H_H^{\varepsilon^{1/4}}(RB|X) + O(\log\varepsilon) - O(1)$$

amount of purity.

To analyse Bob's actions, we again invoke Lemma 6.8 and recall that, for at least $1 - O(\varepsilon^{1/8})$ fraction of indices KL, it holds that

$$H_H^{O(\varepsilon^{1/8})}(\rho_{k,\ell}^B) \leq H_H^{O(\varepsilon)}(B \mid X) + O(\log\frac{1}{\varepsilon}).$$

This implies that, for most indices k and ℓ, there exists a local unitary $V_{k,\ell}^{B\to B_pB_g}$ such that

$$\left\| \mathrm{Tr}_{B_g}\left[V_{k,\ell}\cdot\rho_{k,\ell}^{B}\right] - |0\rangle\langle 0|^{B_p} \right\|_1 \leq O(\sqrt{\varepsilon})$$

where we see that

$$|B_p| \geq |B| - H_H^{O(\varepsilon)}(B|X) + O(\log\varepsilon) - O(1).$$

Then, using the fact that the distribution Q_{KL} is $O(\varepsilon^{1/2})$ close to the uniform distribution on $[K]\times[L]$ and the arguments we used for Alice's actions, we see that the following holds:

$$\left\| \sum_{k,\ell} Q_{KL}(k,\ell)\,|k\rangle\langle k|^K \otimes |\ell\rangle\langle\ell|^{L_A} \otimes \mathrm{Tr}_{A_gB_gR}\left[V_{k,\ell}\otimes U_{k,\ell}\cdot\rho_{k,\ell}^{ARB}\right] \right.$$
$$\left. - |0\rangle\langle 0|^{A_p}\otimes|0\rangle\langle 0|^{B_p}\otimes\left(\sum_{k,\ell} Q_{KL}(k,\ell)\,|k\rangle\langle k|^K\otimes|\ell\rangle\langle\ell|^{L_A}\right)\right\|_1 \leq \varepsilon^{1/16}$$

Tracing out all registers but the systems A_pB_p implies the result, where we see that the *net* number of pure qubits that Protocol B distilled is given by:

$$\log|A| - H_H^{O(\varepsilon)}(RB|X) + \log|B| - H_H^{O(\varepsilon)}(B|X) - I_{\max}^{\varepsilon^4}(X:RB) + O(\log\varepsilon) - O(1).$$

It is not hard to show that for states of the form

$$\sum_x |x\rangle\langle x|^X \otimes \mathbb{I}^{RB}\otimes\Lambda_x\left(|\rho\rangle\langle\rho|^{ABR}\right)$$

it holds that

$$H_H^{\delta}(RB|X) = H_H^{\delta}(A|X).$$

Plugging this in into the above expression, the result follows. The claim about the number of bits of classcial communication used by Protocol B follows directly from its specification. This concludes the proof. □

6.4 Removing the Public Coin from Protocol B: `KD_OneShot`

In this section we derandomise Protocol B by removing the public coin registers K_AK_B to obtain `KD_OneShot`. We show this in the following lemma:

Lemma 6.11. *Given the setting of Proposition 6.10, there exists a subset $\mathcal{T}\subseteq[K]$ of size at least $(1-O(\varepsilon^{1/32}))K$, such that for any $k\in\mathcal{T}$, if Alice runs Protocol B with only the POVM corresponding to this k, the resulting protocol, called* `KD_OneShot`*, distils as many pure qubits as Protocol B.*

Proof. Recall that in Protocol B at the end of Bob's actions, the global state satisfied the following property:

$$\left\| \sum_{k,\ell} Q_{KL}(k,\ell)\,|k\rangle\langle k|^K \otimes |\ell\rangle\langle\ell|^{L} \otimes \mathrm{Tr}_{A_gB_gR}\left[V_{k,\ell}\otimes U_{k,\ell}\cdot\rho_{k,\ell}^{ARB}\right] \right.$$
$$\left. - |0\rangle\langle 0|^{A_p}\otimes|0\rangle\langle 0|^{B_p}\otimes\left(\sum_{k,\ell} Q_{KL}(k,\ell)\,|k\rangle\langle k|^K\otimes|\ell\rangle\langle\ell|^{L}\right)\right\| \leq \varepsilon^{1/16}$$

To derandomise the above protocol, we define

$$\sigma_k^{A_pB_p} := \sum_\ell Q_{L\,|\,k}(\ell\mid k)\,\mathrm{Tr}_{A_gB_gR}\left[V_{k,\ell}\otimes U_{k,\ell}\cdot\rho_{k,\ell}^{ARB}\right]$$

Then, using block diagonality, we see that

$$\sum_k Q_K(k) \left\| \sigma_k^{A_p B_p} - |0\rangle\langle 0|^{A_p} \otimes |0\rangle\langle 0|^{B_p} \right\| \leq \varepsilon^{1/16}.$$

This immediately proves that there exists a k such that if we run the protocol for only that fixed k, Alice and Bob distil the same amount of purity as in the protocol with shared randomness, while making an error at most $\varepsilon^{1/16}$. In fact, since Q_K is $O(\varepsilon^{1/2})$ close to the uniform distribution on $[K]$, this implies that at least $1 - O(\varepsilon^{1/32})$ fraction of k's in $[K]$ satisfy this property. This concludes the proof. □

7 The Protocol With Small Ancilla

In the previous section, we proved a lower bound on $\kappa_\varepsilon^{\rightarrow}(\rho^{AB}, C_{\text{classical}})$. We did this by first fixing a POVM Λ^A on the system A, then using the `KD_OneShot`protocol to extract roughly $\log|AB| - H_H^\varepsilon(A|X) - H_H^\varepsilon(B|X) - I_{\max}^{\varepsilon^4}(X : RB)$ pure qubits. In the process Alice was required to borrow roughly $I_{\max}^{\varepsilon^4}(X : RB)$ many ancilla qubits. The lower bound on $\kappa_\varepsilon^{\rightarrow}(\rho^{AB}, C_{\text{classical}})$ was then obtained by taking an infimum over the POVMs Λ over the subset $\mathcal{S}$ (see Theorem 5.4 for specifics).

In this section we show that, given the same fixed POVM Λ, there exists a protocol which we call `FewQubits` , which uses far fewer ancilla qubits than `KD_OneShot`, yet manages to distil the pure qubits at the same rate as that of `KD_OneShot`. In fact, we show in Corollary 8.1 in Section 8 that `FewQubits` outperforms `KD_OneShot`in terms of the number of qubits borrowed by Alice, as long as ρ^A is not very close to the maximally mixed state. We also show in Corollary 8.2 in Section 8 that in the case when unbounded classical communication is allowed, `FewQubits` always outperforms `KD_OneShot`.

We state the main theorem of this section below, which shows the existence of the `FewQubits` protocol. As in Section 6, we assume that we are given a fixed POVM Λ to state and prove our results.

Theorem 7.1. Main Theorem *Consider a bipartite state ρ^{AB} shared between two parties Alice and Bob, and a POVM $\{\Lambda_x\}_x$ where the symbol x belongs to a set of symbols $\mathcal{X}$. Let $|\rho\rangle^{ABR}$ be a purification of ρ^{AB}. Consider the control state*

$$\rho^{ABRX} := \sum_x |x\rangle\langle x|^X \otimes \Lambda_x^A \left(|\rho\rangle\langle\rho|^{ABR} \right).$$

Let us refer to the number of pure qubits that Alice and Bob can distil as $\text{Purity}_{\text{ALICE}}$ *and* $\text{Purity}_{\text{BOB}}$*. Then, there exists a protocol `FewQubits` where Alice and Bob are allowed only local unitary operations, and one way classical communication from Alice to Bob such that:*

1. **Case I:** *If $I_{\max}^{\varepsilon^4}(RB : X) + H_H^{O(\varepsilon)}(RB|X) - O(\log\varepsilon) \leq \log|A|$, they are able to distil the following number of pure qubits:*

$$\begin{aligned} \text{Purity}_{\text{ALICE}} &\geq \log|A| - I_{\max}^{\varepsilon^4}(RB : X) + H_H^{O(\varepsilon)}(RB|X) + O(\log\varepsilon), \\ \text{Purity}_{\text{BOB}} &\geq \log|B| - H_H^{\varepsilon^2}(B|X) + O(\log\varepsilon). \end{aligned}$$

   while borrowing at most $O(\log\frac{1}{\varepsilon})$ ancilla qubits.

2. **Case II:** *If $I_{\max}^{\varepsilon^4}(RB : X) + H_H^{O(\varepsilon)}(RB|X) - O(\log\varepsilon) > \log|A|$, they are able to distil*

$$\begin{aligned} \text{Purity}_{\text{ALICE}} &= 0, \\ \text{Purity}_{\text{BOB}} &\geq \log|B| - H_H^{\varepsilon^2}(B|X) - \Delta(RB|X) + O(\log\varepsilon). \end{aligned}$$

   while borrowing at most $\Delta(RB|X) := H_H^{\varepsilon^2}(RB|X) - H_{\min}^{O(\varepsilon)}(RB|X) - O(\log\varepsilon)$ many qubits.

All entropic quantities above are computed with respect to the control state.

Proof. The proof is implied by Lemma 7.4 and Lemma 7.5. □

Our main task now is to prove Lemma 7.4 and 7.5. To do this, let us start by examining Alice's actions in Protocol C, as described in the last section (see Lemma 6.11). To recap, Alice and Bob share a public coin register K, and based on the contents of K, Alice implements a POVM $\Theta^A(k)$ coherently on her system A. She stores the outcome of this measurement a system L_A which she creates by borrowing roughly $I^{\varepsilon^4}_{\max}(X:RB)$ qubits, where the entropic quantity is computed with respect to a control state

$$\sum_x |x\rangle\langle x|^X \otimes \Lambda^A_x \left(|\rho\rangle\langle\rho|^{ABR}\right).$$

Alice then performs a locally optimal distillation protocol on the state $\rho^A_{k,\ell}$ using the unitary $U^A_{k,\ell}$. For most values of k and the measurement outcome ℓ, the measurement compression theorem then implies that the number of pure qubits that Alice distils is at least $\log|A| - H^{\varepsilon^2}_H(A|X)$ (suppressing the additive $O(\log\varepsilon)$ term).

Later, we derandomised and showed that the public coin register is actually not necessary and for most settings ($1-O(\varepsilon^{1/32})$ fraction) of the public coin k, the corresponding POVM $\Theta^A(k)$ does as well as the randomised protocol. Alice can then choose any of the k from the set $\mathcal{T}$ (see Lemma 6.11 for the definition of $\mathcal{T}$) and run the protocol suing the fixed POVM $\Theta^A(k)$.

Our goal in this section will be to implement the action of $\Theta^A(k)$ *in place*, that is, by borrowing little to no ancilla qubits. To do this we require $\Theta^A(k)$ to have the property that for *most* outcomes $\ell \in [L]\bigcup\{\perp\}$, corresponding to the POVM element $\Theta^A_\ell(k)$, it holds that:

$$H^\varepsilon_H(\rho^{RB}_{k,\ell}) \leq H^{O(\varepsilon)}_H(RB \mid X) + O(\log\frac{1}{\varepsilon})$$

and

$$H^\varepsilon_H(\rho^{B}_{k,\ell}) \leq H^{O(\varepsilon)}_H(B \mid X) + O(\log\frac{1}{\varepsilon}),$$

where we define:

$$\rho^{RB}_{k,\ell} \coloneqq \mathrm{Tr}_A \frac{\sqrt{\Theta^A_\ell(k)}\cdot|\rho\rangle\langle\rho|^{ABR}}{\mathrm{Tr}\left[\Theta^A_\ell(k)\,|\rho\rangle\langle\rho|^{ABR}\right]}$$

for some purification $|\rho\rangle^{ABR}$ of ρ^{AB}. That such a POVM exists is shown below via Lemma 7.2 and Claim 7.3.

Lemma 7.2. *For the setting of Lemma 6.8, there exists a subset $\mathcal{T}' \subseteq [K]$ of size at least $(1-\varepsilon^{1/16})K$, such that for all $k \in \mathcal{T}'$, there exists a subset $\mathrm{NICE}_{L\mid k} \subseteq [L]$ of size at least $(1-\varepsilon^{1/16})L$ such that for all $k \in \mathcal{T}'$ and $\ell \in \mathrm{NICE}_{L\mid k}$:*

$$H^{O(\varepsilon^{1/8})}_H(RB \mid k,\ell) \leq H^{O(\varepsilon)}_H(RB \mid X) + O(\log\frac{1}{\varepsilon})$$

and

$$H^{O(\varepsilon^{1/8})}_H(B \mid k,\ell) \leq H^{O(\varepsilon)}_H(B \mid X) + O(\log\frac{1}{\varepsilon}).$$

.

Claim 7.3. *Consider the setting of Lemma 6.11. There exists a $k \in \mathcal{T}$ such that the corresponding POVM $\Theta^A(k)$ satisfies the requirements of Lemma 7.2.*

The proofs of Lemma 7.2 and Claim 7.3 can be found in Appendix C. The next idea is that the states $\rho^{AB}_{k,\ell}$, for all $\ell \in \mathrm{NICE}_{L\mid k}$, can be perturbed to a nearby state $\widetilde{\rho}^{RB}_{k,\ell}$ by throwing away the smallest eigenvalues which sum to $O(\varepsilon^{1/8})$. We can then consider a purification $|\widetilde{\rho}_{k,\ell}\rangle^{A_gRB}$ of $\rho^{RB}_{k,\ell}$, where this system A_g requires only $\exp\left(H^{O(\varepsilon)}_H(RB|X)\right)$ dimensions.

The key idea then is to define an embedding of the systems L_A (which holds the measurement outcomes) and the system A_g into a space of dimension roughly $\exp\left(I_{\max}^{\varepsilon^4}(RB:X) + H_H^{O(\varepsilon)}(RB|X)\right)$. We do this by defining an appropriate pure state on the systems $A_p L_A A_g RB$ using the states $|\widetilde{\rho}\rangle^{A_g RB}$, and then using Uhlmann's theorem. This Uhlmann isometry gives us Alice's required unitary. Some care is necessary here since not all $\rho_{k,\ell}^{RB}$ have the nice property we need to define the states $|\widetilde{\rho}\rangle^{A_g RB}$. A detailed exposition of these ideas can be found in the proof of the lemma below:

Lemma 7.4. *Suppose that*

$$I_{\max}^{\varepsilon^4}(RB:X) + H_H^{O(\varepsilon)}(RB|X) - O(\log\varepsilon) \leq \log|A|\,,$$

where all the entropic quantities are computed with respect to the control state:

$$\sum_x |x\rangle\langle x|^X \otimes \Lambda_x^A\left(|\rho\rangle\langle\rho|^{ABR}\right),$$

where $|\rho\rangle^{ABR}$ is a purification of ρ^{AB}. Then there exists a unitary operator $U^{A\to A_p A_g L_A}$ and a system A_g such that:

$$\begin{aligned}
&\left\|\mathrm{Tr}_{A_g B_g}\left[V\circ\mathcal{P}\circ U\left(\rho^{AB}\right)\right] - |0\rangle\langle 0|^{A_p}\otimes|0\rangle\langle 0|^{B_p}\right\|_1 \leq \varepsilon^{1/16},\\
&\log|A_p| \geq \log|A| - I_{\max}^{\varepsilon^4}(RB:X) + H_H^{O(\varepsilon)}(RB|X) + O(\log\varepsilon) - 1,\\
&\log|B_p| \geq \log|B| - H_H^{\varepsilon^2}(B|X) + O(\log\varepsilon).
\end{aligned}$$

where $V^{L_B B\to B_p B_g}$ encapsulates Bob's unitary operations.

Proof. Let us start with the POVM $\Theta^A(k)$ which satisfies the requirements of both Lemma 7.2 and Lemma 6.11. That such a $k\in[K]$ and $\Theta^A(k)$ exist is shown in Claim 7.3.

For ease of notation, we henceforth omit the k throughout this proof. This means that we will refer to $\Theta^A(k)$ simply as Θ^A and the set $\mathrm{NICE}_{L\mid k}$ (as given by Lemma 7.2) simply as NICE_L.

We define:

$$\rho_\ell^{RB} \coloneqq \frac{\mathrm{Tr}_A\left[\Theta_\ell^A|\rho\rangle\langle\rho|^{ABR}\right]}{P_\Theta(\ell)},$$

for all $\ell\in[L]\bigcup\{\perp\}$, where

$$P_\Theta(\ell) \coloneqq \mathrm{Tr}\left[\Theta_\ell^A|\rho\rangle\langle\rho|^{ABR}\right].$$

We will show that the substate $\sum_{\ell\in\mathrm{NICE}_L} P_\Theta(\ell)\rho_\ell^{RB}$ is close to the state $\sum_{\ell\in[L]\bigcup\{\perp\}} P_\Theta(\ell)\rho_\ell^{RB}$. To see this, first recall from the construction of Θ^A in [5] that, for all $\ell\neq\perp$,

$$\Theta_\ell^A = \frac{1}{1+O(\varepsilon^{1/4})}\cdot\frac{1}{L}\cdot\left(\rho^{-1/2}\sigma_\ell\rho^{-1/2}\right)^A$$

where the operator ρ is equivalent to the marginal ρ^A of the state ρ^{AB} on the system A, and σ_ℓ is a state which arises during the construction, but will not be important in the context of this proof. Also recall that by construction,

$$\mathrm{Tr}\left[\Theta_\perp^A|\rho\rangle\langle\rho|^{ABR}\right] \leq O(\varepsilon).$$

Then consider the following:

$$\begin{aligned}
&\left\| \sum_{\ell \in \text{NICE}_L} P_\Theta(\ell) \rho_\ell^{RB} - \sum_{\ell \in [L] \bigcup \{\perp\}} P_\Theta(\ell) \rho_\ell^{RB} \right\|_1 \\
= &\left\| \sum_{\ell \in \text{NICE}_L^c \bigcup \perp} P_\Theta(\ell) \rho_\ell^{RB} \right\|_1 \\
\leq &\sum_{\ell \in \text{NICE}_L^c} \left\| \text{Tr}_A \left[\Theta_\ell^A \left| \rho \right\rangle \left\langle \rho \right|^{ABR} \right] \right\|_1 + \left\| \text{Tr}_A \left[\Theta_\perp^A \left| \rho \right\rangle \left\langle \rho \right|^{ABR} \right] \right\|_1 \\
\overset{(a)}{=} &\sum_{\ell \in \text{NICE}_L^c} \text{Tr} \left[\Theta_\ell^A \rho^A \right] + \text{Tr} \left[\Theta_\perp^A \rho^A \right] \\
= &\sum_{\ell \in \text{NICE}_L^c} \frac{1}{1 + O(\varepsilon^{1/4})} \cdot \frac{1}{L} \cdot \text{Tr} \left[\sigma_\ell \right] + \text{Tr} \left[\Theta_\perp^A \rho^A \right] \\
\overset{(b)}{\leq} &\frac{1}{1 + O(\varepsilon^{1/4})} \cdot \frac{|\text{NICE}_L^c|}{L} + O(\varepsilon) \\
\overset{(c)}{\leq} &\frac{\varepsilon^{1/16}}{1 + O(\varepsilon^{1/4})} + O(\varepsilon) \\
\leq &\ O(\varepsilon^{1/16}),
\end{aligned}$$

where in step (a) we have used the fact that the 1-norm of a positive semidefinite matrix is equal to its trace and subsequently traced out the systems RB, in step (b) we have used the upper bound on $\text{Tr} \left[\Theta_\perp^A \left| \rho \right\rangle \left\langle \rho \right|^{ABR} \right]$ and also the structure of the POVM element Θ_ℓ^A for $\ell \neq \perp$. In step (c) we have used Lemma 7.2 to bound the size of the set NICE_L^c.

An immediate consequence of the above calculation is that:

$$\sum_{\ell \in \text{NICE}_L^c \bigcup \{\perp\}} P_\Theta(\ell) \leq O(\varepsilon^{1/16}).$$

We then define:

$$\widetilde{P}_\Theta(\ell) \coloneqq \frac{P_\Theta(\ell)}{\sum\limits_{\ell \in \text{NICE}_L} P_\Theta(\ell)} \ \ \forall \ell \in \text{NICE}_L.$$

It is then not hard to see via Lemma 2.1 in [9] that:

$$\left\| \sum_{\ell \in \text{NICE}_L} \widetilde{P}_\Theta(\ell) \rho_\ell^{RB} - \sum_{\ell \in [L] \bigcup \{\perp\}} P_\Theta(\ell) \rho_\ell^{RB} \right\|_1 \leq O(\varepsilon^{1/16}).$$

We will now invoke Lemma 7.2 and the definition of the set NICE_L to see that for all $\ell \in \text{NICE}_L$,

$$H_H^{O(\varepsilon^{1/8})}(\rho_\ell^{RB}) \leq H_H^{O(\varepsilon)}(RB \mid X) + O(\log \frac{1}{\varepsilon}).$$

Define

$$\widetilde{\rho_\ell}^{RB} \coloneqq \frac{\sqrt{\Pi_\ell} \rho_\ell^{RB} \sqrt{\Pi_\ell}}{\text{Tr} \left[\Pi_\ell \rho_\ell^{RB} \right]},$$

where the operator Π_ℓ^{RB} arises in the definition of $H_H^{O(\varepsilon^{1/8})}(\rho_\ell^{RB})$. Note that from [16] we know that for any state ρ_ℓ^{RB}, the expression $H_H^{O(\varepsilon^{1/8})}(\rho_\ell^{RB})$ is optimised by an operator which commutes with ρ_ℓ^{RB} and which has all eigenvalues 1 aside from maybe the smallest eigenvalue. Furthermore, the 0 eigenvalues of Π_ℓ coincide with the smallest eigenvalues of ρ_ℓ^{RB} which add up to at most $O(\varepsilon^{1/8})$ (see [16] for a proof of these properties). These properties imply the following:

1. The states $\widetilde{\rho}_\ell^{RB}$ for all $\ell \in \text{NICE}_L$ are close to ρ_ℓ^{RB}. To be precise:

$$\begin{aligned} & \|\widetilde{\rho}_\ell - \rho_\ell\|_1 \\ = & \left\| \frac{\sqrt{\Pi_\ell}\rho_\ell\sqrt{\Pi_\ell}}{\text{Tr}\left[\Pi_\ell\rho_\ell\right]} - \rho_\ell \right\|_1 \\ \overset{(a)}{\leq} & O(\varepsilon^{1/16}), \end{aligned}$$

   where in step (a) we have used the fact that $\text{Tr}\left[\Pi_\ell\rho_\ell\right] \geq 1 - \varepsilon^{1/8}$ (by definition of Π_ℓ for all $\ell \in \text{NICE}_L$) and the Gentle Measurement Lemma.

2. The size support of the support of $\widetilde{\rho}_\ell^{RB}$ is bounded above by $2^{H_H^{O(\varepsilon^{1/8})}(\rho_\ell^{RB})} + 1$. To see this, note that

$$\text{rank}(\Pi_\ell) \leq \text{Tr}\left[\Pi_\ell\right] + 1.$$

   Since $\text{Tr}\left[\Pi_\ell\right] = 2^{H_H^{O(\varepsilon^{1/8})}(\rho_\ell^{RB})}$, the claim follows.

A hybrid argument then shows that:

$$\left\| \sum_{\ell \in \text{NICE}_L} \widetilde{P}_\Theta(\ell)\widetilde{\rho}_\ell^{RB} - \sum_{\ell \in [L] \bigcup \{\perp\}} P_\Theta(\ell)\rho_\ell^{RB} \right\|_1 \leq O(\varepsilon^{1/16}).$$

Finally, note that the action of Θ^A on is that of a CPTP map with Kraus operators $|\ell\rangle^{L_A}\sqrt{\Theta_\ell}^A$, followed by the trace out operation on the system A. Therefore, this cannot change the marginal on the system RB, which implies that

$$\sum_{\ell \in [L] \bigcup \{\perp\}} P_\Theta(\ell)\rho_\ell^{RB} = \rho^{RB}.$$

This gives us the following inequality:

$$\left\| \sum_{\ell \in \text{NICE}_L} \widetilde{P}_\Theta(\ell)\widetilde{\rho}_\ell^{RB} - \rho^{RB} \right\|_1 \leq O(\varepsilon^{1/16}). \tag{3}$$

We will now define, for all $\ell \in \text{NICE}_L$, the purification $|\widetilde{\rho}_\ell\rangle^{A_g RB}$ of the state $\widetilde{\rho}_\ell^{RB}$. Here, the system A_g is a system of dimension $2^{H_H^{O(\varepsilon^{1/8})}(\rho_\ell^{RB})} + 1$, which is sufficient by the arguments presented in Item 2. It is important to point out that we use the same space A_g to purify *all* the states $\widetilde{\rho}_\ell^{RB}$.

We further define the pure state:

$$|\widetilde{\rho}\rangle^{L_A A_g RB} := \sum_{\ell \in \text{NICE}_L} \sqrt{\widetilde{P}_\Theta(\ell)}\, |\ell\rangle^{L_A} |\widetilde{\rho}_\ell\rangle^{A_g RB}.$$

Note that since Θ^A has at most $\exp\left(I_{\max}^{\varepsilon^4}(X:RB) - O(\log\varepsilon)\right)$ outcomes (we absorb the additive 1 due to the $\perp$ outcome in the $O(\log\frac{1}{\varepsilon})$ term), the system $L_A A_g$ is of log dimension:

$$\begin{aligned} \log|L_A A_g| &= I_{\max}^{\varepsilon^4}(X:RB) + H_H^{O(\varepsilon^{1/8})}(\rho_\ell^{RB}) + O(\log\frac{1}{\varepsilon}) + O(1) \\ &\overset{(a)}{\leq} I_{\max}^{\varepsilon^4}(X:RB) + H_H^{O(\varepsilon)}(RB \mid X) + O(\log\frac{1}{\varepsilon}) \\ &\overset{(b)}{\leq} \log|A|, \end{aligned}$$

where for the step (a) we have used Lemma 7.2, and step (b) is by the hypothesis of the lemma. This implies that there exists a system A_p of log dimension at least $\log|A| - I_{\max}^{\varepsilon^4}(X:RB) - H_H^{O(\varepsilon)}(RB \mid X) - O(\log\frac{1}{\varepsilon})$, such that:

$$A_p L_A A_g \cong A.$$

We then define the pure state:

$$|\widetilde{\widetilde{\rho}}\rangle^{L_A A_p A_g RB} := |0\rangle^{A_p} |\widetilde{\rho}\rangle^{L_A A_g RB}.$$

Note that by Equation 3,

$$\widetilde{\widetilde{\rho}}^{RB} \overset{O(\varepsilon^{1/16})}{\approx} \rho^{RB}.$$

Therefore, by Uhlmann's theorem, there exists a *unitary operator* $U_\Theta : A \to L_A A_p A_g$ such that:

$$\begin{aligned} & \left\| |\widetilde{\widetilde{\rho}}\rangle\langle\widetilde{\widetilde{\rho}}|^{L_A A_p A_g RB} - U_\Theta^{A\to L_A A_p A_g} \cdot |\rho\rangle\langle\rho|^{ARB} \right\|_1 \\ = & \left\| |0\rangle\langle 0|^{A_p} \otimes |\widetilde{\rho}\rangle\langle\widetilde{\rho}|^{L_A A_g RB} - U_\Theta^{A\to L_A A_p A_g} \cdot |\rho\rangle\langle\rho|^{ARB} \right\|_1 \\ \leq & \; O(\varepsilon^{1/32}) \end{aligned}$$

Next, Alice sends the L_A system through the channel $\mathcal{P}^{L_A\to L_B}$. Then, the following holds by the monotonicity of the 1-norm:

$$\begin{aligned} & \left(\mathcal{P}^{L_A\to L_B} \circ \mathrm{Tr}_{A_g} \circ U_\Theta^{A\to L_A A_p A_g}\right) \cdot |\rho\rangle\langle\rho|^{ARB} \\ \overset{O(\varepsilon^{1/32})}{\approx} & \; |0\rangle\langle 0|^{A_p} \otimes \left(\sum_{\ell\in \mathrm{NICE}_L} \widetilde{P}_\Theta(\ell) |\ell\rangle\langle\ell|^{L_B} \otimes \widetilde{\rho}_\ell^{RB} \right). \end{aligned}$$

Tracing out the system R and by previous arguments, it is then easy to see that:

$$\begin{aligned} & \sum_{\ell\in \mathrm{NICE}_L} \widetilde{P}_\Theta(\ell) |\ell\rangle\langle\ell|^{L_B} \otimes \widetilde{\rho}_\ell^{B} \\ \overset{O(\varepsilon^{1/16})}{\approx} & \sum_{\ell\in \mathrm{NICE}_L} \widetilde{P}_\Theta(\ell) |\ell\rangle\langle\ell|^{L_B} \otimes \rho_\ell^{B} \end{aligned}$$

Then, invoking Lemma 7.2, we see that for all ρ_ℓ^B in the above expression, it holds that:

$$H_H^{O(\varepsilon^{1/8})}(\rho_\ell^B) \leq H_H^{O(\varepsilon)}(B \mid X) + O(\log\frac{1}{\varepsilon}).$$

Bob can then enact the conditional unitary:

$$V^{L_B B\to B_p B_g} = \sum_{\ell\in[L_B]} |\ell\rangle\langle\ell|^{L_B} \otimes V_\ell^{B\to B_p B_g},$$

where $B \cong B_p B_g$. We define the unitary operators $V_\ell^{B\to B_p B_g}$ as follows:

1. For all $\ell \in \mathrm{NICE}_L$, $V_\ell^{B\to B_p B_g}$ performs the locally optimal purity distillation protocol for the state ρ_ℓ^B with error $O(\varepsilon^{1/16})$.
2. For all $\ell \in [L_B] \setminus \mathrm{NICE}_L$, set $V_\ell^{B\to B_p B_g} = \mathbb{I}$, where $\mathbb{I}$ denotes the natural isomorphism between the spaces B and $B_p B_g$.

Then it holds that:

$$V^{L_B B \to B_p B_g} \cdot \left(\sum_{\ell \in \text{NICE}_L} \widetilde{P}_\Theta(\ell) \left|\ell\right\rangle \left\langle \ell\right|^{L_B} \otimes \rho_\ell^B \right)$$
$$\overset{O(\varepsilon^{1/16})}{\approx} \left|0\right\rangle \left\langle 0\right|^{B_p} \otimes \left(\sum_{\ell \in \text{NICE}_L} \widetilde{P}_\Theta(\ell) \left|\ell\right\rangle \left\langle \ell\right|^{L_B} \otimes \rho_\ell'^{B_g} \right),$$

where $\rho_\ell'^{B_g}$ is the state on the B_g system remnant after the locally optimal protocol has been enacted on ρ_ℓ^B. Note that the system B_p has log dimension at least:

$$\log |B_p| \geq \log |B| - H_H^{O(\varepsilon)}(B \mid X) - O(\log \frac{1}{\varepsilon}).$$

The argument is completed by stringing together all of the above inequalities via a hybrid argument and using the monotonicity of the 1-norm. This concludes the proof. □

We will now deal with the case when

$$I_{\max}^{\varepsilon^4}(RB : X) + H_H^{O(\varepsilon)}(RB|X) - O(\log \varepsilon) > \log |A| .$$

It may happen that for some cases $I_{\max}^{\varepsilon^4}(RB : X) + H_H^{O(\varepsilon)}(RB|X) - O(\log \varepsilon)$ exceeds $\log |A|$. In that case, Alice cannot distil any pure qubits. Indeed, she has to borrow some qubits to even implement the unitary U_Θ^A, which we constructed in Lemma 7.4. However, she will be able to implement the unitary U_Θ^A following the same recipe that we showed in Lemma 7.4 if she borrows:

$$I_{\max}^{\varepsilon^4}(RB : X) + H_H^{O(\varepsilon)}(RB|X) - O(\log \varepsilon) - \log |A|$$

qubits. In this case we will use the following bound on $I_{\max}^{\varepsilon^4}(RB : X)$ which was shown in [2]:

$$\begin{aligned} I_{\max}^{\varepsilon^4}(RB : X) &\leq H_{\max}^{O(\varepsilon^8)}(RB) - H_{\min}^{O(\varepsilon^8)}(RB|X) - O(\log \varepsilon) \\ &\leq H_H^{O(\varepsilon^8)}(A) - H_{\min}^{O(\varepsilon^8)}(RB|X) - O(\log \varepsilon) \\ &\leq \log |A| - H_{\min}^{O(\varepsilon^8)}(RB|X) - O(\log \varepsilon). \end{aligned}$$

Therefore, in this case, Alice would have to borrow at most

$$\Delta(RB|X) \coloneqq H_H^{O(\varepsilon)}(RB|X) - H_{\min}^{O(\varepsilon^8)}(RB|X) - O(\log \varepsilon)$$

many qubits. We state this as a lemma below:

Lemma 7.5. *Given the setting of Lemma 7.4, suppose that*

$$I_{\max}^{\varepsilon^4}(RB : X) + H_H^{O(\varepsilon)}(RB|X) - O(\log \varepsilon) > \log |A| .$$

In this case, Alice can implement the unitary $U^{A \to L_A A_g}$ defined in Lemma 7.4 by borrowing at most

$$\Delta(RB|X) \coloneqq H_H^{O(\varepsilon)}(RB|X) - H_{\min}^{O(\varepsilon^8)}(RB|X) - O(\log \varepsilon)$$

many qubits. Note that there is no A_p system for this case since Alice cannot distil any pure qubits by herself. The net purity that Alice and Bob together distil is given by:

$$\log |B| - H_H^{\varepsilon^2}(B|X) - \Delta(RB|X) + O(\log \varepsilon).$$

Note that in the asymptotic iid limit, the case dealt with in Lemma 7.5 does not occur. We conclude by formally showing that `FewQubits` borrows fewer pure qubits than Protocol C in the worst case, as long as $\log |A| - H_H^{O(\varepsilon)}(A) \geq O(\log \frac{1}{\varepsilon})$, i.e., when the state on the system A is even nominally away from maximally mixed:

8 Comparative Analysis of `FewQubits`

In this section we compare the performance of `FewQubits` to that of `KD_OneShot`, both in the general case of bounded communication and in the case when unbounded classical communication is allowed. We show that in both cases, as long as ρ^A is not too close to the maximally mixed state, `FewQubits` outperforms `KD_OneShot`. We state our results as Corollaries 8.1 and 8.2 below, the proofs of which follow as corollaries from Theorem 7.1.

Corollary 8.1. `FewQubits` *borrows fewer qubits as compared to* `KD_OneShot` *as long as* $\log|A| - H_H^{O(\varepsilon)}(A) \geq O(\log\frac{1}{\varepsilon})$.

Proof. Recall that `KD_OneShot` requires Alice to borrow $C_{\text{borrow}} \coloneqq I_{\max}^{\varepsilon^4}(RB:X) + O(\log\frac{1}{\varepsilon})$ number of pure qubits to function. On the other hand, `FewQubits` requires Alice to borrow

$$D_{\text{borrow}} \coloneqq \max\left\{0, I_{\max}^{\varepsilon^4}(RB:X) + H_H^{O(\varepsilon)}(RB|X) - \log|A| + O(\log\frac{1}{\varepsilon})\right\}$$

qubits. Clearly,

$$\begin{aligned}
&C_{\text{borrow}} - D_{\text{borrow}} \\
\geq&\log|A| - H_H^{\varepsilon}(RB \mid X) - O(\log\frac{1}{\varepsilon}) \\
\geq&\log|A| - H_H^{\varepsilon}(RB) - O(\log\frac{1}{\varepsilon}) \\
=&\log|A| - H_H^{\varepsilon}(A) - O(\log\frac{1}{\varepsilon}).
\end{aligned}$$

where we have used the data-processing inequality to show that $H_H^{O(\varepsilon)}(RB \mid X) \leq H_H^{O(\varepsilon)}(RB)$ and Lemma 2.16 to show that $H_H^{O(\varepsilon)}(RB) = H_H^{O(\varepsilon)}(A)$. Therefore the corollary holds as long as $\log|A| - H_H^{O(\varepsilon)}(A) \geq O(\log\frac{1}{\varepsilon})$. This concludes the proof. □

Corollary 8.2. *Given the setup of Theorem 7.1, suppose that the POVM* $\{\Lambda_x\}_x$ *has rank-*1 *elements. Then* `FewQubits` *guarantees the following:*

$$\begin{aligned}
\text{Purity}_{\text{ALICE}} &\geq \log|A| - H_H^{O(\varepsilon^8)}(A) + O(\log\varepsilon) \\
\text{Purity}_{\text{BOB}} &\geq \log|B| - H_H^{\varepsilon^2}(B) + O(\log\varepsilon)
\end{aligned}$$

and the number of qubits that Alice is required to borrow to run the protocol is at most $O(\log\frac{1}{\varepsilon})$.

Proof. First note that the global state is the pure state $|\rho\rangle^{ABR} = \sum_i s_i |i\rangle^A |i\rangle^{RB}$. It is given that the POVM Λ is rank-1, i.e. it is constituted by operators of the form $\{|\varphi_x\rangle\langle\varphi_x|\}_x$. Note that each vector $|\varphi_x\rangle$ has 2-norm at most 1. This is simply because each operator $|\varphi_x\rangle\langle\varphi_x|^A \leq \mathbb{I}^A$ (by the definition of a POVM). By definition of the action of the POVM, the post measurement state ρ^{XBR} is given by:

$$\begin{aligned}
&\sum_x |x\rangle\langle x|^X \otimes \text{Tr}_A\left[\left(I^{RB} \otimes |\varphi_x\rangle\langle\varphi_x|^A\right)|\rho\rangle\langle\rho|^{ABR}\right] \\
=&\sum_x |x\rangle\langle x|^X \otimes \text{Tr}_A\left[\left(I^{RB} \otimes \sqrt{|\varphi_x\rangle\langle\varphi_x|^A}\right)\cdot|\rho\rangle\langle\rho|^{ABR}\right].
\end{aligned}$$

Note that:

$$\begin{aligned}
\left(I^{RB} \otimes \sqrt{|\varphi_x\rangle\langle\varphi_x|^A}\right)|\rho\rangle^{ABR} &= \left(I^{RB} \otimes \sqrt{|\varphi_x\rangle\langle\varphi_x|^A}\right)\sum_i s_i |i\rangle^A |i\rangle^{BR} \\
&= |\widetilde{\varphi}_x\rangle^A \left(\sum_i s_i \langle\varphi_x|i\rangle |i\rangle^{BR}\right)
\end{aligned}$$

where $|\widetilde{\varphi}_x\rangle^A$ is the normalised version of the vector $|\varphi_x\rangle^A$ and we have used the fact that $\sqrt{|\varphi_x\rangle\langle\varphi_x|^A} = |\widetilde{\varphi}_x\rangle\langle\varphi_x|^A$. It is easy to see that:

$$\begin{aligned}\left\|\sum_i s_i \langle\varphi_x|i\rangle |i\rangle^{BR}\right\|_2^2 &= \sum_i s_i^2 |\langle\varphi_x|i\rangle|^2 \\ &= \mathrm{Tr}\left[|\varphi_x\rangle\langle\varphi_x|^A \rho^A\right] \\ &:= P_X(x).\end{aligned}$$

Then, defining $|\psi_x\rangle^{BR}$ to be the normalised version of the vector $\sum_i s_i \langle\varphi_x|i\rangle |i\rangle^{RB}$ we can rewrite the post measurement state ρ^{XBR} as:

$$\rho^{XBR} = \sum_x P_X(x) |x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^{BR}.$$

We know from Lemma 2.24 that for states of this form, it holds that $H_H^\varepsilon(RB \mid X) \leq 0$. This implies that for this case:

$$\begin{aligned}&I_{\max}^{\varepsilon^4}(RB \mid X) + H_H^{O(\varepsilon)}(RB \mid X) + O(\log\frac{1}{\varepsilon}) \\ &\leq I_{\max}^{\varepsilon^4}(RB \mid X) + O(\log\frac{1}{\varepsilon}) \\ &\leq H_{\max}^{O(\varepsilon^8)}(A) + O(\log\frac{1}{\varepsilon}).\end{aligned}$$

where the last line follows from Lemma B.17, [3]. Therefore, by Theorem 7.1 this implies that in this case Alice needs to borrow at most $O(\log\frac{1}{\varepsilon})$ ancilla qubits for `FewQubits` to work. This concludes the proof. □

9 Acknowledgements

The work of R.J. is supported by the NRF grant NRF2021-QEP2-02-P05 and the Ministry of Education, Singapore, under the Research Centres of Excellence program. This work was done in part while R.J. was visiting the Technion-Israel Institute of Technology, Haifa, Israel, and the Simons Institute for the Theory of Computing, Berkeley, CA, USA. S.C. would like to acknowledge support from the National Research Foundation, including under NRF RF Award No. NRF-NRFF2013-13 and NRF2021-QEP2-02-P05 and the Prime Minister's Office, Singapore and the Ministry of Education, Singapore, under the Research Centres of Excellence program. P.S. would like to acknowledge support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.01-0500, for carrying out this research work.

References

[1] Charles H. Bennett, Péter Gács, Ming Li, Paul M. B. Vitányi, and Wojciech H. Zurek. Thermodynamics of computation and information distance. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '93, page 21–30, New York, NY, USA, 1993. Association for Computing Machinery.

[2] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306(3):579–615, aug 2011.

[3] S. Chakraborty, A. Nema, and F. Buscemi. Generalized resource theory of purity: one-shot purity distillation with local noisy operations and one way classical communication. In *Proceedings of the 2023 IEEE International Symposium on Information Theory (ISIT)*, pages 980–984. IEEE, 2023.

[4] Sayantan Chakraborty, Rahul Jain, and Pranab Sen. One-shot non-catalytic distributed purity distillation. In *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–8, 2023.

[5] Sayantan Chakraborty, Arun Padakandla, and Pranab Sen. Centralised multi link measurement compression with side information. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 61–66, 2022.

[6] I. Devetak. Distillation of local purity from quantum states. *Phys. Rev. A*, 71:062303, Jun 2005.

[7] I. Devetak and A. Winter. Distilling common randomness from bipartite quantum states. *IEEE Transactions on Information Theory*, 50(12):3183–3196, dec 2004.

[8] Gilad Gour, Markus P. Müller, Varun Narasimhachar, Robert W. Spekkens, and Nicole Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Physics Reports*, 583:1–58, 2015. The resource theory of informational nonequilibrium in thermodynamics.

[9] Patrick Hayden, Michał Horodecki, Andreas Winter, and Jon Yard. A decoupling approach to the quantum capacity. *Open Systems amp; Information Dynamics*, 15(01):7–19, March 2008.

[10] Michał Horodecki, Karol Horodecki, Paweł Horodecki, Ryszard Horodecki, Jonathan Oppenheim, Aditi Sen(De), and Ujjwal Sen. Local information as a resource in distributed quantum systems. *Phys. Rev. Lett.*, 90:100402, Mar 2003.

[11] Michał Horodecki, Paweł Horodecki, Ryszard Horodecki, Jonathan Oppenheim, Aditi Sen(De), Ujjwal Sen, and Barbara Synak-Radtke. Local versus nonlocal information in quantum-information theory: Formalism and phenomena. *Phys. Rev. A*, 71:062307, Jun 2005.

[12] Michał Horodecki, Paweł Horodecki, and Jonathan Oppenheim. Reversible transformations from pure to mixed states and the unique measure of information. *Phys. Rev. A*, 67:062104, Jun 2003.

[13] Hari Krovi and Igor Devetak. Local purity distillation with bounded classical communication. *Phys. Rev. A*, 76:012321, Jul 2007.

[14] R. Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

[15] Jonathan Oppenheim, Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Thermodynamical approach to quantifying quantum correlations. *Physical Review Letters*, 89(18), oct 2002.

[16] Pranab Sen. Lecture notes: One shot classical and quantum information theory. TIFR Mumbai, 2021.

[17] Alexander Streltsov, Hermann Kampermann, Sabine Wölk, Manuel Gessner, and Dagmar Bruß. Maximal coherence and the resource theory of purity. *New Journal of Physics*, 20(5):053058, may 2018.

[18] L. Szilard. über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift fur Physik*, 53(11-12):840–856, November 1929.

[19] Marco Tomamichel. A framework for non-asymptotic quantum information theory, 2012.

[20] Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality between smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 56(9):4674–4681, 2010.

[21] Alexander Vitanov, Frédéric Dupuis, Marco Tomamichel, and Renato Renner. Chain rules for smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 59(5):2603–2612, 2013.

[22] Mark M Wilde, Patrick Hayden, Francesco Buscemi, and Min-Hsiu Hsieh. The information-theoretic costs of simulating quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 45(45):453001, oct 2012.

[23] Andreas Winter. Extrinsic and intrinsic data in quantum measurements: Asymptotic convex decomposition of positive operator valued measures. *Communications in Mathematical Physics*, 244(1):157–185, jan 2004.

Appendix A Proof of Lemma 6.8

Before we go on to the main proof, we will prove another lemma which will be useful in the main proof of Lemma 6.8.

Lemma A.1. *Consider the set of classical symbols $\mathcal{X}$ and let X be a register which holds symbols from this set. Let B be a quantum register. Suppose that we are given two classical quantum states ρ^{XB} and σ^{XB} as follows:*

$$\rho^{XB} \coloneqq \sum_x P_X(x) \left|x\right\rangle \left\langle x\right|^X \otimes \rho_x^B$$
$$\sigma^{XB} \coloneqq \sum_x Q_X(x) \left|x\right\rangle \left\langle x\right|^X \otimes \sigma_x^B$$

with the promise that for all $x \in \mathcal{X}$ it holds that $\left\|\sigma_x^B - \rho_x^B\right\|_1 \leq \varepsilon$ and $\left\|P_X - Q_X\right\|_1 \leq \varepsilon$. Next, let K be an integer and suppose that we are given a deterministic function $f : [K] \to \mathcal{X}$. Suppose there exists a distribution Q_K on $[K]$ such that the following holds:

$$\left\|Q_K - \mathbf{Unif}[K]\right\|_1 \leq \delta$$
$$\sum_{k:f(k)=x} Q_K(k) = Q_X(x) \ \ \forall x \in \mathcal{X}.$$

Let us also define, for all $k \in [K]$, the states $\sigma_k^B \coloneqq \sigma_{f(k)}^B$. Then, at least $1 - \varepsilon^{1/8} - \delta$ fraction of $k \in [K]$ satisfy the condition that:

$$2^{H_H^{\varepsilon^{1/8}}(\sigma_k^B)} \leq \frac{2^{H_H^{\varepsilon}(B \mid X)_\rho}}{\varepsilon}.$$

Proof. Let Π_{OPT}^{XB} be the optimising operator in the definition of $H_H^{\varepsilon}(B \mid X)$, where we can assume without loss of generality that Π_{OPT} is of the form

$$\Pi_{\text{OPT}} = \sum_x \left|x\right\rangle \left\langle x\right|^X \otimes \Pi_x^B$$

where each Π_x^B satisfies the condition

$$0^B \leq \Pi_x^B \leq \mathbb{I}^B.$$

We claim that $\left|\operatorname{Tr}\left[\Pi_{\text{OPT}}(\rho^{XB} - \sigma^{XB})\right]\right| \leq 2\varepsilon$. To see this, note that:

$$\begin{aligned}
&\left|\operatorname{Tr}\left[\Pi_{\text{OPT}}(\rho^{XB} - \sigma^{XB})\right]\right| \\
\leq &\left\|\rho^{XB} - \sigma^{XB}\right\|_1 \\
\leq &2\varepsilon,
\end{aligned}$$

where the first inequality is by the definition of the 1-norm and the last inequality can be proved using a standard hybrid argument. This immediately implies that:

$$\begin{aligned}
&\sum_x Q_X(x) \operatorname{Tr}\left[\Pi_x \sigma_x\right] \\
&= \operatorname{Tr}\left[\Pi_{\text{OPT}} \sigma^{XB}\right] \\
&\geq \operatorname{Tr}\left[\Pi_{\text{OPT}} \rho^{XB}\right] - 2\varepsilon \\
&\geq 1 - 3\varepsilon.
\end{aligned}$$

Markov's inequality then implies that there exists a set $\text{GOOD}_1 \subseteq \mathcal{X}$ such that $\Pr_{Q_X}\left[\text{GOOD}_1\right] \geq 1 - \sqrt{3\varepsilon}$ and for all $x \in \text{GOOD}_1$ it holds that:

$$\operatorname{Tr}\left[\Pi_x \sigma_x^B\right] \geq 1 - \sqrt{3\varepsilon}.$$

Since Q_X and P_X are close in the 1-norm, this implies that $\Pr_{P_X}[\text{GOOD}_1] \geq 1 - \sqrt{4\varepsilon}$. Next, note that by definition,

$$\sum_x P_X(x) \operatorname{Tr}[\Pi_x] = 2^{-H_H^\varepsilon(B \mid X)}.$$

Again using Markov's inequality, we see that there exists a set $\text{GOOD}_2 \subseteq \mathcal{X}$ such that $\Pr_{P_X}[\text{GOOD}_2] \geq 1 - \varepsilon$ and for all $x \in \text{GOOD}_2$ it holds that:

$$\operatorname{Tr}[\Pi_x] \leq \frac{2^{H_H^\varepsilon(B \mid X)}}{\varepsilon}.$$

Therefore, we have identified a set $\text{GOOD}_X \coloneqq \text{GOOD}_1 \bigcap \text{GOOD}_2$ of probability at least $1 - \sqrt{5\varepsilon}$ (under P_X) such that for all $x \in \text{GOOD}_X$:

$$\begin{aligned}
\operatorname{Tr}\left[\Pi_x \sigma_x^B\right] &\geq 1 - \sqrt{3\varepsilon} \\
\operatorname{Tr}[\Pi_x] &\leq \frac{2^{H_H^\varepsilon(B \mid X)}}{\varepsilon}.
\end{aligned}$$

Now, let us define the subset $\text{GOOD}_K \subseteq [K]$ as follows:

$$\text{GOOD}_K \coloneqq \left\{k \;\middle|\; f(k) = x, x \in \text{GOOD}_X\right\}.$$

We then define the operator:

$$\Pi'^{KB} \coloneqq \sum_{k \in \text{GOOD}_K} |k\rangle\langle k|^K \otimes \Pi_{f(k)}^B.$$

Then observe that :

$$\begin{aligned}
\operatorname{Tr}\left[\Pi'^{KB} \sigma^{KB}\right] &= \sum_{k \in \text{GOOD}_K} Q_K(k) \operatorname{Tr}\left[\Pi_{f(k)} \sigma_k^B\right] \\
&= \sum_{x \in \text{GOOD}_X} \sum_{k : f(k) = x} Q_K(k) \operatorname{Tr}\left[\Pi_{f(k)} \sigma_k^B\right] \\
&= \sum_{x \in \text{GOOD}_X} Q_X(x) \operatorname{Tr}\left[\Pi_x \sigma_x^B\right] \\
&\geq (1 - \sqrt{3\varepsilon}) \cdot \Pr_{Q_X}[\text{GOOD}_X] \\
&\geq 1 - \varepsilon^{1/4},
\end{aligned}$$

where the last inequality uses the fact that the probabilities of any set under the distributions P_X and Q_X can differ by at most ε. Again, using Markov's inequality we infer that there exists a subset $\text{NICE}_K \subseteq \text{GOOD}_K$ such that $\Pr_{Q_K}[\text{NICE}_K] \geq 1 - \varepsilon^{1/8}$, and for all $k \in \text{NICE}_K$ it holds that:

$$\operatorname{Tr}\left[\Pi_{f(k)} \sigma_k^B\right] \geq 1 - \varepsilon^{1/8}.$$

This implies that for all $k \in \text{NICE}_K$, $\Pi_k^B \coloneqq \Pi_{f(k)}^B$ is a candidate for optimising the expression $2^{H_H^{\varepsilon^{1/8}}(\sigma_k^B)}$. This implies that, for all $k \in \text{NICE}_K$:

$$\begin{aligned}
2^{H_H^{\varepsilon^{1/8}}(\sigma_k^B)} &\leq \operatorname{Tr}\left[\Pi_k^B\right] \\
&= \operatorname{Tr}\left[\Pi_x^B\right] \quad \text{where } x = f(k) \\
&\leq \frac{2^{H_H^\varepsilon(B|X)}}{\varepsilon} \quad \text{since } x \in \text{GOOD}_X.
\end{aligned}$$

Also note that since Q_K and $\mathbf{Unif}[K]$ are close by δ, it holds that $\Pr_{\mathbf{Unif}[K]}[\text{NICE}_K] \geq 1 - \varepsilon^{1/8} - \delta$. This concludes the proof. □

Proof of Lemma 6.8

Proof. It is not hard to see that the states

$$\sum_{k,\ell} Q_{KL}(k,\ell) \left|k,\ell\right\rangle \left\langle k,\ell\right|^{KL} \otimes \sigma^{RB}_{f(k,\ell)} \tag{Bob and Ref}$$

and

$$\sum_{k,\ell} Q_{KL}(k,\ell) \left|k,\ell\right\rangle \left\langle k,\ell\right|^{KL} \otimes \sigma^{B}_{f(k,\ell)} \tag{Bob}$$

both satisfy the requirements of Lemma A.1. For the state in Bob and Ref, we think of the correspondence $B \leftarrow RB$ and $K \leftarrow KL$ with respect to the registers KB in Lemma A.1. Similarly, for the state in Bob the correspondence is $B \leftarrow B$ and $K \leftarrow KL$. Then note that the closeness of $\sigma^{RB}_{f(k,\ell)}$ and $\rho^{RB}_{f(k,\ell)}$ (for $(k,\ell) \in \text{supp}(Q_{KL})$, implied by Fact 6.6) implies the closeness of the marginals $\sigma^{B}_{f(k,\ell)}$ and $\rho^{B}_{f(k,\ell)}$. Also, since $Q_X(x) \coloneqq \sum_{k,\ell: f(k,\ell)=x} Q_{KL}(k,\ell)$, it holds via Fact 6.5 that $\left\|P_X - Q_X\right\|_1 \leq O(\varepsilon)$. Additionally, it also holds that the distribution Q_{KL} is close to the uniform distribution on $[KL]$ by $O(\varepsilon^{1/2})$, as implied by Fact 6.7.

A subtle issue is that the expression in Fact 6.5 the distribution $Q_{L \mid k}$ is supported on $[L] \bigcup \{\perp\}$ for all k. However, from [5] we know that the mass on this element is at most $O(\varepsilon)$ (for our choice of parameters) and can thus be removed from the 1-norm expression with a penalty of at most $O(\varepsilon)$. This allows us to run the argument above for only $\ell \neq \perp$, and the valifity of the closeness of the state $\sigma^{RB}_{f(k,\ell)}$ and $\rho^{RB}_{f(k,\ell)}$ holds.

We will first instantiate parameters (ε, δ) in Lemma A.1, then use Lemma A.1 twice. To that end, set $\delta \leftarrow O(\varepsilon^{1/2})$ and $\varepsilon \leftarrow O(\varepsilon)$. Then, we first apply Lemma A.1 to the state in Bob and Ref to see that there exists a subset

$$\mathcal{S}_1 \subseteq [K] \times [L]$$

with the property that

$$|\mathcal{S}_1| \geq (1 - O(\varepsilon^{1/8})) \cdot KL$$

and for all $(k,\ell) \in \mathcal{S}_1$, it holds that

$$H_H^{O(\varepsilon^{1/8})}(RB \mid k,\ell) \leq H_H^{O(\varepsilon)}(RB \mid X) + O(\log \frac{1}{\varepsilon}).$$

Similarly, applying Lemma A.1 to Equation Bob, we see that there exists a set $\mathcal{S}_2 \subseteq [K] \times [L]$ such that

$$|\mathcal{S}_2| \geq (1 - O(\varepsilon^{1/8}))KL$$

and for all $(k,\ell) \in \mathcal{S}_2$ it holds that

$$H_H^{O(\varepsilon^{1/8})}(B \mid k,\ell) \leq H_H^{O(\varepsilon)}(B \mid X) + O(\log \frac{1}{\varepsilon}).$$

It holds then that for all (k,ℓ) in the set

$$\mathcal{S} \coloneqq \mathcal{S}_1 \bigcap \mathcal{S}_2,$$

where

$$|\mathcal{S}| \geq (1 - O(\varepsilon^{1/8}))KL,$$

it holds that

$$H_H^{O(\varepsilon^{1/8})}(RB \mid k,\ell) \leq H_H^{O(\varepsilon)}(RB \mid X) + O(\log \frac{1}{\varepsilon})$$

and

$$H_H^{O(\varepsilon^{1/8})}(B \mid k,\ell) \leq H_H^{O(\varepsilon)}(B \mid X) + O(\log \frac{1}{\varepsilon}).$$

This concludes the proof. □

Appendix B Proofs of Lemma 2.24 and 2.25

Proof of Lemma 2.24

Proof. Suppose Π_{OPT}^{XB} is the optimising operator that for the quantity $H_H^\varepsilon(B \mid X)$. Without loss of generality we can assume that Π_{OPT} is of the following form:

$$\Pi_{\text{OPT}}^{XB} = \sum_x |x\rangle\langle x|^X \otimes \Pi_x^B.$$

Then by definition Π_{OPT} satisfies the following optimisation problem:

$$\min_{\{\Pi_x\}_x \,:\, 0\leq\Pi_x\leq\mathbb{I}} \sum_x P_X(x) \operatorname{Tr}\left[\Pi_x\right]$$
$$\sum_x P_X(x) \operatorname{Tr}\left[\Pi_x |v_x\rangle\langle v_x|\right] \geq 1-\varepsilon.$$

Let us define for every $x \in \mathcal{X}$ an operator:

$$\Pi'_x := \operatorname{Tr}\left[\Pi_x |v_x\rangle\langle v_x|\right] |v_x\rangle\langle v_x|,$$

and

$$\Pi'^{XB} =:= \sum_x |x\rangle\langle x|^X \otimes \Pi'^B_x.$$

It is clear that $\operatorname{Tr}\left[\Pi'^{XB}\rho^{XB}\right] \geq 1-\varepsilon$ and that $\sum_x P_X(x) \operatorname{Tr}\left[\Pi'^{XB}\right] \leq 1$. This implies that Π'^{XB} is a candidate optimiser for $H_H^\varepsilon(B \mid X)$ and thus:

$$H_H^\varepsilon(B \mid X) \leq 0.$$

This concludes the proof. □

Proof of Lemma 2.25

Proof. We can assume without loss of generality that $|A| \leq |B|$ and that the optimising operator Π_{OPT}^{XA} is of the form:

$$\Pi_{\text{OPT}}^{XA} = \sum_x |x\rangle\langle x|^X \otimes \Pi_x^A.$$

Let us fix $x \in \mathcal{X}$. Let the corresponding $|v_x\rangle^{AB}$ have the following Schmidt decomposition:

$$|v_x\rangle^{AB} = \sum_i \lambda_i |a_i\rangle^A |b_i\rangle^B.$$

Let $V_x^{A\to B}$ be an isometry defined by $|a_i\rangle^A \to |b_i\rangle^B$ for all i. Then, consider the following:

$$\begin{aligned}
&\operatorname{Tr}\left[\Pi_{\text{OPT}}^{XB}\rho^{XAB}\right] \\
=&\sum_x P_X(x) \operatorname{Tr}\left[\Pi_x^A |v_x\rangle\langle v_x|^{AB}\right] \\
=&\sum_x P_X(x) \sum_i \lambda_i^2 \operatorname{Tr}\left[\Pi_x^A |a_i\rangle\langle a_i|^A\right] \\
=&\sum_x P_X(x) \sum_i \lambda_i^2 \operatorname{Tr}\left[\Pi_x^A V_x^\dagger |b_i\rangle\langle b_i|^B V_x\right] \\
=&\sum_x P_X(x) \sum_i \lambda_i^2 \operatorname{Tr}\left[\left(V_x \Pi_x^A V_x^\dagger\right)^B |b_i\rangle\langle b_i|^B\right] \\
=&\sum_x P_X(x) \operatorname{Tr}\left[\left(V_x \Pi_x^A V_x^\dagger\right)^B |v_x\rangle\langle v_x|^{AB}\right].
\end{aligned}$$

Therefore, we can define an operator $\widetilde{\Pi}^{XB} = \sum_x P_X(x) \left|x\right\rangle\left\langle x\right| \otimes \widetilde{\Pi}_x^B$ where for each $x \in \mathcal{X}$, $\widetilde{\Pi}_x^B \coloneqq \left(V_x \Pi_x^A V_x^\dagger\right)^B$, such that:

$$\begin{aligned} &\operatorname{Tr}\left[\widetilde{\Pi}^{XB} \rho^{XB} \rho^{XB}\right] \\ =&\operatorname{Tr}\left[\Pi_{\text{OPT}}^{XA} \rho^{XA} \rho^{XA}\right] \\ \geq& 1 - \varepsilon. \end{aligned}$$

Therefore, $\widetilde{\Pi}^{XB}$ is a candidate optimiser for the quantity $H_H^\varepsilon(B \mid X)$, which implies that:

$$H_H^\varepsilon(B \mid X) \leq H_H^\varepsilon(A \mid X).$$

A similar analysis shows that:

$$H_H^\varepsilon(A \mid X) \leq H_H^\varepsilon(B \mid X).$$

This concludes the proof. □

Appendix C Proofs of Lemma 7.2 and Claim 7.3

C.1 Proof of Lemma 7.2

Proof. From Lemma 6.8, we know that the entropic inequalities in the statement of the lemma hold for at least $(1 - O(\varepsilon^{1/8}))$ fraction of all index pairs (k, ℓ). Define $\mathbb{1}_{k,\ell}$ as the indicator that the entropic inequalities hold for the fixed index pair (k, ℓ). Then,

$$\sum_{k,\ell} \frac{1}{KL} \mathbb{1}_{k,\ell} \geq (1 - O(\varepsilon^{1/8})).$$

Define

$$\texttt{prob}_k \coloneqq \sum_{k,\ell} \frac{1}{L} \mathbb{1}_{k,\ell}$$

Then, it holds by Markov's inequality that for $(1 - \varepsilon^{1/16})$ fraction of k's,

$$\text{PROB}_k \geq 1 - \varepsilon^{1/16}.$$

We define the set $\mathcal{T}'$ to be that set of k's where the above condition holds. Let $k \in \mathcal{T}'$. Then by the fact that $\texttt{prob}_k$ is an average of indicator functions, we can conclude that, for at least $1 - \varepsilon^{1/16}$ fraction of ℓ's in $[L]$, it holds that

$$\mathbb{1}_{k,\ell} = 1.$$

We define the set where the above condition holds to be $\text{NICE}_{L \mid k}$. This concludes the proof. □

C.2 Proof of Claim 7.3

Proof. Recall that Corollary 7.2 shows the existence of the set $\mathcal{T}' \subseteq [K]$ of size at least $(1 - \varepsilon^{1/16})K$. On the other hand, Lemma 6.11 shows that the set $\mathcal{T}$ is of size at least $(1 - O(\varepsilon^{1/32}))K$. This implies that $|\mathcal{T} \bigcap \mathcal{T}'| > 0$. This concludes the proof. □