Optimal State Merging without Decoupling

Jean-Christian Boileau[1] and Joseph M. Renes[2]

[1] Center for Quantum Information and Quantum Control, University of Toronto
[2] Institut für Angewandte Physik, Technical University of Darmstadt

Abstract. We construct an optimal state merging protocol by adapting a recently-discovered optimal entanglement distillation protcol [Renes and Boileau, Phys. Rev. A. 73, 032335 (2008)]. The proof of optimality relies only on directly establishing sufficient "amplitude" and "phase" correlations between Alice and Bob and not on usual techniques of decoupling Alice from the environment. This strengthens the intuition from quantum error-correction that these two correlations are all that really matter in two-party quantum information processing.

1 Introduction

Quantum state merging is an important primitive protocol in the hierarchy of quantum communication protocols, also known as the quantum information family tree. Given two parties Alice and Bob and a mixed bipartite state ψ^{AB}, the goal of state merging is simply for Alice to send her half of the state to Bob. One option, of course, is to compress ψ^A into as few qubits as possible and send it over a quantum channel. However, this ignores the information Bob has about the state in the form of ψ^B. Although it might seem that a quantum channel is essential for state merging to work, Bob's side information can be such that only *classical* communication from Alice is required.

Reasoning about the protocol is made somewhat easier by considering the purification $|\psi\rangle^{ABR}$ of ψ^{AB} to a reference system R, that is $\psi^{AB} = \mathrm{Tr}_R\left[\psi^{ABR}\right]$. The goal of state merging is then to arrange for Bob to hold the purification of R. In some cases quantum communication will clearly be required, for instance when $|\psi\rangle^{ABR} = |\Phi\rangle^{AR}|\xi\rangle^B$, where $|\xi\rangle$ is arbitrary while $|\Phi\rangle^{AR} = \frac{1}{\sqrt{d}}\sum_k |k,k\rangle^{AR}$ is the canonical maximally entangled state for a fixed basis $\{|k\rangle\}$ and d is the minimum dimension of A and R. Bob's state is clearly irrelevant, and Alice must simply send her whole system, as it is incompressible. On the other hand, when Alice and Bob share $|\Phi\rangle^{AB}$, no communication is required at all! This is simply due to the fact that now the state of R is by itself pure, so neither Alice nor Bob hold its purification.

Horodecki, Oppenheim, and Winter [1,2] consider the asymptotic setting of many copies of ψ^{ABR} and show that classical communication suffices when the quantum conditional entropy $S(A|B) = S(AB) - S(B)$ is negative, where $S(A) = -\mathrm{Tr}\left[\rho^A \log_2 \rho^A\right]$ is the von Neumann entropy. In fact, when $S(A|B) < 0$ their state merging protocol produces entangled pairs at the rate $-S(A|B)$ and uses classical communication at the rate $I(A{:}R)$, where $I(A{:}R) = S(A) + S(R) -$

A. Childs and M. Mosca (Eds.): TQC 2009, LNCS 5906, pp. 76–84, 2009.
© Springer-Verlag Berlin Heidelberg 2009

$S(AR)$ is the quantum mutual information. These rates are also shown to be optimal. When $S(A|B) > 0$ on the other hand, any state merging protocol requires quantum communication at the rate $S(A|B)$, or equivalently consumes entangled pairs at this rate. This fact gives an operational meaning to the conditional entropy in terms of entanglement consumption or production, which due to its possible negativity is quite unlike its classical counterpart.

In this paper we construct a state merging protocol operating at the optimal rates by focusing on the classical information that Bob has about complementary observables "amplitude" and "phase" on Alice's system and showing how classical communication is sufficient to transfer the necessary quantum correlations. This approach is substantially different from the original proof, which is based on the technique of decoupling Alice's system from the reference system R [3], and follows our recent work on entanglement distillation (ED) quite closely [4]. Indeed, state merging is actually achieved in that protocol as well, but at the cost of too much classical communication. We rectify this problem here, showing that if Alice first compresses her system and then runs the ED protocol, a small modification suffices to make this an optimal state merging protocol.

The remainder of the paper is outlined as follows. We first review the known results for the state merging protocol in the next section, and then recapitulate the important parts of the proof of the ED protocol appearing in [4] in the following section. Section 4 contains the new contribution of this paper, showing how to modify the ED protocol to use only the minimum necessary classical communication. Finally, we conclude with a summary of the results and comment on the connections to the quantum noisy channel coding theorem.

2 State Merging Defined

As with most protocols in quantum information theory, we are concerned here with the rate at which Alice and Bob can transform an asymptotically-large number of copies of the state $|\psi\rangle^{ABR}$ into a good approximation of n copies in which Bob holds system A. To keep the accounting simple, we assume that any necessary quantum communication is performed by teleportation through pre-shared entangled pairs, so that the protocol uses only classical communication in any case, and either produces or consumes entanglement depending on the circumstances. We then define an (n, ϵ) state merging protocol for ψ^{ABR} to be a series of local operations involving only classical communication (LOCC operations) such that application to $|\Psi\rangle^{ABR} = (|\psi\rangle^{ABR})^{\otimes n}$ produces an output Υ^{DBR} in which Bob holds the system D such that $||\Upsilon^{DBR} - \Psi^{DBR}||_1 \leq \epsilon$. If there exists an (n, ϵ_n) protocol using K_n bits of classical communication and consuming E_n ebits of entanglement for every n such that $\lim_{n\to\infty} \epsilon_n = 0$, then the rates of communication and entanglement consumption of the protocol are given by

$$R_K = \lim_{n\to\infty} \frac{K_n}{n} \qquad \text{and} \qquad R_E = \lim_{n\to\infty} \frac{E_n}{n}. \tag{1}$$

Horodecki, Oppenheim, and Winter showed in [1,2] that

$$\inf R_K = I(A{:}R) \qquad \text{and} \qquad \inf R_E = S(A|B), \tag{2}$$

where a negative R_E indicates the amount of entanglement produced. The proof of these statements has two parts, the direct part showing the rates are achievable, and the converse part showing they cannot be surpassed. Here we will give a new proof of the direct part, borrowing our techniques from [4] which were used to give a new proof of the hashing inequality [5] on the achievable rate of entanglement distillation. In the next section we sketch the important parts of that proof.

3 Entanglement Distillation Revisited

A maximally entangled pair in one for which Bob can predict the measurement of either of the two observables, "amplitude" $Z^A = \sum_k (-1)^k |k\rangle\langle k|^A$ and its Fourier conjugate "phase" $X = \sum_k |k \oplus 1\rangle\langle k|$. Here we are assuming that Alice's system has dimension 2, but what follows can be easily extended to higher dimensions. Since this is the desired output of the distillation procedure, the idea behind the protocol given in Theorem 6 of [4] is to determine what information Bob already has about these observables from his system B and then arrange for Alice to send him the rest. This is classical information, since it refers to the measurement outcomes, and therefore only classical communication will be required. However, since Alice needs to send information pertaining to both X and Z, one must ensure that both parts of her message simultaneously exist. This is achieved by measuring the X- and Z-type stabilizers of a Calderbank-Shor-Steane (CSS) code [6,7,8] to generate the message. The amount of information is governed by the "static" version of the Holevo-Schumacher-Westmoreland (HSW) theorem [9,10], which solves the problem of classical data compression with quantum side information [11] and which we review in the appendix.

Greatly simplified, the protocol starts by Alice picking a random CSS code of a given size for her Hilbert space. She then measures the stabilizers to obtain the syndromes α (for X) and β (for Z) and communicates them to Bob. The syndromes are such that he can find measurements $\Lambda^B_{\alpha,\mathbf{x}}$ and $\Gamma^B_{\beta,\mathbf{z}}$ on B which enable him to predict (with high probability) the outcome of measuring either X^A or Z^A, respectively. The existence of such measurements is guaranteed by the (static) HSW theorem, using Bob's marginal states generated by Alice's measurement as the ensemble and the code syndrome as the side information. It implies that the CSS code must have roughly $m_Z = nS(Z^A|B)$ Z-type syndromes and $m_X = nS(X^A|CB)$ X-type, where C is an additional quantum register containing a copy of Alice's system in the Z basis, and $S(Z^A|B) = S(\bar{\psi}^{AB}_Z) - S(\psi^B)$ for $\bar{\psi}^{AB}_Z$ the shared state after Alice measures the observable Z. Once this process is complete, Bob can (in principle) predict either X^A or Z^A on each pair, and therefore can perform a quantum operation on his systems to create entangled pairs (to good approximation). Since Alice is left with only the code subspace

given by α and β, whose size is $n - m_X - m_X$, this is the number of entangled pairs they can create.

To see how this works in more detail, begin with the individual shared state $|\psi\rangle^{ABR}$ and write it as $|\psi\rangle^{ABR} = \sum \sqrt{p_k}|k\rangle^A|\varphi_k\rangle^{BR}$, where $|k\rangle$ is the eigenbasis of ψ^A and also defines the operator Z, the $|\varphi_k\rangle$ are a set of arbitrary orthonormal states, and p_k is a probability distribution. The n-fold version $|\Psi_0\rangle^{ABR} = (|\psi\rangle^{ABR})^{\otimes n}$ we write like so, using bold-faced symbols $\mathbf{k}$ to denote strings $(k_1, k_2, \ldots, k_n)$:

$$|\Psi_0\rangle^{ABR} = \sqrt{p_{\mathbf{k}}} \sum_{\mathbf{k}} |\mathbf{k}\rangle^A |\varphi_{\mathbf{k}}\rangle^{BR}. \tag{3}$$

We'll also need to consider the associated state in which Bob has a copy of Alice's system in the Z basis:

$$|\psi_c\rangle^{ACBR} = \sum_k \sqrt{p_k}|k,k\rangle^{AC}|\varphi_k\rangle^{BR} = \tfrac{1}{\sqrt{2}} \sum_x |\widetilde{x}\rangle^A |\vartheta_x\rangle^{CBR}.$$

Here $|\widetilde{x}\rangle$ is an eigenstate of X and the $|\vartheta_x\rangle$ are again a arbitrary set of orthonormal states. Observe that $|\vartheta_0\rangle^{CBR} = |\psi\rangle^{CBR}$.

Denote the projections onto the stabilizers of the chosen CSS code by $\widetilde{\Pi}_\alpha^A$ and Π_β^A, which commute by the CSS nature of the code. The result of Alice measuring the stabilizers and sending them to Bob is

$$|\Psi_1\rangle^{ABRP} = \sum_{\alpha,\beta} \widetilde{\Pi}_\alpha^A \Pi_\beta^A |\Psi_0\rangle^{ABR} |\alpha,\beta\rangle^P. \tag{4}$$

The system label P, for "public", is shorthand for having arbitrarily many copies $P_1, P_2, \ldots$ of the values α, β, and mimics the information being classically-transmitted. Given β, Bob can coherently perform the measurement $\Gamma_{\beta,\mathbf{k}}^B$ to extract the value of k in A to an auxiliary system C with high probability. One can show that this implies the state is very nearly identical to

$$|\Psi_2\rangle = \sum_{\alpha,\beta,\mathbf{k}} \widetilde{\Pi}_\alpha^A \Pi_\beta^A |\mathbf{k}\mathbf{k}\rangle^{AC} |\varphi_{\mathbf{k}}\rangle^{BR} |\alpha,\beta\rangle^P = \sum_{\alpha,\beta} \widetilde{\Pi}_\alpha^A \Pi_\beta^A |\Psi_c\rangle |\alpha,\beta\rangle^P. \tag{5}$$

Next, Bob can coherently measure $\Lambda_{\alpha,\mathbf{x}}^B$ to extract $\mathbf{x}$ in the conjugate basis of A to a further auxiliary system D, again with high probability. The resulting state is nearly identical to

$$|\Psi_3\rangle = \tfrac{1}{\sqrt{2^n}} \sum_{\alpha,\beta,\mathbf{x}} \widetilde{\Pi}_\alpha^A \Pi_\beta^A |\widetilde{\mathbf{x}}\rangle^A |\widetilde{\mathbf{x}}\rangle^D |\vartheta_{\mathbf{x}}\rangle^{CBR} |\alpha,\beta\rangle^P. \tag{6}$$

Owing to the properties of X and Z and the two forms of $|\psi_c\rangle$, we have the relation $|\vartheta_{\mathbf{x}}\rangle^{CBR} = \sum_{\mathbf{k}} \sqrt{p_{\mathbf{k}}}\, \omega^{\mathbf{k}\cdot\mathbf{x}} |\mathbf{k}\rangle^C |\varphi_{\mathbf{k}}\rangle^{BR} = (Z^{\mathbf{x}})^C |\Psi_0\rangle^{CBR}$. Inserting this into equation 6 gives

$$|\Psi_3\rangle = \tfrac{1}{\sqrt{2^n}} \sum_{\alpha,\beta,\mathbf{x}} \widetilde{\Pi}_\alpha^A \Pi_\beta^A |\widetilde{\mathbf{x}}\rangle^A |\widetilde{\mathbf{x}}\rangle^D (Z^{\mathbf{x}})^C |\Psi_0\rangle^{CBR} |\alpha,\beta\rangle^P. \tag{7}$$

Finally, a controlled-Z operation from D to C inverts the $Z^{\mathbf{x}}$ operator, leaving the desired output

$$|\Psi_4\rangle = \sum_{\alpha,\beta,\mathbf{x}} \widetilde{\Pi}^A_\alpha \Pi^A_\beta |\Phi_n\rangle^{AD} |\alpha,\beta\rangle^P \otimes |\Psi_0\rangle^{CBR}, \tag{8}$$

where $|\Phi_n\rangle = |\Phi\rangle^{\otimes n}$. Observe that the purification of R is now solely in Bob's possession, so state merging has been accomplished. Furthermore, since $n[S(Z^A|B)+S(X^A|CB)]$ CSS stabilizers leave $n[1-S(Z^A|B)-S(X^A|CB)]$ encoded logical operators, Alice and Bob share this many entangled pairs in systems A and D. In [4] it is shown that this equals $-nS(A|B)$, so provided this quantity is positive ($S(A|B) < 0$), the protocol achieves the rate R_E.

Of course, $|\Psi_4\rangle$ is not precisely the output of the protocol, since the two coherent measurement operations by Bob were not perfect. The details of the approximation are given in [4], the result being that if Alice chooses a random code having $n[S(Z^A|B)+\delta]$ Z-type stabilizers and $n[S(X^A|CB)+\delta]$ X-type stabilizers for some $\delta > 0$, then the output will be within $\exp(-O(n\delta^2))$ of $|\Psi_4\rangle$, as measured by the trace-distance.

If $S(A|B) > 0$, we can use the same trick as [1,2]. Adding $n[S(A|B)+2\delta]$ entangled pairs, each of which has $S(A|B) = -1$, the conditional entropy of the overall state $|\Psi\rangle^{ABR}|\Phi_{n[S(A|B)+2\delta]}\rangle^{A'B'}$ is $-2n\delta$. Using this as the individual input into the above protocol accomplishes the state merging and outputs no entanglement. In this way R_E can be achieved when $S(A|B) > 0$.

The above protocol requires too much classical communication, however, $n[1-S(A|B)]$ bits. This is generally greater than $I(A{:}E)$, and is only equal for $S(A) = 1$. The fact that the protocol is optimal when ψ^A is maximally mixed suggests that for a general input Alice should first compress her system and then run the protocol. However, the compression procedure will disturb the conjugate observable X and its eigenbasis, so there is no longer any guarantee that Bob's $\Lambda_{\alpha,\mathbf{x}}$ measurement will work as intended. The next section shows how to fix this problem.

4 Classical Communication Reduced

Fortunately, the ensemble of states $\vartheta^{CB}_{\mathbf{x}}$ which Bob would like to distinguish is invariant under the action of the group $(Z^{\mathbf{x}})^C$, which will enable us to adapt the original $\Lambda_{\alpha,\mathbf{x}}$ measurement for use after Alice compresses her state. This will reduce the number of X syndromes she needs to communicate to Bob to the optimal level.

The modified protocol begins as before with the state $|\Psi_0\rangle$. Alice then makes a measurement projecting her systen onto the typical subspace $\mathcal{T}^n_\delta$, which is the subspace spanned by eigenvectors $|\mathbf{k}\rangle$ whose $\mathbf{k}$ are in the typical set $T^n_\delta = \{\mathbf{k} : |-\frac{1}{n}\log p_{\mathbf{k}} - S(\psi^A)| \leq \delta\}$ for a fixed $\delta > 0$ [12,13]. The probability $\mathcal{N}^n_\delta = \Pr[\mathbf{k} \in T^n_\delta]$ that $\mathbf{k}$ is typical is greater than $1-2^{-cn\delta^2} := 1-\epsilon$, for some constant c [5] and

therefore the projection succeeds with probability exponentially close to unity; otherwise the protocol aborts. When it succeeds, it *prunes* the state $|\Psi_0\rangle$, leaving

$$|\Psi_0'\rangle^{ABR} = \frac{1}{\sqrt{\mathcal{N}_\delta^n}} \sum_{\mathbf{k}\in T_\delta^n} \sqrt{p_{\mathbf{k}}}|\mathbf{k}\rangle^A|\varphi_{\mathbf{k}}\rangle^{BR} = \sum_{\mathbf{k}\in T_\delta^n} \sqrt{p'_{\mathbf{k}}}|\mathbf{k}\rangle^A|\varphi_{\mathbf{k}}\rangle^{BR}, \tag{9}$$

where we have implicitly defined new probability weights $p'_{\mathbf{k}} = p_{\mathbf{k}}/\mathcal{N}_\delta^n$. Importantly, $D_\delta^n := \dim(\mathcal{T}_\delta^n) \leq 2^{n[S(\psi^A)+\delta]}$, and a simple calculation shows that $\langle\Psi_0|\Psi_0'\rangle = \sqrt{\mathcal{N}_\delta^n}$. This implies that two states are close in trace distance, $||\Psi_0 - \Psi_0'||_1 \leq \sqrt{\epsilon}$, using the relationship between fidelity and trace distance $||\rho-\sigma||_1 \leq \sqrt{1-F(\rho,\sigma)^2}$ [14].

The protocol proceeds just as before, measuring X'- and Z'-type stabilizers of a random CSS code on the pruned state and communicating the results to Bob. Here Z' is the analog of Z for the typical subspace, and X' is its Fourier conjugate. Now, however, we have no direct way of setting the number of stabilizers, since the state is no longer i.i.d. and therefore the HSW theorem no longer applies. This is not really a problem for the Z'-type stabilizers, since the typical projection is done in the $|\mathbf{k}\rangle$ basis, the basis which generates the $\varphi_{\mathbf{k}}^B$. By design, the measurement constructed in the HSW theorem does not attempt to identify $\varphi_{\mathbf{k}}^B$ for nontypical $\mathbf{k}$, so Bob can just reuse it in this case. The probability of error will only decrease by *explicitly* rejecting nontypical $\mathbf{k}$. Hence $m_z \approx nS(Z^A|B)$ as before.

However, the original measurement will not work for the conjugate basis $|\mathbf{x}'\rangle$, the Fourier transform of the typical subspace basis, since the states $\vartheta_{\mathbf{x}'}'^{CB}$ have no *a priori* relation to the original $\vartheta_{\mathbf{x}}^{CB}$. However, the former states stem from the related state

$$|\Psi_c'\rangle = \sum_{\mathbf{k}\in T_\delta^n} \sqrt{p'_{\mathbf{k}}}|\mathbf{k}\mathbf{k}\rangle^{AC}|\varphi_{\mathbf{k}}\rangle^{BR} = \frac{1}{\sqrt{D_\delta^n}} \sum_{\mathbf{x}'} |\widetilde{\mathbf{x}}'\rangle^A|\vartheta'_{\mathbf{x}'}\rangle^{CBR}, \tag{10}$$

and this fact, coupled with the group covariance of both sets, gives us a means to transform $\Lambda_{\alpha,\mathbf{x}}^{CB}$ into a measurement $\Lambda_{\alpha,\mathbf{x}'}'^{CB}$ suitable for distinguishing the $\vartheta_{\mathbf{x}'}'^{CB}$.

To see how this works, it is easiest to go back to the proof of the HSW theorem, which for convenience is stated in the appendix. In the original i.i.d. case, the projectors $P_{\mathbf{x}}$ and P^{CB} onto the typical subspaces of $\vartheta_{\mathbf{x}}^{CB}$ and $\bar{\vartheta}^{CB} = \frac{1}{2^n}\sum_{\mathbf{x}} \vartheta_{\mathbf{x}}^{CB}$, respectively, fulfill the five conditions needed in the proof of the theorem, equations 17 through 21. Since $\vartheta_{\mathbf{x}}^{CB} = (Z^{\mathbf{x}})^C \Psi_0^{CB} (Z^{\mathbf{x}})^C$, the same holds for $P_{\mathbf{x}}^{CB}$, and the five conditions become

$$\mathrm{Tr}[\bar{\vartheta}^{CB}(\mathbb{1}^{CB} - P^{CB})] \leq \epsilon \tag{11}$$

$$\mathrm{Tr}[\Psi_0^{CB}(\mathbb{1}^{CB} - P_0^{CB})] \leq \epsilon \tag{12}$$

$$P_0^{CB} \leq r \cdot \Psi_0^{CB} \tag{13}$$

$$\sum_{\mathbf{x}} \vartheta_{\mathbf{x}}^{CB} \leq d \cdot \bar{\vartheta}^{CB} \tag{14}$$

$$||P^{CB}\bar{\vartheta}^{CB}P^{CB}||_\infty \leq \lambda, \tag{15}$$

with $\epsilon =$, $r = 2^{n[S(\psi^{CB})+\delta]}$, $d = 2^n$ (and the condition is an equality since all $\mathbf{x}$ are typical), $\lambda = 2^{-n[S(\vartheta^{CB})-\delta]}$. Our aim is now to find a set of new projectors $P'^{CB}_{\mathbf{x}}$ and P'^{CB} fulfilling these conditions for the states $\vartheta'^{CB}_{\mathbf{x}'}$ and $\bar{\vartheta}'^{CB} = \frac{1}{D^n_\delta}\sum_{\mathbf{x}'} \vartheta'^{CB}_{\mathbf{x}'}$.

To start, use the fact that $\mathrm{Tr}[(\Psi'^{CB}_0 - \Psi^{CB}_0)P^{CB}_{\mathbf{0}}] \leq ||\Psi'^{CB}_0 - \Psi^{CB}_0||_1 \leq \sqrt{\epsilon}$, since the trace distance is equal to the maximum of the lefthand side, maximized over all projectors [8]. Then we have

$$\mathrm{Tr}\left[(\mathbb{1} - P^{CB}_{\mathbf{0}})\Psi'^{CB}_0\right] \leq \mathrm{Tr}\left[(\mathbb{1} - P^{CB}_{\mathbf{0}})\Psi^{CB}_0\right] + ||\Psi'^{CB}_0 - \Psi^{CB}_0||_1 \leq \epsilon + \sqrt{\epsilon},$$

and so we can define $P^{CB}_{\mathbf{x}'} = (Z'^{\mathbf{x}})^C P^{CB}_0 (Z^{\mathbf{x}})^C$ to satisfy the first condition. The second condition follows analogously upon noting that $\bar{\vartheta}^{CB} = \sum_{\mathbf{k}} p_{\mathbf{k}} |\mathbf{k}\rangle\langle\mathbf{k}|^C \otimes \varphi^B_{\mathbf{k}}$ (and similarly for the pruned version) and therefore $||\bar{\vartheta}' - \bar{\vartheta}|| \leq 2(1 - \mathcal{N}^n_\delta) \leq 2\epsilon$. The third condition remains as is, since we're using the same $P_{\mathbf{0}}$, and the fourth is an equality when $d = D^n_\delta$. For the fifth condition, observe that

$$\tfrac{1}{\mathcal{N}^n_\delta}\bar{\vartheta}^{CB} - \bar{\vartheta}'^{CB} = \tfrac{1}{\mathcal{N}^n_\delta} \sum_{\mathbf{k}\notin T^n_\delta} p_{\mathbf{k}} |\mathbf{k}\rangle\langle\mathbf{k}|^C \otimes \varphi^B_{\mathbf{k}} \geq 0.$$

Therefore, $P^{CB}\bar{\vartheta}'^{CB}P^{CB} \leq \frac{1}{\mathcal{N}^n_\delta} P^{CB}\bar{\vartheta}^{CB}P^{CB}$, which leads immediately to $||P^{CB}\bar{\vartheta}'^{CB}P^{CB}||_\infty \leq \lambda/\mathcal{N}^n_\delta \leq \lambda(1+2\epsilon)$.

We thus have all the ingredients needed to construct the required measurement, with $\epsilon' = 2\sqrt{\epsilon}$, $r' = r$, $d' = D^n_\delta$, and $\lambda' = \lambda(1+2\epsilon)$. The number of syndromes Bob needs from Alice is given by $m'_X \geq n[S(\psi^{AB}) + S(\psi^A) - S(\vartheta^{CB}) + 3\delta] + \log(1+2\epsilon)$, which works out to be $m'_X \approx n[S(\psi^R) - \sum_k p_k S(\varphi^R_k)]$. Since the pruned state is nearly identical to the original state, the remainder of the protocol goes through as before, outputting roughly $nS(A) - m_Z - m'_X$ entangled pairs. A simple calculation (along the lines of lemma 2 in [4]) gives $m'_X + m_Z = I(A{:}E)$ and $nS(A) - m'_X - m_Z = -S(A|B)$, and thus the protocol is optimal.

5 Conclusion

We have shown how to construct an optimal state merging protocol by following the intuition from quantum error-correction that what really matters in two-party quantum information processing is information about amplitude and phase measurements. Combining entanglement distillation with teleportation, our results also imply a new proof of the direct part of the noisy channel coding theorem [5], one not following the usual route of decoupling Alice's system from the purification R (e.g. all the fully fleshed-out proofs to date [15,16,17,18,19]). It would be interesting to apply these techniques to more protocols, and see how far this intuition about quantum information extends.

Acknowledgments. JMR received support from the European IST project SECOQC and JCB from Quantumworks and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Horodecki, M., Oppenheim, J., Winter, A.: Nature 436(7051), 673–676 (2005)
2. Horodecki, M., Oppenheim, J., Winter, A.: Communications in Mathematical Physics 269(1), 107–136 (2007)
3. Schumacher, B., Westmoreland, M.D.: Quantum Information Processing, vol. 1, pp. 5–12 (2002)
4. Renes, J.M., Boileau, J.-C.: Physical Review A. 78(3), 032335–12 (2008)
5. Devetak, I., Winter, A.: Proceedings of the Royal Society A 461(2053), 207–235 (2053)
6. Calderbank, A.R., Shor, P.W.: Physical Review A 54(2), 1098 (1996)
7. Steane, A.: Proceedings of the Royal Society A 452(1954), 2551–2577 (1996)
8. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)
9. Holevo, A.: IEEE Transactions on Information Theory 44(1), 269–273 (1998)
10. Schumacher, B., Westmoreland, M.D.: Physical Review A 56(1), 131 (1997)
11. Devetak, I., Winter, A.: Physical Review A 68(4), 042301 (2003)
12. Schumacher, B.: Physical Review A 51(4), 2738 (1995)
13. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley-Interscience, Hoboken (2006)
14. Fuchs, C., van de Graaf, J.: IEEE Transactions on Information Theory 45(4), 1216–1227 (1999)
15. Devetak, I.: IEEE Transactions on Information Theory 51(1), 44–55 (2005)
16. Hayden, P., Horodecki, M., Winter, A., Yard, J.: Open Systems & Information Dynamics 15(1), 7–19 (2008)
17. Klesse, R.: Open Systems & Information Dynamics 15(1), 24–45 (2008)
18. Horodecki, M., Lloyd, S., Winter, A.: Open Systems & Information Dynamics 15(1), 47–69 (2008)
19. Hayden, P., Shor, P.W., Winter, A.: Open Systems & Information Dynamics 15(1), 71–89 (2008)
20. Carter, J.L., Wegman, M.N.: Journal of Computer and System Sciences 18(2), 143–154 (1979)
21. Hsieh, M., Devetak, I., Winter, A.: IEEE Transactions on Information Theory 54(7), 3078–3090 (2008)

A Static HSW Theorem

Here we are interested in the "static" setting of the HSW theorem, which is concerned with the following. Given n samples from an ensemble $\{p_k, \rho_k\}_{k=1}^d$ with average $\rho = \sum_k p_k \rho_k$, what is the smallest amount of side information $t = f(\mathbf{k})$ required in order to reliably construct a measurement $\Lambda_{t,\mathbf{k}}$ which will identify $\mathbf{k}$ from $\rho_{\mathbf{k}}$ with only a small probability of error? In order to match the setting in the main text, we can think of the ensemble as arising from the state $\psi^{AB} = \sum_k p_k |k\rangle\langle k|^A \otimes \rho_k^B$, a measurement of $|k\rangle$ (or Z^A) on A generating state ρ_k. For random CSS codes f is a random linear function, resulting from measuring the stabilizer observables on the state $|\mathbf{k}\rangle$. However, in what follows we will consider *universal hashing* [20], since it is no more difficult to do so. In universal (or 2-universal) hashing, the function $f : \{0,1\}^n \to \{0,1\}^m$ generating

the side information is chosen at random from a universal family of hash functions in which the probability of collision $f(x) = f(y)$ but $x \neq y$ is the same as for random functions: $\Pr_f[f(x) = f(y)|x \neq y] \leq 1/2^m$.

In [4] we proved that for a fixed $\delta > 0$, choosing $m = n[S(Z^A|B) + 4\delta]$ is sufficient to guarantee the existence of a measurement having elements $\Lambda_{f(\mathbf{k}),\ell}$ such that the probability of error P_e is exponentially small (see also [11]):

$$\mathrm{P}_e = \left\langle \sum_{\ell \neq \mathbf{k}} \left[\Lambda_{f(\mathbf{k}),\ell} \rho_{\mathbf{k}}\right] \right\rangle_{f,\mathbf{k}} \leq 6 \times 2^{-n\delta^2/2}. \tag{16}$$

A crucial step in the proof is to show the existence of projectors $Q_{\mathbf{k}}$ and Q such that

$$\mathrm{Tr}[\langle \rho_{\mathbf{k}} \rangle_{\mathbf{k}} (\mathbb{1} - Q)] \leq \epsilon \tag{17}$$

$$\langle \mathrm{Tr}[\rho_{\mathbf{k}} (\mathbb{1} - Q_{\mathbf{k}})] \rangle_{\mathbf{k}} \leq \epsilon \tag{18}$$

$$Q_{\mathbf{k}} \leq r \cdot \rho_{\mathbf{k}} \tag{19}$$

$$\sum_{k \in T_\delta^n} \rho_{\mathbf{k}} \leq d \cdot \langle \rho_{\mathbf{k}} \rangle_{\mathbf{k}} \tag{20}$$

$$||Q \langle \rho_{\mathbf{k}} \rangle_{\mathbf{k}} Q||_\infty \leq \lambda, \tag{21}$$

after which it can be shown that $m \geq \lfloor \frac{1}{\gamma} \log r d \lambda \rfloor$ for $0 \leq \gamma \leq 1$ suffices to construct the measurement.[1] In the i.i.d. case of the HSW theorem, the $Q_{\mathbf{k}}$ and Q are projectors onto the typical subspaces of $\rho_{\mathbf{k}}$ (for typical $\mathbf{k}$) and $\rho^{\otimes n}$, respectively, for which $\epsilon = 2^{-cn\delta^2}$, $r = 2^{n[\sum_k p_k S(\rho_k) + \delta]}$, $d = 2^{n[H(p_k) + \delta]}$, and $\lambda = 2^{-n[S(\rho) - \delta]}$. Thus, one chooses $m \geq n[H(p_k) - S(\rho) + \sum_k p_k S(\rho_k) + 4\delta] = n[S(Z^A|B) + 4\delta]$.

[1] Breaking up the proof in this way is similar to the packing lemma of [21].