One-Shot Non-Catalytic Distributed Purity Distillation

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

[1] Centre for Quantum Technologies, National University of Singapore
[2] Tata Institute of Fundamental Research, Mumbai

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 [11], 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, 13] proves 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. The problem first appeared in the works [12, 7, 8], who considered a local and a distributed version of the problem. Informally, in the local purity distillation scenario, one is provided a state ρ and only allowed local unitary operations. One can of course also discard, i.e., trace out certain subsystems which they already possess. The goal is to maximise the number of pure qubits that can be extracted given these resources. In the distributed version of the problem, there are two parties Alice and Bob who each possess one part of a bipartite state. They are allowed local unitary operations and may communicate using a completely dephasing channel. This paradigm is termed CLOCC (*closed* local operations and classical communication), is a modification of the more familiar LOCC paradigm, which does not allow use of pure ancilla qubits to implement the local operations. The works mentioned above provided some preliminary bounds for the distributed purity distillation problem in this regime. A tight lower bound for the local distillation problem was provided in [9].

Aside from the preliminary works mentioned above, the first detailed treatment for the purity distillation problem appeared in the work of Devetak [5]. In an important relaxation of the CLOCC paradigm, Devetak considered the problem of distributed purity distillation, in the CLOCC$'$ paradigm, where 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, with one-way communication. In particular, Devetak showed that, given n iid copies of a bipartite state ρ^{AB}, where A and B are shared between two parties, it is possible to distill pure qubits at (roughly) a rate:

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

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.

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* correlations between the two systems A and B (see [6] for more details). These classical 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. However, one may ask whether it is possible to design a purity distillation protocol, in the CLOCC′ paradigm with one-way communication, where one uses an *arbitrary* POVM, at the cost of a lower rate of distilled pure qubit states. This question was answered in the paper by Krovi and Devetak [10], where not only did the authors provide a protocol which can use any POVM, but also simplified Devetak's original protocol and proof to a large extent.

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 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 [5] original techniques to the one-shot setting, it is not immediately clear how one can adapt them to the case of general POVMs.

Our Contribution

In this paper we present a one-shot protocol for distributed purity distillation which has two important properties:

1. It is much simpler than the original proof techniques used in [3] and can be used to describe a general distributed distillation protocol which uses any arbitrary POVM, along the lines of Krovi and Devetak [10].

2. The second and perhaps most important feature of our protocol, is that we improve upon the number of ancilla qubits used by the Krovi-Devetak protocol. Indeed, we show that in the one-shot setting, for most instances of the underlying state and choice of POVM, the number of ancilla qubits required is 0, barring a few pathological cases. We show that for these pathological cases, the number ancilla qubits required is very small, and goes to 0 in the asymptotic iid limit.

3. We present one-shot upper bounds which show that all our protocols are almost optimal. Such upper bounds were not known for the one-shot setting prior to this work.

Note that reducing the number of ancilla qubits used for the original asymptotic iid Krovi-Devetak protocol is possible in certain situations. 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. This is of course assuming that the party who enacts the POVM on their state has some positive rate of producing pure qubits.

Of course, these ideas completely fail in the one-shot setting where only *one* copy of the underlying state is available, and one cannot do any bootstrapping. Our protocol wins over the original Krovi-Devetak protocol in this sense, since it requires very little to no ancilla to run even in the one-shot setting. Also, our protocol requires 0 amount of ancilla in the asymptotic iid limit, for all setting of the underlying state and the chosen POVM.

We divide the paper into four parts: Section 2 contains important definitions of the information theoretic tasks that we consider as well as the definitions of important one-shot quantities. Section 3 contains the description of the locally optimal protocol for local purity distillation. Section 4 contains the description of a naïve protocol which nonetheless helps in highlighting some of the ideas that we will use later. Section 5 contains the one-shot version of the Krovi-Devetak protocol, which works with ancilla qubits. Finally, we present the protocol which uses very little to no ancilla in Section 6.

2 Definitions and Relevant Quantities

2.1 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 2.1 (**Smoothed Support Max Entropy**). *Given a quantum state ρ^A, let us denote its eigenvalues by $\lambda_1, \ldots \lambda_{|\text{supp}(\rho)|}$ (in ascending order) corresponding to the eigenvectors $v_1, \ldots, v_{|\text{supp}(\rho)|}$. Let $\lambda_1, \ldots \lambda_k$ denote the smallest eigenvalues such that $\sum_i \lambda_i \leq \varepsilon$. Then, we define the* **smoothed support max entropy** *of ρ^A as*

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

Definition 2.2 (**Smoothed Norm Max Entropy**). *Given the setup of in Definition 2.1, we define the* **smoothed norm max entropy** *of the state ρ^A as*

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

Definition 2.3 (**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:

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

Definition 2.4. (Conditional Smooth Max Entropy) *Given a bipartite quantum state ρ^{AB}, we define the* **smooth max entropy** *as*

$$H^{\varepsilon}_{\max}(A|B)_\rho \coloneqq \min_{\rho'\in\mathcal{B}^{\varepsilon}(\rho)} \max_{\sigma^B} 2\log F\left(\rho'^{AB}, \mathbb{1}^A \otimes \sigma^B\right)$$

Definition 2.5. (Conditional Smooth Min Entropy) *Given a bipartite quantum state ρ^{AB}, the* **conditional smooth min entropy** *is defined as:*

$$H^{\varepsilon}_{\min}(A|B)_\rho \coloneqq -\log\min\left\{\text{Tr}\left[\sigma^B\right] \mid \sigma^B \geq 0, \rho^{AB} \leq \mathbb{I}^A \otimes \sigma^B\right\}.$$

2.2 Some Properties of One-Shot Quantities

Lemma 2.6. *Given a state ρ^A and an arbitrary purification $|\rho\rangle^{RA}$, it holds that*

$$H^{\varepsilon}_H(A) = H^{\varepsilon}_H(R).$$

Proof. We will first show that we can assume that the optimising operator in the definition of $H^{\varepsilon}_H(A)$ commutes with ρ. Let this operator be Π. Suppose that

$$\rho^A = \sum_x \lambda_x |x\rangle\langle x|.$$

Then,

$$\text{Tr}\left[\Pi\rho\right] = \sum_x \lambda_x \langle x|\Pi|x\rangle$$
$$\text{Tr}\left[\Pi\right] = \sum_x \langle x|\Pi|x\rangle$$

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:

$$\widetilde{\Pi} \coloneqq \sum_{|x\rangle\langle x|\in\text{supp}(\rho)} \langle x|\Pi|x\rangle |x\rangle\langle x|$$

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, consider the Schmidt decomposition of $|\rho\rangle^{AR}$:

$$|\rho^{AR}\rangle = \sum \sqrt{\lambda_x}\, |x\rangle^A |\zeta_x\rangle^R.$$

We wish to compute the quantity $H_H^\varepsilon(R)$. As before, we can assume that the optimising operator commutes with ρ^R, i.e., it diagonalises in the basis $\left\{|\zeta_x\rangle^A\right\}$ 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:

$$\Sigma^R = \sum_{\zeta_x \in \text{supp}(\rho^R)} \alpha_x \left|\zeta_x\right\rangle \left\langle\zeta_x\right|$$

Finally, consider the following LP which is the definition of $H_H^\varepsilon(R)$:

$$\min \sum_x \alpha_x$$

$$\sum_x \alpha_x \lambda_x \geq 1 - \varepsilon$$
$$0 \leq \alpha_x \leq 1$$

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

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

This concludes the proof. □

Fact 2.7 (cite NemaSen)**.** *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 2.8 (Equivalence of the Smoothed Norm Max and Smooth Hypothesis Testing Entropies). *For any quantum state ρ^A it holds that*

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

Proof. To prove this lemma, we first observe that without loss of generality we can assume that the optimising operator for $H_H^\varepsilon(A)$ diagonalises in the same basis as ρ^A. This immediately implies the upper bound since we obtain $\widetilde{H}_{\max}^\varepsilon(A)$ 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 $H_H^\varepsilon(A)$ diagonalises in the same basis as ρ^A, once can assume that the following holds for all such candidate operators Π^A:

$$\rho^A = \sum_a P_A(a) \left|a\right\rangle \left\langle a\right|^A$$
$$\Pi^A = \sum_a \lambda(a) \left|a\right\rangle \left\langle a\right|^A$$

Then, it holds that the problem of finding $H_H^\varepsilon(\varepsilon)(A)$ can be reduced to solving the following LP:

$$\min \sum_a \lambda(a)$$
$$\sum_a P_A(a)\lambda(a) \geq 1 - \varepsilon$$

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

Lemma 2.9. (Subadditivity of the Smooth Hypothesis Testing Entropy) *Given a bipartite quantum state ρ^{AB}, it holds that*

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

Proof. To see that this holds, let Π^A and Π^B be the optimising operators for $H_H^\varepsilon(A)$ and $H_H^\varepsilon(B)$ 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

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

Therefore we can conclude that,

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

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

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

This concludes the proof. □

Lemma 2.10. *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 $H_H^\varepsilon(A)$. Since $|0\rangle\langle 0|$ has trace 1, the result follows. This concludes the proof. □

Lemma 2.11. *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*

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

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

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

By definition, this operator has the property that:

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

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

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

Again, by definition, it holds that:

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

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:

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

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. □

2.3 Definitions of Local and Distributed Purity Distillation

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

Definition 2.12. (Local Purity Distillation) *Given a quantum state ρ^A we define a ε local purity distillation code as a unitary map*

$$U : A \to A_p A_g$$

such that

$$\left\| \operatorname{Tr}_{A_g}\left[U \cdot \rho\right] - |0\rangle\langle 0|^{A_p} \right\|_1 \leq \varepsilon$$

The rate of the code is given by

$$R_{\text{local}}^\varepsilon \coloneqq \log |A_p|$$

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 sucht that

$$R_{\text{local}}^\varepsilon = R$$

The supremum over all achievable rates is defined to be the *purity* of the state ρ^A and is defined as

$$\kappa_\varepsilon(\rho^A) \coloneqq \sup R_{\text{local}}^\varepsilon$$

Definition 2.13. (Distributed Purity Distillation) *Given a bipartite quantum state ρ^{AB} to two parties Alice and Bob, a distributed purity distillation protocol with error ε is a protocol consisting of :*

1. *A local unitary $U_{\text{ALICE}}^{AC\to A_p A_g X_A}$ which Alice can apply to her local register A and ancilla C.*
2. *A completely dephasing channel $\mathcal{P}^{X_A \to X_B}$, where X_B is a classical register belonging to Bob.*
3. *A local unitary $U_{\text{ALICE}}^{AC\to A_p A_g X_A}$ which Bob can apply to his local registers $X_B B$.*

such that, at the end of the protocol, the following condition holds:

$$\left\| \operatorname{Tr}_{A_g B_g}\left[V \circ \mathcal{P} \circ U\left(\rho^{AB} \otimes |0\rangle\langle 0|^C\right)\right] - |0\rangle\langle 0|^{A_p} \otimes |0\rangle\langle 0|^{B_p} \right\|_1 \leq \varepsilon$$

The rate of the protocol is defined as

$$R_{\text{dist}}^\varepsilon \coloneqq \log |A_p| + \log |B_p| - \log |C|$$

A rate R is said to be achievable with error ε for distributed purity distillation if there exists a distributed purity distillation protocol with error ε for that state. We define the ε-distributed purity of the state ρ^{AB} as the supremum over all achievable rates:

$$\kappa_{\text{dist}}^\varepsilon(\rho^{AB}) \coloneqq \sup R_{\text{dist}}^\varepsilon$$

3 Optimal Protocols for Local Purity Distillation

In this section we show that the protocol for local purity distillation presented in [3] is optimal. To that end, we will require the following lemma:

Lemma 3.1. (Upper Bound for Purity of a State) *Given a a quantum state ρ^A and error $\varepsilon > 0$, it holds that*

$$\kappa_\varepsilon(\rho^A) \leq \log|A| - H_H^{3\sqrt{\varepsilon}}(A)_\rho - \log\varepsilon$$

Proof. Recall that any ε local purity distillation code consists of a unitary operator

$$U : A \to A_p A_g$$

such that

$$\left\| \mathrm{Tr}_{A_g}\left[U \cdot \rho\right] - |0\rangle\langle 0|^{A_p} \right\|_1 \leq \varepsilon$$

This implies that

$$\log|A| = \log|A_p| + \log|A_g|$$

We will lower bound $\log|A_g|$ which in turn will give us an upper bound for $\log|A_p|$. To that end note the following sequence of inequalities

$$\begin{aligned}
\log|A_g| &\geq H_H^\varepsilon(A_g) + \log\varepsilon \\
&\geq H_H^{3\sqrt{\varepsilon}}(A_p A_g) - H_H^\varepsilon(A_p) + \log\varepsilon \\
&= H_H^{3\sqrt{\varepsilon}}(A) + \log\varepsilon
\end{aligned}$$

The first inequality uses the fact that

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

The second inequality uses the subadditivity of the smooth hypothesis testing entropy. The last equality uses two facts:

1. $H_H^{3\sqrt{\varepsilon}}(A_p A_g) = H^{3\sqrt{\varepsilon}}(A)$ by the unitary invariance of the smooth hypothesis testing entropy.
2. Since, by assumption,
$$\left\| \mathrm{Tr}_{A_g}\left[U \cdot \rho\right] - |0\rangle\langle 0|^{A_p} \right\|_1 \leq \varepsilon$$
the entropy
$$H_H^\varepsilon(A_p)_{\mathrm{Tr}_{A_g}[U\cdot\rho]} = 0$$

Collating the above arguments we see that, for any ε local purity distillation protocol,

$$\log|A_p| \leq \log|A| - H_H^{3\sqrt{\varepsilon}}(A) - \log\varepsilon$$

This concludes the proof.

□

The following fact can be proved using a straightforward application of Theorem 1.7, [3] and Fact 2.8:

Fact 3.2. Lower Bound for Purity of a State *Given a quantum state ρ^A, there exists an ε purity distillation code with rate*

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

Lemma 3.1 and Fact 3.2 together imply the following theorem:

Theorem 3.3. *Given a quantum state ρ^A and $\varepsilon > 0$, the purity of the state $\kappa_\varepsilon(\rho^A)$ satisfies the following bounds:*

$$\log|A| - H_H^{\varepsilon^2/9}(A)_\rho + O(\log\varepsilon) - 1 \leq \kappa_\varepsilon(\rho^A) \leq \log|A| - H_H^{3\sqrt{\varepsilon}}(A)_\rho - \log\varepsilon$$

In the following section, we will thus refer to the *locally optimal* protocol, in reference to Theorem 3.3.

4 A First Protocol and 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 distills, but nevertheless helps in understanding some of the key ideas that lead to the other optimal protocol construction that follow in later sections. To setup the protocol, we recall the setup of the distributed purity distillation problem:

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 $\left\{\Lambda_x^{A\to X}\right\}$ which has outcomes x from the set of set of symbols $\mathcal{X}$.

2. Alice can borrow any number of qubits 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 distills on her end.

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 is 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 distills.

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 then goes as follows:

Protocol A

1. Consider a purification $|\rho\rangle^{ABR}$ of the state ρ^{AB}.

2. Alice first acts the POVM $\Lambda^{A\to X}$ on the system A coherently, by borrowing $\log|\mathcal{X}|$ number of ancilla qubits. To be precise, consider the isometry:
$$\sum_{x\in\mathcal{X}} |x\rangle^{X_A} \sqrt{\Lambda_x^A}.$$
This isometry can always be extended to a full unitary on the system $X_A A$, where we identify X with X_A. The global state after this stage of the protocol is then
$$\sum_{x\in\mathcal{X}} |x\rangle^{X_A} \sqrt{\Lambda_x^A}\, |\rho\rangle^{ABR}.$$

3. Define the state:
$$|\rho_x\rangle^{ABR} := \frac{1}{\sqrt{\mathrm{Tr}\left[\Lambda_x^A \rho^A\right]}} \sqrt{\Lambda_x} \, |\rho\rangle^{ABR}$$
and its marginals ρ_x^A and ρ_x^B appropriately. Let $U_x^{A\to A_p A_g}$ be the unitary which performs the locally optimal protocol on ρ_x^A.

4. Alice performs the controlled unitary
$$\sum_{x\in\mathcal{X}} |x\rangle\langle x|^{X_A} \otimes U_x^{A\to A_p A_g}$$
on her systems $X_A A$.

5. Alice then sends the system X_A through the completely dephasing channel $\mathcal{P}$ to Bob.

6. At this point, the state on the systems $X_B B$ that are in Bob's possession look like:
$$\sum_x P_X(x) \, |x\rangle\langle x|^{X_B} \otimes \rho_x^B.$$

7. Define the unitary $V_x^{B\to B_p B_g}$ to be the unitary which performs the locally optimal distillation protocol on the state ρ_x^B. Bob then uses the controlled unitary:
$$\sum_x |x\rangle\langle x|^{X_B} \otimes V_x^{B\to B_p B_g}$$
on his systems $X_B B$.

—-x—-

Let us analyse Protocol A. We claim the following proposition:

Proposition 4.1. *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:
$$\langle s|s'\rangle^{A_g} = \langle s|s'\rangle^A.$$
We can then write:
$$U_x^{A\to A_p A_g} |\rho_x\rangle^{ARB} = \sum_{s:\lambda_s\notin \text{BAD}} \lambda_s \, |0\rangle^{A_p} |s\rangle^{A_g} |s\rangle^{RB} + |\text{JUNK}\rangle^{A_p A_g RB}.$$

It is then not hard to see that

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

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

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

Now, consider the cq state:

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

where $P_X(x)$ is the probability of the outcome x when ρ^A is measured with the POVM Λ. Then, using Lemma 2.11 we see that there exists a subset of x's, which we call $\mathcal{S}_{\text{ALICE}}$ such that

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

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:

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

where the system A_p is constituted by $H_H^{\varepsilon^2}(A|X) - \log\varepsilon$ qubits. Another applpication of Lemma 2.11 shows us that there exists a set $\mathcal{S}_{\text{BOB}}$ such that

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

where the entropic quantities in the expression above are computed with respect to the cq state $\sum_x P_X(x) |x\rangle\langle x|^{X_B} \otimes \rho_x^B$. 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:

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

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:

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

This concludes the proof. □

As mentioned earlier, the number of pure qubit states that Protocol A distills 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 distill $\log|A| - H_H^\varepsilon(A|X)$ 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.

5 An Optimal Protocol with Ancilla

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 distills. However, an issue with this strategy is that this new POVM may not allow Alice and Bob to individually distill $\log|A| - H_H^\varepsilon(A|X)$ and $\log|B| - H_H^\varepsilon(B|X)$ pure qubits. The measurement compression theorem comes to our aid here. We take a small detour from our exposition to state the measurement compression theorem.

5.1 Measurement Compression

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 $|\rho\rangle^{ABR}$. It can be shown (see [14]) that the global state, after the action of the POVM on the system A, looks like

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

where

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

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 $k \overset{R}{\leftarrow} [K]$.
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 [4]:

Fact 5.1. *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

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

there exist POVMs $\Theta^A(1), \Theta^A(2), \ldots \Theta^A(K)$, where each POVM Θ^A_k 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

$$\left\| \rho^{XBR} - \sum_x \sum_{k,\ell} Q_{KL}(k,\ell) \cdot \mathtt{1}_{f(k,\ell)=x} \left|x\right\rangle \left\langle x\right|^X \otimes \sigma^{RB}_{f(k,\ell)} \right\|_1 \leq c_3 \varepsilon^{1/4}$$

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

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

We will make us of another useful fact about measurement compression, specifically that the distribution Q_{KL} is close to the uniform distribution on $[K] \times [L]$. We state this fact below:

Fact 5.2. *Given the setup of Fact 5.1, it holds that*

$$\|Q_{KL} - \mathbf{Unif}[K] \times \mathbf{Unif}[L]\|_1 \leq O(\varepsilon^{1/8}).$$

5.2 The Protocol

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 §. Thus, one would expect, via similar reasoning as that which we used to prove Lemma 2.11, 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 5.3. *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

$$H_H^{O(\varepsilon^{1/8})}(RB \mid k, \ell) \leq H_H^{\varepsilon^{1/4}}(RB \mid X) + O(\log \frac{1}{\varepsilon}) + O(1)$$

and

$$H_H^{O(\varepsilon^{1/8})}(B \mid k, \ell) \leq H_H^{\varepsilon^{1/4}}(B \mid X) + O(\log \frac{1}{\varepsilon}) + O(1).$$

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 5.1:

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

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

   .

3. Alice possesses the POVMs $\Theta^A(1), \Theta^A(2), \ldots, \Theta^A(K)$ whose existence is implied by Fact 5.1. Each of these POVMs produces outputs in the set $[L] \bigcup \{\perp\}$.

4. We will overload notation and use the letter K to denote a public coin register that is available to both Alice and Bob. 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 5.1.

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.

Protocol B

1. Alice borrows $I_{\max}^{\varepsilon}(X : RB) + O(\log \frac{1}{\varepsilon}) + o(1)$ number of qubits to store the measurement result. These will constitute the system L_A.

2. Alice will apply the isometry

$$\sum_k |k\rangle \langle k|^K \otimes \sum_\ell |\ell\rangle^{L_A} \sqrt{\Theta_\ell^A(k)}$$

   to the system A. In actuality Alice is only allowed to use unitaries on her side. However, it is always possible to complete the description of the above isometry to a full unitary.

3. Alice then applies the unitary operator

$$\sum_{\ell} |\ell\rangle\langle\ell|^{L_A} \otimes U_{k,\ell}^{A\to A_p A_g}$$

to her systems $L_A A$, where the each unitary $U_{k,\ell}$ enacts the locally optimal distillation protocol on the state $\rho_{k,\ell}^A$.

4. Alice sends the system L_A through the completely dephasing channel $\mathcal{P}^{L_A\to L_B}$.

5. After receiving Alice's classical message, Bob simply conditions on the contents of the register K and his register L_B to perform the locally optimal protocol on his system B.

—-x—-

Proposition 5.4. *Protocol B distills*

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

number of pure qubits with error $O(\varepsilon^{1/16})$. 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. Using the same arguments as we saw in the proof of Proposition 4.1, we can show that

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

We will now invoke Lemma 5.3 to note that, for at least $(1-\varepsilon^{1/8})$ fraction of indices KL, it holds that

$$H_H^{O(\varepsilon^{1/8})}(\rho_{k,\ell}^{RB}) \le H_H^{\varepsilon^{1/4}}(RB|X) + O(\log\frac{1}{\varepsilon}) + O(1).$$

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

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

amount of purity. Thus, using the fact that for pure states $|\rho_{k,\ell}\rangle^{ARB}$

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

(see Lemma 2.6) we can conclude that the global state after Alice sends the system L_A through the dephasing channel satisfies the following condition:

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

where we have used the fact that the distribution Q_{KL} is $O(\varepsilon^{1/4})$ to the uniform distribution on $[K]\times[L]$ (see Fact 5.2). Thus, on her side, Alice distills at least

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

amount of purity.

To analyse Bob's actions, we again invoke Lemma 5.3 and recall that, for at least $1 - O(\varepsilon^{1/8})$ fraction of indices KL, it holds that

$$H_H^{O(\varepsilon^{1/8})}(\rho^B_{k,\ell}) \leq H_H^{\varepsilon^{1/4}}(B \mid X) + O(\log \frac{1}{\varepsilon}) + O(1).$$

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

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

where we see that

$$|B_p| \geq |B| - H_H^{\varepsilon^{1/4}}(B|X) + O(\log \varepsilon) - O(1).$$

Then, using the fact that the distribution Q_{KL} is $O(\varepsilon^{1/4})$ close to the uniform distribution on $[K] \times [L]$ and the arguments we used for Alice's actions, we see that the following holds:

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

Tracing out all registers but the systems $A_p B_p$ implies the result, where we see that the *net* number of pure qubits that the protocol distilled is given by:

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

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. This concludes the proof. □

5.3 Removing the Public Coin from Protocol B

In this section we derandomise Protocol B by removing the public coin register K. We show this in the following lemma:

Lemma 5.5. *Given the setting of Proposition 5.4, there exists at least one $k \in [K]$, such that if Alice runs Protocol B with only the POVM corresponding to this k, the resulting protocol, called Protocol C, distills 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:

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

To derandomise the above protocol, we define

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

Then, using block diagonality, we see that

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

This immediately proves that there exists a k such that if we run the protocol for only that fixed k, Alice and Bob distill the same amount of purity as in the protocol with shared randomness, while making an error at most $O(\varepsilon^{1/16})$. In fact, since Q_K is $O(\varepsilon^{1/4})$ 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. □

5.4 Almost Optimality of Protocol B

In this section we prove a one-shot upper bound on the number of qubit states that Alice and Bob can hope to distill, given the setting of the distributed purity distillation problem. Our bounds show that Protocol B is *almost* optimal. We will require the following lemma:

Lemma 5.6. *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*

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

while making an error of at most ε, and

Proof. By the definition of a distributed purity distillation protocol, it holds that

$$\log|A_pB_p| - \log|C| = \log|AB| - \log|A_gB_g|$$

We will lower bound $\log|A_gB_g|$ which will in turn allow us to upper bound $\log|A_pB_p| - \log|C|$. 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_BB after Alice sends the contents of the register X_A through the channel is a cq state, 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 afct that both $H_H^{\varepsilon}(A_p)$ and $H_H^{\varepsilon}(B_p)$ are 0. Thus, LHS is

$$\begin{aligned}&= H_H^{3\sqrt{\varepsilon}}(A_pA_g) + H_H^{3\sqrt{\varepsilon}}(BX_B) + 2\log\varepsilon\\ &\geq H_{\max}^{3\sqrt{\varepsilon}}(A_pA_g) + H_{\max}^{3\sqrt{\varepsilon}}(X_B) + H_{\max}^{3\sqrt{\varepsilon}}(BX_B) - H_{\max}^{3\sqrt{\varepsilon}}(X_B) + 2\log\varepsilon\\ &\geq H_{\max}^{3\sqrt{\varepsilon}}(A_pA_g) + H_{\max}^{3\sqrt{\varepsilon}}(X_A) + H_{\min}^{f(\varepsilon)}(B\mid X_B) + 2\log\varepsilon\end{aligned}$$

In the last inequality 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 cite Marco thesis. We have also used the chain rules for smooth min and max entropies proved in cite Dupuis. Next, we will use the subadditivity of the max entropy to see that the LHS is:

$$\begin{aligned}&\geq H_{\max}^{g(\varepsilon)}(A_pA_gX_A) + H_{\min}^{f(\varepsilon)}(B\mid X_B) + 2\log\varepsilon\\ &= H_{\max}^{g(\varepsilon)}(AC) + H_{\min}^{f(\varepsilon)}(B\mid X_B) + 2\log\varepsilon\\ &= H_{\max}^{g(\varepsilon)}(A) + H_{\min}^{f(\varepsilon)}(B\mid X_B) + 2\log\varepsilon.\end{aligned}$$

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

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

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 5.7. (Upper Bound for Distributed Purity of a State) *Given a quantum state ρ^{AB}, the ε-distributed purity of this state $\kappa^{\varepsilon}_{\text{dist}}(\rho^{AB})$ is at most*

$$\kappa^{\varepsilon}_{\text{dist}}(\rho^{AB}) \leq \log|A| + \log|B| - H^{g(\varepsilon)}_{\max}(A) - \min_{\Lambda:A\to X} H^{f(\varepsilon)}_{\min}(B \mid X)_{\Lambda^A\otimes\mathbb{I}^B(\rho^{AB})} + 2\log\varepsilon$$

where $\Lambda : A \to X$ is a POVM with outcomes in the set $\mathcal{X}$.

Proof. From Lemma 5.6, 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

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

Recall that we obtained the system X_B by:

1. Using the local unitary $U^{AC\to A_pA_gX_A}_{\text{ALICE}}$ to map the systems AC to $A_pA_gX_A$. We can view the the action of this unitary as that of a CPTP map from the system $A \to X_A$ by hardcoding the ancilla C and tracing out the systems A_pA_g.
2. Sending X_A through the completely dephasing channel $\mathcal{P}$.

It is clear that we can concatenate the two operations described above to define a POVM:

$$\begin{aligned}\Lambda &: A \to X_B \\ &: \rho^A \to \mathcal{P} \circ \text{Tr}_{A_pA_g} \circ U_{\text{ALICE}}(\rho^A)\end{aligned}$$

Renaming X_B to X and taking the supremum over all achievable rates $R^{\varepsilon}_{\text{dist}}$ we see that

$$\kappa^{\varepsilon}_{\text{dist}}(\rho^{AB}) \leq \log|A| + \log|B| - H^{g(\varepsilon)}_{\max}(A) - \min_{\Lambda:A\to X} H^{f(\varepsilon)}_{\min}(B \mid X)_{\Lambda^A\otimes\mathbb{I}^B(\rho^{AB})} + 2\log\varepsilon$$

This concludes the proof. □

6 The Protocol Without Ancilla

In this section we will show how one can implement a distributed distillation protocol while using little to no ancilla qubits. To do this, let us start by examining Alice's actions in Protocol C, as described in the last section (see Lemma 5.5). 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 $I^{\varepsilon}_{\max}(X:RB)$ qubits, where the entropic quantity is computed with respect to a reference 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 distills 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 there exists a setting of the public coin for which the corresponding POVM $\Theta^A(k)$ does as well as the randomised protocol. Since we will be using the POVM corresponding to this setting of the public coin in the following sections, we will refer to this POVM simply as Θ^A.

Our goal in this section will be to implement the action of Θ^A *in place*, that is, by borrowing little to no ancilla qubits. To do this, we will need to make a couple of assumptions on the POVM that Alice used to do the measurement:

1. The number of outcomes of the POVM is at most $I^{\varepsilon}_{\max}(X : RB) - O(\log \varepsilon)$.
2. Barring the POVM element corresponding to the failure event $\perp$, all other POVM elements have rank at most $\frac{1}{\varepsilon} 2^{H_H^{\varepsilon^2}(RB|X)}$.

Although Θ^A satisfies the first assumption by design, in general it will *not* satisfy the second assumption. Therefore, we will first show that Θ^A can be perturbed slightly, such that it satisfies both assumptions above. Let us call this perturbed POVM $\widetilde{\Theta}^A$. However, recall that after she is done with her portion of the protocol, Alice sends her L_A system through the dephasing channel to Bob. For Bob to extract the maximal amount of pure qubits at his end, it is then necessary that the post measurement states created by Θ^A and $\widetilde{\Theta}^A$ be close. We show that there exists a perturbed POVM $\widetilde{\Theta}^A$ which satisfies all of these requirements in the lemma below:

Lemma 6.1. *Consider the setting of Lemma 5.5. Suppose the the POVM which is provided to Alice by Lemma 5.5 is called Θ^A. Then, there exists a POVM $\widetilde{\Theta}^A$ such that:*

1. *Aside from the element corresponding to the failure outcome $\perp$, all other elements of $\widetilde{\Theta}^A$ have rank at most $\frac{1}{\varepsilon} 2^{H_H^{\varepsilon^2}(RB|X)}$.*
2. *The states after measuring the A system ρ^{AB} with Θ^A and $\widetilde{\Theta}^A$ are close, i.e.:*

$$\left\| \sum_{\ell} |\ell\rangle\langle\ell|^{L_A} \otimes \mathrm{Tr}_A \, \Theta_\ell \left(|\rho\rangle\langle\rho|^{ARB} \right) - \sum_{\ell} |\ell\rangle\langle\ell|^{L_A} \otimes \mathrm{Tr}_A \, \widetilde{\Theta}_\ell \left(|\rho\rangle\langle\rho|^{ARB} \right) \right\|_1 \leq O(f(\varepsilon))$$

where $f(\varepsilon)$ is some rational power of ε.

The proof of this lemma can be found in Appendix B. We are now ready to describe the protocol which uses very little ancilla. The key idea stems from the observation that once Alice distills her pure qubits, she discards the system A_g. This system only serves to describe the correlations between the classical register in Alice's possession and Bob's quantum system B. As long as these correlations are preserved, we do not really care how the vectors in the system A_g are represented. This leads to the idea that we can embed the range o fevery POVM element $\widetilde{\Theta}^A_\ell$ using a unitary map U_ℓ into the *same space* A_g and attach a label $|\ell\rangle^{L_A}$ to it. This action requires that the rank of each POVM element $\widetilde{\Theta}^A_\ell$ be bounded above by (roughly) $2^{H_H^{\varepsilon^2}(RB|X)}$, since we will require $I^{\varepsilon}_{\max}(RB : X)$ qubits to store the classical labels. This is *almost* the guarantee that Lemma 6.1 gives us, since we can control the rank of every POVM element but the one that corresponds to the element $\perp$. This issue needs care and will be explained in greater detail in the proofs. Finally, if Alice is a able to carry out the embedding that we roughly described above, she can then send her L_A system to Bob via the classical channel, as before. Bob's actions remain the same as in Protocol C, and indeed distill the same number of pure qubits, since Lemma 6.1 also guaranteed us that the post measurement states are close. This will complete the description of the protocol.

The key technical idea therefore is the way Alice will carry out her embedding. To that end, consider the following lemma:

Lemma 6.2. *Consider the settings of Lemmas 5.5 and 6.1, and let $\widetilde{\Theta}^A$ be as given by Lemma 6.1. Assume that*

$$\max \left\{ I^{\varepsilon}_{\max}(RB : X) + H_H^{\varepsilon^2}(RB|X), H_H^{\varepsilon}(A) \right\} - O(\log \varepsilon) \leq \log |A| \,,$$

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

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

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:

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

where $V^{L_B B \to B_p B_g}$ encapsulates Bob's unitary operations as given in Protocol C, and $f(\varepsilon)$ is some rational function of ε.

Proof. First, consider the Schmidt decomposition of $|\rho\rangle^{ARB}$:

$$|\rho\rangle^{ARB} = \sum_i s_i |i\rangle^A |i\rangle^{RB}.$$

Consider the quantity $H_H^\varepsilon(A)$ and let Π^A be the positive semi-definite operator which optimises this expression. Consider the state

$$|\widetilde{\rho}\rangle^{ABR} := \sum_{|i\rangle \in \text{supp}(\Pi^A)} s'_i |i\rangle^A |i\rangle^{RB}.$$

where we define

$$s'_i := \frac{s_i}{\sqrt{\sum_{|i\rangle \in \text{supp}(\Pi^A)} s_i^2}}.$$

It is easy to see that

$$\| |\rho\rangle\langle\rho| - |\widetilde{\rho}\rangle\langle\widetilde{\rho}| \|_1 \leq 2\sqrt{\varepsilon}.$$

All our arguments from here will be based on $\widetilde{\rho}$. To ease the burden on notation, we neglect to explicitly mention that fact that $\sum_i$ implies summing over only those $|i\rangle$ which are in the support of $\widetilde{\rho}$. Notice that due to the closeness of ρ and $\widetilde{\rho}$, if Alice and Bob had carried out Protocol C on $\widetilde{\rho}$ instead of ρ, they would only incur an extra additive error of $2\sqrt{\varepsilon}$. Because of this, we can basse all our arguments on $\widetilde{\rho}$ and its marginals, instead of ρ.

Let us begin by first writing down the state if we were to act the POVM $\widetilde{\Theta}^A$ coherently on it, via the Stinespring dilation of the POVM, $V_{\widetilde{\Theta}}^{A \to AL_A}$:

$$V_{\widetilde{\Theta}}^{A \to AL_A} |\widetilde{\rho}\rangle^{ABR} = \sum_i \sum_\ell s'_i |\ell\rangle^{L_A} \sqrt{\widetilde{\Theta}_\ell} |i\rangle^A |i\rangle^{RB}$$

Note that, upon measuring the L_A system in the computational basis and tracing out the system A, we get the state

$$\sum_\ell \sum_{i,j} s'_i \cdot s'_j |\ell\rangle\langle\ell|^{L_A} \otimes \text{Tr}\left[\sqrt{\widetilde{\Theta}_\ell} |i\rangle\langle j|^A \sqrt{\widetilde{\Theta}_\ell}\right] \otimes |i\rangle\langle j|^{RB}$$

Further tracing out the system R gives the desired post measurement state $\widetilde{\rho}^{L_A B}$. Recall that our goal is to design a unitary operator which preserves the correlations between the L_A and the B systems. Notice that these correlations are controlled by the inner products between the vectors $\sqrt{\widetilde{\Theta}_\ell} |i\rangle$ and $\sqrt{\widetilde{\Theta}_\ell} |j\rangle$, for all pairs i, j. Using the fact that the rank of every POVM element (aside from $\widetilde{\Theta}_\perp$) is bounded above by $\frac{1}{\varepsilon} 2^{H_H^{\varepsilon^2}(RB|X)}$, we can then do the following:

1. For every $\ell \neq \perp$, there exists a unitary $U_\ell^{A \to A}$, which maps the eigenspace of the operator $\widetilde{\Theta}_\ell$ to some *fixed* space A_g, where we $\log |A_g| \geq H_H^{\varepsilon^2}(RB|X) + \log \frac{1}{\varepsilon}$. The precise dimension of A_g will be set later.

2. We then consider the following modified state:

$$\sum_i s'_i \left(\sum_{\ell \neq \perp} |\ell\rangle^{L_A} U_\ell \sqrt{\widetilde{\Theta}_\ell} |i\rangle^A + |\perp\rangle^{L_A} \sqrt{\widetilde{\Theta}_\perp} |i\rangle^A \right) |i\rangle^{RB}$$

3. The above state has several desirable properties. For one, it is easy to see that, the above state preserves the correlations between the classical system L_A (post measurement) and the quantum system B. This is because the unitaries U_ℓ preserve the angles between the vectors $\sqrt{\widetilde{\Theta}_\ell} |i\rangle$ and $\sqrt{\widetilde{\Theta}_\ell} |j\rangle$.

4. However, the vectors $U_\ell \sqrt{\widetilde{\Theta}_\ell}\,|\ell\rangle$ are now embedded in the space A_g.

The above discussion shows how we might go about designing our ideal unitary. We could simply define our unitary operator in the following way:

$$|i\rangle \to \sum_{\ell \neq \perp} |\ell\rangle^{L_A} \left(U_\ell \sqrt{\widetilde{\Theta}_\ell}\,|i\rangle \right)^{A_g} + |\perp\rangle^{L_A} \sqrt{\widetilde{\Theta}_\perp}\,|i\rangle$$

where the vector

$$\left(U_\ell \sqrt{\widetilde{\Theta}_\ell}\,|i\rangle \right)^{A_g}$$

is the low dimensional embedding of the original vector

$$U_\ell \sqrt{\widetilde{\Theta}_\ell}\,|i\rangle^{A}.$$

However, the above construction doesn't work since we have no control on the dimension of the subspace in which the vector $|\perp\rangle \sqrt{\widetilde{\Theta}_\perp}\,|i\rangle^A$ lives. What this means is that the vector

$$\sum_{\ell \neq \perp} |\ell\rangle^{L_A} \left(U_\ell \sqrt{\widetilde{\Theta}_\ell}\,|i\rangle \right)^{A_g} + |\perp\rangle^{L_A} \sqrt{\widetilde{\Theta}_\perp}\,|i\rangle$$

could potentially require more than $|A|$ dimensions to describe. This happens because we cannot provide any guarantees about the rank of the operator $\sqrt{\widetilde{\Theta}_\perp}$. To make the problem clear, consider the following:

1. We wish to embed the vector $|\perp\rangle \sqrt{\widetilde{\Theta}_\perp}\,|i\rangle^A$ in a space of dimension $|L_A A_g|$, for some appropriately chosen space A_g. Clearly, embedding this vector in the $L_A A_g$ system directly is not possible, since it may not be possible to embed the vector $\sqrt{\widetilde{\Theta}_\perp}\,|i\rangle$ into a space of dimension $|A_g|$, if our chosen A_g is of dimension smaller than A.

2. Another issue is that to ensure the unitarity, we must preserve the pairwise inner products among the vectors in the set $\left\{ \sqrt{\widetilde{\Theta}_\perp}\,|i\rangle^A \right\}$.

As a first step towards solving this issue, let us *assume* for now that we have in hand some vectors $\{w_i\}$ in the space $L_A A_g$ such that these vectors satisfy the following properties:

$$\langle w_i | w_j \rangle = \langle i |\, \widetilde{\Theta}_\perp\, |j\rangle \quad \forall i, j$$
$$\left\langle |w_i\rangle, |\ell\rangle \left(U_\ell \sqrt{\widetilde{\Theta}_\ell}\,|j\rangle \right) \right\rangle = 0 \;\; \forall \ell \neq \perp, \text{ and } i, j.$$

To ensure this, we must choose A_g large enough such that $L_A A_g$ can accommodate both the vectors of the form $|\ell\rangle \left(U_\ell \sqrt{\widetilde{\Theta}_\ell}\,|i\rangle \right)$ (for $\ell \neq \perp$) and also the vectors $|w_i\rangle$ defined above. We claim that we can always find such vectors $\{|w_i\rangle\}$ as long as

$$\log |L_A A_g| \geq \max \left\{ I^\varepsilon_{\max}(RB : X) + H^{\varepsilon^2}_H(RB|X), H^\varepsilon_H(A) \right\} - O(\log \varepsilon).$$

To see this, first note that there are $2^{H^\varepsilon_H(A)}$ many vectors in the set $\left\{ \sqrt{\widetilde{\Theta}_\perp}\,|i\rangle \right\}$. Let

$$\mathcal{B} \coloneqq \{u_1, u_2, \ldots, u_m\}$$

be an orthonormal basis for

$$\text{span}\left\{\sqrt{\widetilde{\Theta}_\perp}\,|i\rangle\right\}.$$

It is clear that

$$m \leq 2^{H_H^\varepsilon(A)} =: n.$$

To be precise, it holds that

$$\begin{aligned}\sqrt{\widetilde{\Theta}_\perp}\,|1\rangle &= \alpha_{11}u_1 + \alpha_{12}u_2 + \ldots + \alpha_{1m}u_m\\ &\vdots\\ \sqrt{\widetilde{\Theta}_\perp}\,|n\rangle &= \alpha_{n1}u_1 + \alpha_{n2}u_2 + \ldots + \alpha_{nm}u_m\end{aligned}$$

The problem is that, although we only need *at most* n many orthonormal basis vectors to describe the vectors of interest, the basis vectors themselves maybe embedded in a space of large dimension. However, it not hard to see that one can define vectors $\{w_1, w_2, \ldots, w_n\}$ such that:

$$\begin{aligned}w_1 &= \alpha_{11}\mathbf{e}_1 + \alpha_{12}\mathbf{e}_2 + \ldots + \alpha_{1m}\mathbf{e}_m\\ &\vdots\\ w_n &= \alpha_{n1}\mathbf{e}_1 + \alpha_{n2}\mathbf{e}_2 + \ldots + \alpha_{nm}\mathbf{e}_m\end{aligned}$$

where $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\}$ is the standard basis for $\mathbb{C}^m$. It is easy to see that for any i, j

$$\langle w_i|w_j\rangle = \langle i|\widetilde{\Theta}_\perp|j\rangle\,.$$

Note that, to ensure the second desired property, that for all i, j and $\ell \neq \perp$, $|w_i\rangle$ is orthogonal to $|\ell\rangle\left(U_\ell\sqrt{\widetilde{\Theta}_\ell}\,|j\rangle\right)$, we require that the vectors $|w_i\rangle$ be embedded in a space that is orthogonal to the span of the $|\ell\rangle\left(U_\ell\sqrt{\widetilde{\Theta}_\ell}\,|j\rangle\right)$'s. Therefore, to complete our construction, we require the system $L_A A_g$ to have at least

$$\frac{1}{O(\varepsilon)}\left(2^{I_{\max}^\varepsilon(RB:X)+H_H^{\varepsilon^2}(RB|X)} + 2^{H_H^\varepsilon(A)}\right)$$

dimensions. The description of the unitary U^A can then be given by:

$$\begin{aligned}U^A &: A \to L_A A_p A_g\\ &: |i\rangle^A \to \left(\sum_{\ell\neq\perp}|\ell\rangle^{L_A}\left(U_\ell\sqrt{\widetilde{\Theta}_\ell}\,|i\rangle\right)^{A_g} + |w_i\rangle^{L_A A_g}\right)|0\rangle^{A_p} \quad \forall\,|i\rangle \in \text{supp}(\Pi^A).\end{aligned}$$

The unitary can be completed in an arbitrary manner on the rest of A. To count the dimensions of A_g and A_p, note that the condition for the existence of U^A implies that we can set

$$\log|L_A A_g| = \max\left\{I_{\max}^\varepsilon(RB:X) + H_H^{\varepsilon^2}(RB|X), H_H^\varepsilon(A)\right\} - O(\log\varepsilon).$$

Since by design $L_A A_p A_g \cong A$, it then holds that

$$\log|A_p| = \log|A| - \max\left\{I_{\max}^\varepsilon(RB:X) + H_H^{\varepsilon^2}(RB|X), H_H^\varepsilon(A)\right\} + O(\log\varepsilon).$$

Next, note that

$$\begin{aligned}U^A|\widetilde{\rho}\rangle^{ABR} &= |0\rangle^{A_p}\sum_i s_i'\sum_{\ell\neq\perp}|\ell\rangle^{L_A}\left(U_\ell\sqrt{\widetilde{\Theta}_\ell}\,|i\rangle\right)^{A_g}|i\rangle^{RB} + |0\rangle^{A_p}\sum_i s_i'\,|w_i\rangle^{L_A A_g}\,|i\rangle^{RB}\\ &:= \sqrt{p}\cdot|v_{\text{GOOD}}\rangle^{L_A A_p A_g RB} + \sqrt{1-p}\,|\text{JUNK}\rangle^{L_A A_p A_g RB}\,.\end{aligned}$$

By design $|v_{\text{GOOD}}\rangle$ and $|\text{JUNK}\rangle$ are orthogonal. Also note that:

$$\begin{aligned} p &= \sum_i (s'_i)^2 \sum_{\ell \neq \perp} \langle i | \widetilde{\Theta}_\ell | i \rangle \\ &\geq 1 - g(\varepsilon), \end{aligned}$$

where $g(\varepsilon)$ is again some rational power of ε. This is because by design, the POVM $\widetilde{\Theta}$ has very small probability of giving the outcome $\perp$ when measuring the state ρ^A. This directly implies that:

$$\begin{aligned} \left\| U^A \left| \widetilde{\rho} \right\rangle \left\langle \widetilde{\rho} \right|^{ABR} U^{\dagger A} - \left| v_{\text{GOOD}} \right\rangle \left\langle v_{\text{GOOD}} \right| \right\|_1 &= \sqrt{2} \cdot \sqrt{1 - \left| \langle \widetilde{\rho} | U^A | v_{\text{GOOD}} \rangle \right|^2} \\ &\leq \sqrt{2g(\varepsilon)}. \end{aligned}$$

therefore, after Alice applies the unitary U^A on her system A, we can pretend as if the global state is $|v_{\text{GOOD}}\rangle$. The utility of this step is that we do not have to deal with the result of the action of the channel on the vectors $|w_i\rangle^{L_A A_g}$. Indeed, if we pretend that the global state is $|v_{\text{GOOD}}\rangle$, and Alice then sends her L_A register through the channel, the global state is exactly what it would be if Alice were to measure her A system with the POVM $\widetilde{\Theta}^A$ and send the classical register to Bob after conditioning on the $\perp$ event *not* occurring. Of course, since the $\perp$ event has very little probability of occurring, we can finally replace the state which occurs after the channel acts on $|v_{\text{GOOD}}\rangle$ with the honest to god state that would have arisen if Alice had measured her system A with $\widetilde{\Theta}$ and sent the entire classical register to Bob without any conditioning. What we have argued here can be encapsulated by the following calculation:

$$\begin{aligned} & \left\| \left(\mathcal{P}^{L_A \to L_B} \circ \text{Tr}_{A_g} \circ U^A \right) \cdot \widetilde{\rho}^{ABR} - |0\rangle \langle 0|^{A_p} \otimes \left(\mathcal{P}^{L_A \to L_B} \circ \widetilde{\Theta}^{A \to L} \right) \widetilde{\rho}^{ABR} \right\|_1 \\ \leq & \left\| U^A \cdot \widetilde{\rho}^{ABR} - |v_{\text{GOOD}}\rangle \langle v_{\text{GOOD}}| \right\|_1 \\ + & \left\| \left(\mathcal{P}^{L_A \to L_B} \circ \text{Tr}_{A_g} \right) \cdot |v_{\text{GOOD}}\rangle \langle v_{\text{GOOD}}| - |0\rangle \langle 0|^{A_p} \otimes \left(\mathcal{P}^{L_A \to L_B} \circ \widetilde{\Theta}^{A \to L} \right) \widetilde{\rho}^{ABR} \right\|_1 \\ \leq & O(\sqrt{g(\varepsilon)}). \end{aligned}$$

To complete the proof, note that we know from Lemma 6.1 that the post measurement state of Θ^A (which we get from Lemma 5.5) and that of $\widetilde{\Theta}^A$ are close by some rational power of ε. Therefore, Bob's actions in Protocol C on his registers produce the same number of qubit states on his end (with some additional additive error as a function of ε) if Alice had carried out Protocol C with $\widetilde{\Theta}^A$ instead of Θ^A. A further additive $O(\sqrt{\varepsilon})$ error is incurred by replacing $|\rho\rangle^{ARB}$ with $|\widetilde{\rho}\rangle^{ARB}$. Thus, the total error incurred by Alice and Bob is again some rational power of ε, while distilling $\log |A| - \max \left\{ I^\varepsilon_{\max}(RB:X) + H^{\varepsilon^2}_H(RB|X), H^\varepsilon_H(A) \right\} + O(\log \varepsilon)$ on Alice's end and $\log |B| - H^{\varepsilon^2}_H(B|X) + O(\log \varepsilon)$. This concludes the proof. □

We will now deal with the case when

$$\max \left\{ I^\varepsilon_{\max}(RB:X) + H^{\varepsilon^2}_H(RB|X), H^\varepsilon_H(A) \right\} - O(\log \varepsilon) > \log |A| .$$

Clearly, this is a pathological case, since one would expect that the entropic quantities within the max would always be less than or equal to the dimension of the entire space. Indeed, it is not hard to see that $H^\varepsilon_H(A)$ does obey this property. However, it may happen that for some pathological case $I^\varepsilon_{\max}(RB:X) + H^{\varepsilon^2}_H(RB|X) - O(\log \varepsilon)$ exceeds $\log |A|$. In that case, Alice cannot distill any pure qubits. Indeed, she has to borrow some qubits to even implement the unitary U^A, which we constructed in Lemma 6.2. However, she will be able to implement the unitary U^A following the same recipe that we showed in Lemma 6.2 if she borrows:

$$\max \left\{ I^\varepsilon_{\max}(RB:X) + H^{\varepsilon^2}_H(RB|X), H^\varepsilon_H(A) \right\} - O(\log \varepsilon) - \log |A|$$

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

$$\begin{aligned} I^\varepsilon_{\max}(RB:X) &\leq H^{O(\varepsilon^2)}_{\max}(RB) - H^{O(\varepsilon^2)}_{\min}(RB|X) - O(\log \varepsilon) \\ &\leq H^{O(\varepsilon^2)}_H(A) - H^{O(\varepsilon^2)}_{\min}(RB|X) - O(\log \varepsilon) \\ &\leq \log |A| - H^{O(\varepsilon^2)}_{\min}(RB|X) - O(\log \varepsilon). \end{aligned}$$

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

$$\Delta(RB|X) \coloneqq H_H^{\varepsilon^2}(RB|X) - H_{\min}^{O(\varepsilon^2)}(RB|X) - O(\log \varepsilon)$$

many qubits. We state this as a lemma below:

Lemma 6.3. *Given the setting of Lemma 6.2, suppose that*

$$\max\left\{I_{\max}^{\varepsilon}(RB:X) + H_H^{\varepsilon^2}(RB|X), H_H^{\varepsilon}(A)\right\} - O(\log \varepsilon) > \log|A|\,.$$

In this case, Alice can implement the unitary $U^{A\to L_A A_g}$ defined in Lemma 6.2 by borrowing at most

$$\Delta(RB|X) \coloneqq H_H^{\varepsilon^2}(RB|X) - H_{\min}^{O(\varepsilon^2)}(RB|X) - O(\log \varepsilon)$$

many qubits. Note that there is no A_p system for this case since Alice cannot distill any pure qubiits by herself. The net purity that Alice and Bob together distill 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 6.3 does not occur.

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, Arun Padakandla, and Pranab Sen. Centralised multi link measurement compression with side information, 2022.

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

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

[7] 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.

[8] 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.

[9] 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.

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

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

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

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

[14] 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.

Appendix A Towards Proving Lemma 5.3

Lemma A.1. *Let ρ^{XB} be a cq state of the form*

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

Then it holds that there exists a set $\text{GOOD}_X \subseteq \mathcal{X}$ such that the following conditions hold

$$\Pr_{P_X}\left[\text{GOOD}_X\right] \geq 1 - 2\sqrt{\varepsilon}$$

and

$$H_H^{\sqrt{\varepsilon}}(\rho_x^B) \leq H_H^{\varepsilon}(B \mid X)_{\rho^{XB}} + \log\frac{1}{\varepsilon} \quad \forall x \in \text{GOOD}_X$$

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

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

Then, by definition, it holds that

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

and

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

By Markov's inequality, it is easy to see that:

1. There exists a set $\mathcal{S}_1 \subseteq \mathcal{X}$ of probability at least $1 - \sqrt{\varepsilon}$ such that, for all $x \in \mathcal{S}_1$ it holds that

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

2. There exists a set $\mathcal{S}_2 \subseteq \mathcal{X}$ of probability at least $1 - \varepsilon$ such that for all $x \in \mathcal{S}_2$ it holds that

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

We now define the set

$$\text{GOOD}_X \coloneqq \mathcal{S}_1 \bigcap \mathcal{S}_2$$

and note that the probability of this set is at least $1 - 2\sqrt{\varepsilon}$. Fix a member x of $\mathcal{S}$. Then for this fixed x, it holds that

$$\begin{aligned} 2^{H_H^{\sqrt{\varepsilon}}(\rho_x^B)} &\leq \operatorname{Tr}\left[\Pi_x^B\right] \\ &\leq \frac{2^{H_H^{\varepsilon}(B \mid X)}}{\varepsilon} \end{aligned}$$

Taking logs on both sides concludes the proof. □

Lemma A.2. *For the setting of Lemma A.1, consider the cq state*

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

where we are promised that for all $x \in \mathcal{X}$,

$$\left\|\sigma_x^B - \rho_x^B\right\|_1 \leq \varepsilon$$

Then there exists a set $\text{GOOD}'_X \subseteq \mathcal{X}$ such that

$$\Pr_{P_X}\left[\text{GOOD}'_X\right] \geq 1 - 4\sqrt{\varepsilon}$$

and

$$H_H^{\sqrt{2\varepsilon}}(\sigma_x^B) \leq H_H^{\varepsilon}(B \mid X) + \log \frac{1}{\varepsilon}$$

Proof. The proof is very similar to that of Lemma A.1. recall that we had defined the operator Π_{OPT} as the optimiser for the quantity $H_H^{\varepsilon}(B \mid X)$. First note that

$$\begin{aligned} \sum_x P_X(x) \operatorname{Tr}\left[\Pi_x^B \sigma_x^B\right] &\geq \sum_x P_X(x) \operatorname{Tr}\left[\Pi_x^B \rho_x^B\right] - \sum_x P_X(x) \left\|\rho_x - \sigma_x\right\| \\ &\geq \sum_x P_X(x) \operatorname{Tr}\left[\Pi_x^B \rho_x^B\right] - \varepsilon \\ &\geq 1 - 2\varepsilon \end{aligned}$$

We then define the set $\mathcal{S}'_1$ to be the set of those x such that

$$\operatorname{Tr}\left[\Pi_x^B \sigma_x^B\right] \geq 1 - \sqrt{2\varepsilon}$$

and it is also easy to see that the probability of $\mathcal{S}'_1$ is at least $1 - \sqrt{2\varepsilon}$. We define $\mathcal{S}_2$ as before. We then define

$$\text{GOOD}'_x \coloneqq \mathcal{S}'_1 \bigcap \mathcal{S}_2$$

which implies that the probability of GOOD'_x is at least $1 - 4\sqrt{\varepsilon}$. Then, fixing an element $x \in \text{GOOD}'_X$ we see that

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

This completes the proof. □

Lemma A.3. *Consider the settings of Lemma A.1 and Lemma A.2. Let the cq state ρ^{XB} be as in Lemma A.1. Let ρ'^{KB} be a cq state of the form*

$$\rho'^{KB} \coloneqq \sum_k Q_K(k) \left|k\right\rangle \left\langle k\right|^K \otimes \rho_k'^B$$

Let f be a deterministic function such that

$$\begin{aligned} f : [K] &\to \mathcal{X} \\ : k &\to x \end{aligned}$$

and define

$$Q_X(x) \coloneqq \sum_k Q_K(k) \mathtt{1}_{f(k)=x}$$

We are promised that

$$\rho_k^{\prime B} = \sigma_{f(k)}^B$$

and

$$\|Q_X - P_X\|_1 \leq \varepsilon$$

Where σ_x^B is as in Lemma A.2. Then it holds that

$$2^{H_H^{\varepsilon'}(\rho^{\prime KB})} \leq K \cdot \frac{2^{H_H^{\varepsilon}(B \mid X)}}{\varepsilon}$$

Proof. We will work with the set GOOD'_X as given in Lemma A.2. Recall that,

$$\begin{aligned} &\Pr\left[\text{GOOD}'_X\right] \geq 1 - 4\sqrt{\varepsilon} \\ &H_H^{\sqrt{2\varepsilon}}(\sigma_x^B) \leq H_H^{\varepsilon}(B \mid X) + \log \frac{1}{\varepsilon} \end{aligned}$$

Define the set $\text{GOOD}_K \subseteq [K] \times [L]$ such that

$$\text{GOOD}_K \coloneqq \left\{k \mid f(k) \in \text{GOOD}'_X\right\}$$

It is then easy to see that

$$\begin{aligned} \Pr_{Q_K}\left[\text{GOOD}_K\right] &= \sum_{x \in \text{GOOD}_X} \sum_k Q_K(k) \mathtt{1}_{f(k)=x} \\ &= \sum_{x \in \text{GOOD}_X} Q_X(x) \\ &\geq \sum_{x \in \text{GOOD}_X} P_X(x) - \varepsilon \\ &\geq 1 - 3\sqrt{\varepsilon} \end{aligned}$$

Now, recall that we had defined the operator

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

as the optimiser for the quantity $H_H^{\varepsilon}(B \mid X)_{\rho^{XB}}$. Define the operator

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

The note that

$$\begin{aligned}
\operatorname{Tr}\left[\Pi'^{KB}\rho'^{KB}\right] &= \sum_{k\in \text{GOOD}_K} Q_K(k)\operatorname{Tr}\left[\Pi^B_{f(k)}\rho'^B_k\right] \\
&= \sum_{x\in \text{GOOD}'_X} \operatorname{Tr}\left[\Pi^B_x\sigma^B_x\right] \sum_{f(k)=x} Q_K(k) \\
&\geq (1-\sqrt{2\varepsilon})\cdot \Pr_{Q_K}\left[\text{GOOD}_K\right] \\
&\geq (1-\sqrt{2\varepsilon})\cdot(1-3\sqrt{\varepsilon}) \\
&\geq 1-6\sqrt{\varepsilon}
\end{aligned}$$

It then follows that

$$\begin{aligned}
2^{H_H^{6\sqrt{\varepsilon}}(KB)_{\rho'^{KB}}} &\leq \operatorname{Tr}\left[\Pi'^{KB}\right] \\
&\leq K\cdot\frac{2^{H_H^{\varepsilon}(B\mid X)}}{\varepsilon}
\end{aligned}$$

This concludes the proof. □

Lemma A.4. *Given the setting of Lemma A.3, let the marginal distribution Q_K be close to the uniform on the set $[K]$, i.e.,*

$$\|Q_K - \mathbf{Unif}[K]\|_1 \leq \delta$$

Then, at least $1-3\varepsilon'$ fraction of $k\in[K]$ satisfy the condition that

$$2^{H_H^{\varepsilon'}(B\mid k)_{\rho'}} \leq \frac{1}{1-\varepsilon'}\cdot L\cdot\frac{2^{H_H^{\varepsilon}(B\mid X)}}{\varepsilon}$$

where $\varepsilon' \coloneqq 2\left(6\varepsilon^{1/2}+\delta\right)$.

Proof. Repeating the calculation in the proof of Lemma A.3, we see that

$$\begin{aligned}
1-6\sqrt{\varepsilon} &\leq \operatorname{Tr}\left[\Pi'^{KB}\rho'^{KB}\right] \\
&= \sum_{k\in\text{GOOD}_K} Q_K(k)\operatorname{Tr}\left[\Pi^B_{f(k)}\rho'^B_k\right] \\
&\leq \sum_{k\in\text{GOOD}_K} \frac{1}{K}\operatorname{Tr}\left[\Pi^B_{f(k)}\rho'^B_k\right] + \delta \\
&\leq \sum_{k\in\text{GOOD}_K} \frac{1}{|\text{GOOD}_K|}\operatorname{Tr}\left[\Pi^B_{f(k)}\rho'^B_k\right] + \delta
\end{aligned}$$

This implies, by Markov's inequality, that there exists a subset $\text{NICE}_K \subseteq \text{GOOD}_K \subseteq [K]$ which contains at least

$$\begin{aligned}
&\left(1-\left(6\varepsilon^{1/2}+\delta\right)^{1/2}\right)\cdot\left(1-3\sqrt{\varepsilon}\right) \\
&\geq 1-2\cdot\left(6\varepsilon^{1/2}+\delta\right)^{1/2}
\end{aligned}$$

fraction of k's, such that, for all $k\in\text{NICE}_K$,

$$1-2\left(6\varepsilon^{1/2}+\delta\right)^{1/2} \leq \operatorname{Tr}\left[\Pi^B_{f(k)}\rho'^B_k\right]$$

Note that this implies that, for all $k\in\text{GOOD}_K$ the operators

$$\Pi^{B\mid k} \coloneqq \Pi^B_{f(k)}$$

are candidates to optimise the expressions

$$H_H^{\varepsilon'}(B \mid k) \coloneqq H_H^{\varepsilon'}(\rho_k'^B)$$

where

$$\varepsilon' \coloneqq 2\left(6\varepsilon^{1/2} + \delta\right)^{1/2}$$

Then, the following holds

$$\begin{aligned}
\frac{1}{|\text{NICE}_K|} \sum_{k \in \text{NICE}_K} 2^{H_H^{\varepsilon'}(B \mid k)} &\leq \frac{1}{1-\varepsilon'} \cdot \frac{1}{K} \sum_{k \in \text{NICE}_K} 2^{H_H^{\varepsilon'}(B \mid k)} \\
&\leq \frac{1}{1-\varepsilon'} \cdot \sum_{k \in \text{NICE}_K} \frac{1}{K} \operatorname{Tr}\left[\Pi^{B \mid k}\right] \\
&= \frac{1}{1-\varepsilon'} \cdot \frac{1}{K} \operatorname{Tr}\left[\sum_{k \in \text{NICE}_K} |k\rangle\langle k|^K \otimes \Pi^{B \mid k}\right] \\
&\leq \frac{1}{1-\varepsilon'} \cdot \frac{1}{K} \operatorname{Tr}\left[\Pi'^{KB}\right] \\
&\leq \frac{1}{1-\varepsilon'} \cdot \frac{1}{K} \cdot K \cdot \frac{2^{H_H^{\varepsilon}(B \mid X)}}{\varepsilon} \\
&= \frac{1}{1-\varepsilon'} \cdot \frac{2^{H_H^{\varepsilon}(B \mid X)}}{\varepsilon} \qquad (2)
\end{aligned}$$

Again, Markov's inequality implies that $1-\varepsilon$ fraction of the $k \in \text{NICE}_K$ (which is at least $1-3\varepsilon'$ fraction of all $k \in [K]$) satisfy the condition that

$$2^{H_H^{\varepsilon'}(B \mid k)} \leq \frac{1}{1-\varepsilon'} \frac{2^{H_H^{\varepsilon}(B \mid X)}}{\varepsilon^2}$$

This concludes the proof. □

Proof of Lemma 5.3

Proof. Collating the ideas in the previous discussion, we see that the state

$$\sum_{k,\ell} Q_{KL}(k,\ell) |k,\ell\rangle\langle k,\ell|^{KL} \otimes \sigma_{f(k,\ell)}^{RB} \qquad \text{(Bob and Ref)}$$

and the state

$$\sum_{k,\ell} Q_{KL}(k,\ell) |k,\ell\rangle\langle k,\ell|^{KL} \otimes \sigma_{f(k,\ell)}^{B} \qquad \text{(Bob)}$$

both satisfy the requirements of Lemma A.4. The closeness of $\sigma_{f(k,\ell)}^{RB}$ and $\rho_{f(k,\ell)}^{RB}$ (for $(k,\ell) \in \text{supp}(Q_{KL})$) implies the closeness of the marginals $\sigma_{f(k,\ell)}^{B}$ and $\rho_{f(k,\ell)}^{B}$. Also, since

$$Q_X(x) \coloneqq \sum_{k,\ell : f(k,\ell)=x} Q_{KL}(k,\ell),$$

it holds that

$$\|P_X - Q_X\|_1 \leq c_3 \varepsilon^{1/4}.$$

We will first instantiate parameters, then use Lemma A.4 twice. Set

$$\delta \leftarrow c_7 \varepsilon^{1/4}$$
$$\varepsilon \leftarrow c_8 \varepsilon^{1/4}$$

where $c_8 \geq \max c_3, c_4$. Then, we first apply Lemma A.4 to Eq. (Bob and Ref) to see that there exists a subset

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

with the property that

$$|\mathcal{S}_1| \geq (1 - c_9 \varepsilon^{1/8}) \cdot KL$$

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

$$H_H^{O(\varepsilon^{1/8})}(RB \mid k, \ell) \leq H_H^{\varepsilon^{1/4}}(RB \mid X) + O(\log \frac{1}{\varepsilon}) + O(1).$$

Similarly, applying Lemma A.4 to Eq. (Bob), we see that there exists a set $\mathcal{S}_2 \subseteq [K] \times [L]$ such that

$$|\mathcal{S}_2| \geq (1 - c_{10} \varepsilon^{1/8}) KL$$

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

$$H_H^{O(\varepsilon^{1/8})}(B \mid k, \ell) \leq H_H^{\varepsilon^{1/4}}(B \mid X) + O(\log \frac{1}{\varepsilon}) + O(1).$$

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

$$H_H^{O(\varepsilon^{1/8})}(RB \mid k, \ell) \leq H_H^{\varepsilon^{1/4}}(RB \mid X) + O(\log \frac{1}{\varepsilon}) + O(1)$$

and

$$H_H^{O(\varepsilon^{1/8})}(B \mid k, \ell) \leq H_H^{\varepsilon^{1/4}}(B \mid X) + O(\log \frac{1}{\varepsilon}) + O(1).$$

□

Appendix B A Perturbed POVM: Proof of Lemma 6.1

For our main result, we will need to work with a POVM which has the property that most of the POVM elements have rank at most $2^{H_H^\varepsilon(RB \mid X)}$. This property will be crucial for our main protocol. To find such a POVM, consider th following lemma, whose proof follows in a straightforward manner from Lemma 5.3.

Lemma B.1. *For the setting of Lemma 5.3, there exists an index $k^* \in [K]$ such that, for at least $(1 - \varepsilon^{1/16})$ fraction of $\ell \in [L]$, it holds that*

$$H_H^{O(\varepsilon^{1/8})}(RB \mid k^*, \ell) \leq H_H^{\varepsilon^{1/4}}(RB \mid X) + O(\log \frac{1}{\varepsilon}) + O(1)$$

and

$$H_H^{O(\varepsilon^{1/8})}(B \mid k^*, \ell) \leq H_H^{\varepsilon^{1/4}}(B \mid X) + O(\log \frac{1}{\varepsilon}) + O(1).$$

.

Proof. From Lemma 5.3, we know that the entropic inequalities in the statement of the lemma hold for at least $(1-\varepsilon^{1/8})$ 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,

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

Define

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

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

$$\textsc{prob}_k \geq 1-\varepsilon^{1/16}.$$

Let k^* be such that

$$\textsc{prob}_{k^*} \geq 1-\varepsilon^{1/16}.$$

Then by the fact that $\mathtt{prob}_{k^*}$ 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

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

□

Definition B.2. *For the setup of Lemma B.1, define the set of all ℓ for which the conditions in the lemma holds to be the set*

$$\textsc{nice}_{L \mid k^*}.$$

B.1 Remarks about Measurement Compression

Before we proceed, we need to say a few words about the post measurement state in the measurement compression theorem. Recall that, post measurement, it holds that

$$\left\| \rho^{XRB} - \sum_x \sum_{k,\ell} Q_{KL}(k,\ell) \cdot \mathtt{1}_{f(k,\ell)=x} \left|x\right\rangle \left\langle x\right|^X \otimes \sigma^{RB}_{f(k,\ell)} \right\|_1 \leq c_3 \varepsilon^{1/4}$$

Both states inside the 1-norm above are the post measurement states that are created by measuring the pure state $\left|\varphi\right\rangle^{ABR}$ by the operators in the original and the simulating POVM respectively. However, while designing the simulated POVM, it is easier to consider a purification $\left|\rho\right\rangle^{A\widetilde{R}}$ of the marginal ρ^A. Note that by Uhlmann's theorem, there exists a unitary map which maps $\widetilde{R} \to RB$. Thus, using the unitary invariance of the 1-norm, the closeness condition of the post measurement state can be stated as:

$$\left\| \rho^{X\widetilde{R}} - \sum_x \sum_{k,\ell} Q_{KL}(k,\ell) \cdot \mathtt{1}_{f(k,\ell)=x} \left|x\right\rangle \left\langle x\right|^X \otimes \sigma^{\widetilde{R}}_{f(k,\ell)} \right\|_1 \leq c_5 \varepsilon^{1/4}$$

where we define,

$$\rho^{\widetilde{R}}_x = U^{RB\to\widetilde{R}} \cdot \rho^{RB}_x$$
$$\sigma^{\widetilde{R}}_{f(k,\ell)} = U^{RB\to\widetilde{R}} \cdot \sigma^{RB}_{f(k,\ell)}.$$

and that

$$\left\| \rho^{\widetilde{R}}_{f(k,\ell)} - \sigma^{\widetilde{R}}_{f(k,\ell)} \right\| \leq c_4 \varepsilon^{1/4}$$

for all (k,ℓ) in the support of Q_{KL}. The advantage of working with the $\widetilde{R}$ system becomes clear when we consider the fact that, due to the invariance of the smooth hypothesis testing entropy under the action of an isometry, it holds that,

$$\begin{aligned} H_H^{\sqrt{\varepsilon'}}(\rho'^{RB}_{k,\ell}) &= H_H^{\sqrt{\varepsilon'}}(U^{RB} \cdot \rho'^{RB}_{k,\ell}) \\ &= H_H^{\sqrt{\varepsilon'}}(\rho'^{\widetilde{R}}_{k,\ell}) \end{aligned}$$

B.2 New POVM

We will work with the POVM

$$\Theta(k^*) = \{\Theta_1(k^*), \ldots, \Theta_L(k^*), \Theta_\perp(k^*)\}.$$

Consider the fact that (CPS'22), for each $\ell \in [L]$, the operator $\Theta_\ell(k^*)$ has the following form:

$$\Theta_\ell(k^*) = \frac{1}{1 - c_9\varepsilon^{1/4}} \cdot \frac{1}{L} \cdot \rho^{-1/2} \rho'_{k^*,\ell} \rho^{-1/2}$$

where the operator ρ is the marginal of ρ^{AB} on the system A and the state $\rho'_{k^*,\ell}$ is defined on the system $\widetilde{R}$. The reader is referred to cite CPS'22for details. We will perturb these POVM elements slightly to form a new POVM $\widetilde{\Theta}(k^*)$. To do this, first consider all the indexes

$$\ell \in \text{NICE}_{L \mid k^*}$$

and their corresponding POVM elements $\Theta_\ell(k^*)$. For all such ℓ, we define

$$\widetilde{\Theta}_\ell(k^*) \coloneqq \frac{1}{1 - c_9\varepsilon^{1/4}} \cdot \frac{1}{L} \cdot \rho^{-1/2} \Pi_{k^*,\ell} \rho'_{k^*,\ell} \Pi_{k^*,\ell} \leq \rho^{-1/2}$$

where the projector $\Pi_{k^*,\ell}$ is defined as the projector which optimises the expression $H_H^{\sqrt{\varepsilon'}}(\rho'^{RB}_{k^*,\ell}) = H_H^{\sqrt{\varepsilon'}}(\rho'^{\widetilde{R}}_{k^*,\ell})$. Recall that this projector commutes with $\rho'_{k^*,\ell}$ and removes at most $\sqrt{\varepsilon'}$ amount of mass from its support.

Let us first consider the effect this perturbation has on the post measurement state. Consider the purification $|\rho\rangle^{A\widetilde{R}}$. Then,

$$\begin{aligned}
&\left\| \operatorname{Tr}_A \Theta_\ell(k^*) |\rho\rangle \langle\rho|^{A\widetilde{R}} - \operatorname{Tr}_A \widetilde{\Theta}_\ell(k^*) |\rho\rangle \langle\rho|^{A\widetilde{R}} \right\|_1 \\
=&\frac{1}{1 - c_9\varepsilon^{1/4}} \cdot \frac{1}{L} \left\| \rho'_{k^*,\ell} - \Pi_{k^*,\ell} \cdot \rho'_{k^*,\ell} \right\|_1 \\
\leq&\frac{1}{1 - c_9\varepsilon^{1/4}} \cdot \frac{1}{L} \cdot \sqrt{\varepsilon'}
\end{aligned}$$

Remark B.3. The above inequality still holds true when we use the unitary $U^{RB \to \widetilde{R}}$ to map the $\widetilde{R}$ system into the RB system.

Remark B.4. An important point to note is that conjugating $\rho'_{k^*,\ell}$ with the projector $\Pi_{k^*,\ell}$ also preserves the operator inequality which is at the heart of the POVM construction. This is because:

1. Conjugation is a CP map which preserves operator inequalities.
2. The projector commutes with $\rho'_{k^*,\ell}$. This ensures that

$$\begin{aligned}
\frac{1}{1 - c_9\varepsilon^{1/4}} \cdot \frac{1}{L} \Pi_{k^*,\ell} \rho'_{k^*,\ell} \Pi_{k^*,\ell} &\leq \frac{1}{1 - c_9\varepsilon^{1/4}} \cdot \frac{1}{L} \rho'_{k^*,\ell} \\
&\leq \rho
\end{aligned}$$

Define

$$\widetilde{\Theta}_\perp(k^*) \coloneqq \mathbb{I}^A - \sum_{\ell \in \text{NICE}_{L \mid k^*}} \widetilde{\Theta}_\ell(k^*)$$

Then,

$$\begin{aligned}
\operatorname{Tr}\left[\widetilde{\Theta}_{\perp}(k^*)\rho^A\right] &= \operatorname{Tr}\left[\rho^A - \sum_{\ell\in \text{NICE}_{L\mid k^*}} \widetilde{\Theta}_{\ell}(k^*)\rho^A\right] \\
&= 1 - \sum_{\ell\in \text{NICE}_{L\mid k^*}} \operatorname{Tr}\left[\widetilde{\Theta}_{\ell}(k^*)\rho^A\right] \\
&\leq 1 + \frac{\sqrt{\varepsilon'}}{1-c_9\varepsilon^{1/4}}\cdot\frac{\left|\text{NICE}_{L\mid k^*}\right|}{L} - \sum_{\ell\in \text{NICE}_{L\mid k^*}} \operatorname{Tr}\left[\Theta_{\ell}(k^*)\rho^A\right] \\
&\leq \sqrt{\varepsilon'}\cdot\frac{1}{1-c_9\varepsilon^{1/4}} + \operatorname{Tr}\left[\Theta_{\perp}(k^*)\rho^A\right] \\
&\leq \sqrt{\varepsilon'}\cdot\frac{1}{1-c_9\varepsilon^{1/4}} + c_0\varepsilon^{1/4}
\end{aligned}$$

For an appropriately small ε, the LHS is then less than

$$\leq c_{10}\varepsilon'^{1/2}$$

Thus, it holds that

$$\begin{aligned}
&\left\|\sum_{\ell}|\ell\rangle\langle\ell|^L\otimes \operatorname{Tr}_A\Theta_{\ell}(k^*)|\varphi\rangle\langle\varphi|^{ARB} - \sum_{\ell\in\text{NICE}_{L\mid k^*}\cup\{\perp\}}|\ell\rangle\langle\ell|^L\otimes\operatorname{Tr}_A\widetilde{\Theta}_{\ell}(k^*)|\varphi\rangle\langle\varphi|^{ARB}\right\|_1 \\
&\leq \sum_{\ell\in\text{NICE}_{L\mid k^*}}\left\|\operatorname{Tr}_A\Theta_{\ell}(k^*)|\varphi\rangle\langle\varphi|^{ARB} - \operatorname{Tr}_A\widetilde{\Theta}_{\ell}(k^*)|\varphi\rangle\langle\varphi|^{ARB}\right\|_1 \\
&+ \sum_{\substack{\ell\notin\text{NICE}_{L\mid k^*}\\ \ell\in\operatorname{supp}(Q_{L\mid k^*})\\ \ell\neq\perp}}\left\|\operatorname{Tr}_A\Theta_{\ell}(k^*)|\varphi\rangle\langle\varphi|^{ARB}\right\|_1 \\
&\qquad + \left\|\widetilde{\Theta}_{\perp}(k^*)|\varphi\rangle\langle\varphi|^{ARB}\right\|_1 + \left\|\Theta_{\perp}(k^*)|\varphi\rangle\langle\varphi|^{ARB}\right\|_1 \\
&\leq \frac{\sqrt{\varepsilon'}}{1-c_9\varepsilon^{1/4}}\frac{\left|\text{NICE}_{L\mid k*}\right|}{L} + \frac{1}{1-c_9\varepsilon^{1/4}}\frac{\left|\text{NICE}^c_{L\mid k*}\right|}{L} + c_{10}\varepsilon'^{1/2} + c_0\varepsilon^{1/4} \\
&\leq 3c_{10}\varepsilon'^{1/2} + \frac{1}{1-c_9\varepsilon^{1/4}}\frac{2c_7\varepsilon'\cdot L}{L} \\
&\leq c_{11}\varepsilon'^{1/2}
\end{aligned}$$

We will use the POVM $\widetilde{\Theta}(k^*)$ instead of the POVM $\Theta(k^*)$. The result above shows that the post measurement states created by these two POVMs are not very far apart. The main reason we consider this perturbation however is that, aside from the element $\widetilde{\Theta}_{\perp}(k^*)$, every other POVM element of $\widetilde{\Theta}(k^*)$ has rank at most $2^{H_H^{\sqrt{\varepsilon'}}(\rho'^{RB}_{k^*,\ell})}$. Then, invoking Lemma 5.3, we see that each such POVM element has rank at most

$$\exp\left(H_H^{\varepsilon}(RB\mid X) + 2\log\frac{1}{c_8\cdot(1-\varepsilon'')\cdot\varepsilon^2}\right).$$

We will use this property extensively to design the actual unitary operator that Alice uses on her system.