On the Distributed Compression of Quantum Information

Charlene Ahn, Andrew C. Doherty, Patrick Hayden, *Member, IEEE*, and Andreas J. Winter, *Member, IEEE*

***Abstract*—The problem of distributed compression for correlated quantum sources is considered. The classical version of this problem was solved by Slepian and Wolf, who showed that distributed compression could take full advantage of redundancy in the local sources created by the presence of correlations. Here it is shown that, in general, this is not the case for quantum sources, by proving a lower bound on the rate sum for irreducible sources of product states which is stronger than the one given by a naive application of Slepian–Wolf. Nonetheless, strategies taking advantage of correlation do exist for some special classes of quantum sources. For example, Devetak and Winter demonstrated the existence of such a strategy when one of the sources is classical. Optimal nontrivial strategies for a different extreme, sources of Bell states, are presented here. In addition, it is explained how distributed compression is connected to other problems in quantum information theory, including information-disturbance questions, entanglement distillation and quantum error correction.**

***Index Terms*—Compression, distributed, quantum information, Slepian–Wolf.**

I. Introduction

THE insights that have come from efforts to study quantum mechanics from an information-theoretic point of view are profound and wide-ranging, demonstrating that quantum information can be compressed [1], [2], stabilised [3] and usefully processed [4]. Schumacher's theorem [1], [2], [5], in particular, demonstrated the fungibility of quantum states by quantifying their compressibility, justifying the use of the *qubit* as the fundamental unit of quantum information.

In this paper we consider a distributed variant of the problem posed by Schumacher. Namely, we suppose that a source distributes quantum states to two or more parties, who independently compress the states before sending them on to a receiver, who is required to be able to reconstruct the original inputs. Since many ideas for the design of quantum computers and other quantum information processing devices envision a network of relatively small quantum processors sending quantum information between nodes [6], [7], finding good protocols for distributed compression of quantum data could conceivably have important practical benefits. More generally, much of quantum information theory is concerned with the manipulation of data under locality constraints [8], so our problem connects naturally to these investigations.

We present two main results. First, we show that, in stark contrast to the classical case, independent encoders frequently can take relatively little advantage of the correlations present between their states: we prove this via a bound on the achievable rate sum for sources generating irreducible sets of product vectors. On the other hand, it is possible to do much better for some special classes of sources. We show, in particular, that for sources of Bell states, independent encoders *can* take full advantage of correlations. The achievable rates, however, are governed by different formulas than in the classical case, reflecting the quantum nature of the correlations in the input states.

The paper is structured as follows. Section II gives a formal definition of the distributed compression problem and shows how questions about cloning, imprinting [9] and quantum error correction can be formulated in that framework. It also gives a statement of the classical theorem governing distributed compression due to Slepian and Wolf before summarizing previous work on the quantum version. Section III contains the statement and proof of our tighter bound for irreducible sources of product states. Section IV finds the achievable rate region for sources generating Bell states. Section V then provides some further examples, where it seems likely that the optimal rates lie somewhere between full utilization of correlations and no utilization at all. We end with a discussion and some open problems.

We use the following conventions throughout the paper. If $\mathcal{E}_{AB} = \{p_i, \varphi_i^{AB}\}$ is an ensemble of bipartite states then we write $\mathcal{E}_A$ for the ensemble $\{p_i, \varphi_i^A\}$ of reduced states on system A. Sometimes we omit subscripts (or superscripts) labeling subsystems, in which case the largest subsystem on which the ensemble (or state) has been defined should be assumed: $\mathcal{E} = \mathcal{E}_{AB}$ and $\varphi_i = \varphi_i^{AB}$. We identify states with their density operators and if $|\varphi\rangle$ is a pure state vector, we use the notation $\varphi = |\varphi\rangle\langle\varphi|$ for its density operator. The function $S(\rho)$ is the von Neumann entropy $S(\rho) = -\mathrm{Tr}\rho \log \rho$ and $S(\mathcal{E})$ the von Neumann entropy of the average state of the ensemble $\mathcal{E}$. Functions like $S(A|B)_\rho$

Manuscript received September 24, 2004; revised September 19, 2005. This work was supported by the U.S. National Science Foundation under Grant EIA-0086038. The work of C. Ahn was also supported by an Institute for Quantum Information Fellowship; the work of P. Hayden was supported by the Sherman Fairchild Foundation, CIAR, NSERC and the Canada Research Chairs Program; the work of A. C. Doherty was supported by the Caltech MURI Center for Quantum Networks (DAAD19-00-1-0374); and the work of A. J. Winter was supported by the U.K. Engineering and Physical Sciences Research Council.

C. Ahn was with the Institute for Quantum Information at the California Institute of Technology (Caltech), Pasadena, CA 91125 USA. She is now with the Toyon Research Corporation, Goleta, CA 93117 USA (e-mail: cahn@toyon.com).

A. C. Doherty is with the School of Physical Sciences, University of Queensland, Brisbane 4072, Australia (e-mail: doherty@physics.uq.edu.au).

P. Hayden is with the School of Computer Science, McGill University, Montreal, QC H3A 2A7, Canada (e-mail: patrick@cs.mcgill.ca).

A. J. Winter is with the Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K. (e-mail: a.j.winter@bris.ac.uk).

Communicated by E. Knill, Associate Editor for Quantum Information Theory.

Digital Object Identifier 10.1109/TIT.2006.881734

and $S(A:B|C)_\rho$ are defined in the same way as their classical counterparts:

$$S(A:B|C)_\rho = S(\rho^{AC}) + S(\rho^{BC}) - S(\rho^{ABC}) - S(\rho^{C}), \tag{1}$$

for example. $\chi(\mathcal{E})$ is the Holevo χ quantity of $\mathcal{E}$ [10]. Throughout, log and exp are taken base 2.

II. *Definition and Examples*

We now give a more formal definition of the distributed compression problem. For convenience, our definition will refer to the case of two encoders, henceforth known as Alice and Bob. The extension to any finite number of parties is straightforward. Our receiver will be named Charlie. Consider an ensemble of bipartite quantum states $\mathcal{E}_{AB} = \{p_i, |\varphi_i\rangle^{AB}\}$ on a finite-dimensional Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ and the product ensemble $\mathcal{E}^{\otimes n} = \{p_{i^n}, |\varphi_{i^n}\rangle^{AB}\}$ on $\mathcal{H}_{AB}^{\otimes n}$, where

$$\begin{aligned} i^n &= i_1 i_2 \ldots i_n, \\ p_{i^n} &= p_{i_1} p_{i_2} \ldots p_{i_n} \quad \text{and} \\ |\varphi_{i^n}\rangle &= |\varphi_{i_1}\rangle \otimes |\varphi_{i_2}\rangle \otimes \cdots \otimes |\varphi_{i_n}\rangle. \end{aligned}$$

A source provides Alice and Bob with the state $|\varphi_{i^n}\rangle$, drawn with probability p_{i^n}. Alice and Bob then perform their respective encoding operations E_A and E_B. These are quantum operations, that is, completely positive, trace-preserving (CPTP) maps, with outputs on quantum systems C_A and C_B of dimensions d_A and d_B, respectively. The joint encoding operation is $E_A \otimes E_B$ since Alice and Bob are required to act independently. The systems C_A and C_B are then sent to Charlie, who performs a decoding operation D, again a CPTP map, producing the output state $\tilde{\varphi}_{i^n} = D \circ (E_A \otimes E_B)(\varphi_{i^n})$. We say the *encoding-decoding scheme* has fidelity $1 - \epsilon$ if

$$\sum_{i^n} p_{i^n} \langle \varphi_{i^n} | \tilde{\varphi}_{i^n} | \varphi_{i^n} \rangle \geq 1 - \epsilon \tag{2}$$

and that (R_A, R_B) is an achievable rate pair if for all $\delta, \epsilon > 0$ there exists an integer N such that for all $n > N$ there is an encoding–decoding scheme with fidelity $1 - \epsilon$ satisfying

$$\frac{1}{n} \log d_A \leq R_A + \delta \quad \text{and} \quad \frac{1}{n} \log d_B \leq R_B + \delta. \tag{3}$$

This scenario is formulated in analogy to the asymptotically lossless setting of classical block compression, as opposed to lossless variable-length coding.

We remark here that we may easily allow Alice and Bob the use of prior shared randomness without affecting any of our conclusions. Indeed, randomness is unnecessary, as a look at the fidelity criterion (2) shows: the fidelity is an ensemble expectation of quantities linear in the output state $\tilde{\varphi}_{i^n}$. Hence the fidelity of a randomized scheme, regardless of whether it uses shared or private randomness, is the average of fidelities of the schemes obtained by picking particular instances of the random data. So, at least one of the randomness-free schemes has a fidelity at least as good as the randomized version.

The classical correlated source compression problem has a beautiful solution, due to Slepian and Wolf [11]. This remarkable theorem shows that Alice and Bob can *always* take advantage of any correlations that exist between their data.

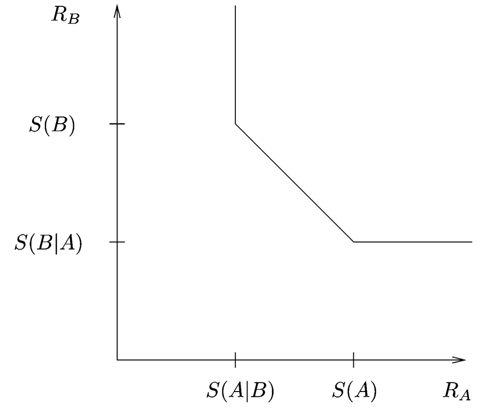

Fig. 1. Achievable rate region for Slepian–Wolf encoding.

Theorem II.1 (Slepian–Wolf. See Also [12], p. 407): Let $\mathcal{E}_{AB} = \{p_i, |\varphi_i\rangle_A |\psi_i\rangle_B\}$ such that $|\langle\varphi_i|\varphi_j\rangle|, |\langle\psi_i|\psi_j\rangle| \in \{0, 1\}$. Then (R_A, R_B) is an achievable rate pair if and only if

$$\begin{aligned} R_A + R_B &\geq S(A, B) && (4) \\ R_A &\geq S(A|B) && (5) \\ R_B &\geq S(B|A). && (6) \end{aligned}$$

The entropies here and in our subsequent theorems are taken with respect to the average state of the ensemble $\mathcal{E}_{AB}$. We will refer to inequalities (4)–(6) as the Slepian–Wolf bounds. Note that by time sharing and resource wasting, achievability of the region defined by the Slepian–Wolf bounds follows from the achievability of just two rate points: $(S(A), S(B|A))$ and $(S(A|B), S(B))$. The region is depicted in Fig. 1.

It is straightforward to show that the Slepian–Wolf bounds hold for all sources of quantum states [13], [14] but we will see in Section III that in the general case they are freqently not achievable. In fact, achievability of the Slepian–Wolf bounds appears to be a singular phenomenon. Nonetheless, Devetak and Winter have generalized the coding portion of the Slepian–Wolf theorem to the situation where the states given to one party, say Alice, are quantum mechanical while those given to the other party are classical, meaning pure and perfectly distinguishable. For such a source, they show that $(S(A), S(B|A))$ is an achievable rate pair [15]. (In Section V-B we will combine the technique they used with a type of superdense coding to develop a coding procedure for partially entangled states.) Whether the point $(S(A|B), S(B))$ is achievable in their scenario remains unknown.

Example (Cloning and information-disturbance): Let us move on to a purely quantum mechanical scenario, in which we will be able to relate the distributed compression problem to no-cloning and information-disturbance ideas. Suppose that the source generates pairs $|\varphi\rangle|\varphi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ according to the uniform distribution over qubit states. If Alice is given a noiseless quantum channel with a rate of one qubit per signal state to Charlie while Bob is given no channel at all, then perfect reconstruction of the input by Charlie is simply cloning. This situa-

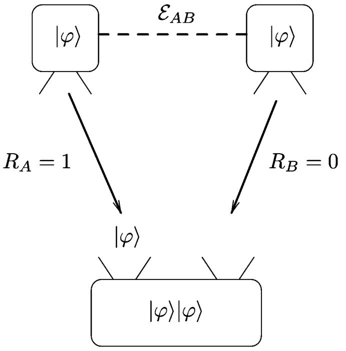

Fig. 2. Cloning as distributed compression. Solid lines represent noiseless quantum channels and dashed lines correlation in the ensemble $\mathcal{E}_{AB}$. The encoders are each given a copy of $|\varphi\rangle$ while the decoder tries to produce the state $|\varphi\rangle|\varphi\rangle$.

tion is illustrated in Fig. 2. In the approximate setting, the rate pair $(1, 0)$ is achievable if and only if there exists a sequence of CPTP maps D_n such that

$$\lim_{n\to\infty} \int \langle\varphi_1 \cdots \varphi_n| D_n(|\varphi_1 \ldots \varphi_n\rangle) |\varphi_1 \cdots \varphi_n\rangle \, d\varphi_1 \cdots d\varphi_n \tag{7}$$

is equal to 1.

Similarly, if we replace the uniform ensemble over states $|\varphi\rangle|\varphi\rangle$ by some other ensemble $\{p_i, |\varphi_i\rangle_A |\psi_i\rangle_B\}$ and again do not give Bob any capacity to communicate with Charlie, then studying distributed compression is simply an information-disturbance problem. A graphical depiction is given in Fig. 3. On the other hand, if Alice is given a full qubit's worth of capacity and Bob is given some capacity greater than zero but less than a full qubit, then we are in the regime of information-disturbance relations with prior correlation [14], since we can assume that Charlie receives a state of the form $|\varphi_{i^n}\rangle\langle\varphi_{i^n}| \otimes \rho_{i^n}$ for some density operator ρ_{i^n} and would like to use a CPTP map to convert it to a state close to $|\varphi_{i^n}\rangle|\psi_{i^n}\rangle$.

Example (Erasure codes): Our final example hints that a full theory of the distributed compression of entangled states may be related to the analysis of quantum error correcting codes. Consider the following states:

$$\begin{aligned}
|\psi_{00}\rangle &= \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle) \\
|\psi_{01}\rangle &= \frac{1}{\sqrt{2}}(|0011\rangle + |1100\rangle) \\
|\psi_{10}\rangle &= \frac{1}{\sqrt{2}}(|0101\rangle + |1010\rangle) \\
|\psi_{11}\rangle &= \frac{1}{\sqrt{2}}(|1001\rangle + |0110\rangle).
\end{aligned} \tag{8}$$

Let $\mathcal{E}_{AB}$ be the uniform ensemble for the subspace they span, giving Alice the first two qubits and Bob the last two. The subspace is, in fact, a type of quantum error correcting code known as an *erasure code*, capable of correcting for one error at a known position [16], [17]. Thus, $(2, 1)$ is an achievable rate pair: Alice sends all of her qubits while Bob throws away half of his. Meanwhile, the Slepian–Wolf bounds only require that $R_A + R_B \geq 2$ with no conditions on R_A and R_B individually. Whether the pair $(2, 1)$ is optimal, then, is actually a question about the approximate performance of a quantum error correcting code.

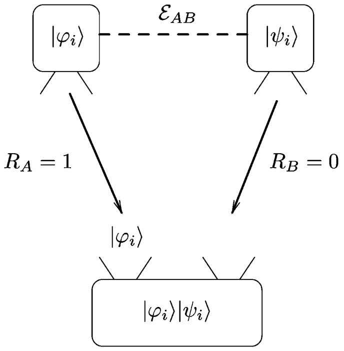

Fig. 3. Measurement without disturbance as distributed compression. This time the encoders are given the states $|\varphi_i\rangle$ and $|\psi_i\rangle$, and the decoder attempts to produce $|\varphi_i\rangle|\psi_i\rangle$.

III. A Bound for Irreducible Product State Sources

The case of irreducible product state ensembles provides what is perhaps the most striking example of the unattainability of the Slepian–Wolf conditions. A set $\mathcal{S}$ of state vectors is called *reducible* if its elements fall into two or more orthogonal subspaces. Otherwise $\mathcal{S}$ is called *irreducible*. For more details on the definition and some of its equivalent formulations, see [18] and [14]. Intuitively, an irreducible set of state vectors is one for which all nontrivial measurements induce at least some disturbance. We say that an ensemble is irreducible if the corresponding underlying set of states is. The main result of this section is a lower bound on the attainable rate sums $R_A + R_B$ for irreducible ensembles. We will use two results that have been proved elsewhere [18] which express the fact that an irreducible ensemble which some quantum operation leaves almost invariant cannot leak much quantum information to the environment of the map. These statements can be thought of as approximate and asymptotic formulations of the no-cloning and information-disturbance principles.

Lemma III.1 (Barnum et al. [18], Lemma 6.1): Suppose that $\mathcal{E} = \{p_i, |\sigma_i\rangle\}$ is an irreducible ensemble with K states. Suppose that the states $|\sigma_i\rangle$ are provided in a register A with state space $\mathcal{H}_A$ and let register B be an ancilla with state space $\mathcal{H}_B$ (with Hilbert space dimensions d_A and d_B); we will refer to B as the environment. Let

$$\begin{aligned}
\Gamma : \mathcal{H}_A \otimes \mathcal{H}_B &\longrightarrow \mathcal{H}_A \otimes \mathcal{H}_B \\
|\sigma_i\rangle_A |0\rangle_B &\longmapsto |\xi_i\rangle_{AB}
\end{aligned} \tag{9}$$

be a unitary map such that

$$\sum_i p_i F\big(|\sigma_i\rangle, \mathrm{Tr}_B |\xi_i\rangle\langle\xi_i|\big) = 1 - \epsilon.$$

Let $\{p_i, \rho_i = \mathrm{Tr}_A |\xi_i\rangle\langle\xi_i|\}$ be the environment ensemble and let

$$\chi\big(\{p_i, \rho_i\}\big) = S\left(\sum_i p_i \rho_i\right) - \sum_i p_i S(\rho_i)$$

be the Holevo quantity of the environment. Then if $\mathcal{E}$ is kept fixed, but ϵ, Γ and d_B are allowed to vary, we have $\chi \leq f(\epsilon)$

where the function f satisfies $f(\epsilon) \to 0$ as $\epsilon \to 0$. In fact we may take $f(\epsilon) = \alpha\sqrt{\epsilon} + \beta\sqrt{\epsilon}\log\sqrt{\epsilon}$ where α and β are constants.

Proposition III.2 (Barnum et al. [18], Lemma 6.4): Consider, for $\epsilon > 0$, the following:

1) an ensemble $\mathcal{E} = \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$, where each $\mathcal{E}_i$ is an irreducible ensemble on a state space of dimension at most k and with at most K signal states;
2) an encoding-decoding scheme (E, D) on $\mathcal{E}$ with average fidelity $\geq 1 - \epsilon$, leaving the environment in a state ρ_{i^n} for input state labelled by $i^n = i_1 \ldots i_n$.

Then $\frac{1}{n}\chi\big(\{p_{i^n}, \rho_{i^n}\}\big) < g(\epsilon)$ where g is a function satisfying $g(\epsilon) \to 0$ as $\epsilon \to 0$. Hence the amount of information per position tends to zero as the fidelity tends to 1, for large block lengths n.

With these tools, we can prove our main result:

Theorem III.3: Let $\mathcal{E}_{AB} = \{p_i, |\varphi_i\rangle_A|\psi_i\rangle_B\}$ be an irreducible ensemble of product states. Then a necessary condition for the rate pair (R_A, R_B) to be achievable for $\mathcal{E}_{AB}$ is that

$$R_A + R_B \geq \frac{S(\mathcal{E}_A) + S(\mathcal{E}_B) + S(\mathcal{E}_{AB})}{2}. \tag{10}$$

Proof: The basic idea is that if (R_A, R_B) fail to satisfy (10), then there is not enough room in the compressed data to absorb all the distinguishability present in the input. Some must, therefore, be left behind in the environments of Alice and Bob. The amount of distinguishability allowed there, however, is governed by Proposition III.2.

Suppose that, for some $\delta, \epsilon > 0$, Alice and Bob have a distributed encoding–decoding scheme $(E_A \otimes E_B, D)$ for blocks of size n, with $\frac{1}{n}\log d_A \leq R_A + \delta$, $\frac{1}{n}\log d_B \leq R_B + \delta$ and fidelity $1 - \epsilon$. There exists a unitary extension of Alice's encoding operation E_A in which the output Hilbert space factors as $\mathcal{H}_A = \mathcal{H}_{W_A} \otimes \mathcal{H}_{C_A}$, where $\mathcal{H}_{W_A}$ is waste and $\mathcal{H}_{C_A}$ represents her noiseless quantum channel. Thus, $E_A(\rho) = \mathrm{Tr}_{W_A} U_A(\rho \otimes |0\rangle\langle 0|)U_A^\dagger$ for some unitary U_A and fixed ancilla state $|0\rangle\langle 0|$ on W_A. Likewise, we can factor Bob's Hilbert space as $\mathcal{H}_B = \mathcal{H}_{W_B} \otimes \mathcal{H}_{C_B}$ and write $E_B(\rho) = \mathrm{Tr}_{W_B} U_B(\rho \otimes |0\rangle\langle 0|)U_B^\dagger$. Now, let $\rho^A = \sum_{i^n} p_{i^n}|\varphi_{i^n}\rangle\langle\varphi_{i^n}|$ and $\rho^B = \sum_{i^n} p_{i^n}|\psi_{i^n}\rangle\langle\psi_{i^n}|$. Then, by the subadditivity and unitary invariance of the von Neumann entropy, we find

$$\begin{aligned} S(\mathcal{E}_A^{\otimes n}) &= S(\rho^A) \\ &= S\Big(U_A(\rho^A \otimes |0\rangle\langle 0|)U_A^\dagger\Big) \\ &\leq S(\rho^{W_A}) + S(\rho^{C_A}) \end{aligned} \tag{11}$$

where

$$S(\rho^{W_A}) = S\Big(\sum_{i^n} p_{i^n}\mathrm{Tr}_{C_A}\big[U_A(|\varphi_{i^n}\rangle\langle\varphi_{i^n}| \otimes |0\rangle\langle 0|)U_A^\dagger\big]\Big)$$

is the average density operator for the reduced state of Alice's waste area and where

$$S(\rho^{C_A}) = S\Big(\sum_{i^n} p_{i^n}\mathrm{Tr}_{W_A}\big[U_A(|\varphi_{i^n}\rangle\langle\varphi_{i^n}| \otimes |0\rangle\langle 0|)U_A^\dagger\big]\Big).$$

If we define

$$\rho_{i^n}^{W_A} = \mathrm{Tr}_{C_A}\big[U_A(|\varphi_{i^n}\rangle\langle\varphi_{i^n}| \otimes |0\rangle\langle 0|)U_A^\dagger\big]$$

and

$$\rho_{i^n}^{C_A} = \mathrm{Tr}_{W_A}\big[U_A(|\varphi_{i^n}\rangle\langle\varphi_{i^n}| \otimes |0\rangle\langle 0|)U_A^\dagger\big]$$

and note that $n(R_A + \delta) \geq S(\rho^{C_A})$, since ρ^{C_A} is a state on a Hilbert space of dimension at most $2^{n(R_A+\delta)}$, we can then use (11) to conclude that

$$\begin{aligned} R_A + \delta &\geq S(\mathcal{E}_A) - \frac{1}{n}\left(\chi\big(\{p_{i^n}, \rho_{i^n}^{W_A}\}\big) - \sum_{i^n} p_{i^n} S(\rho_{i^n}^{W_A})\right) \\ &= S(\mathcal{E}_A) - \frac{1}{n}\left(\chi\big(\{p_{i^n}, \rho_{i^n}^{W_A}\}\big) - \sum_{i^n} p_{i^n} S(\rho_{i^n}^{C_A})\right) \end{aligned} \tag{12}$$

where in the last line we have used that $S(\rho_{i^n}^{W_A}) = S(\rho_{i^n}^{C_A})$. An analogous inequality obviously holds for B. At this point, we have come close to isolating the distinguishability left behind in the Alice waste area, in the form of $\frac{1}{n}\chi\big(\{p_{i^n}, \rho_{i^n}^{W_A}\}\big)$, which goes to 0 as $n \to \infty$ and $\epsilon \to 0$ by Proposition III.2. But our expression also depends on the average mixedness of the channel states $\rho_{i^n}^{C_A}$. We can control this through a series of inequalities that follow from the properties of χ, however

$$\begin{aligned} \chi\big(\{p_{i^n}, \rho_{i^n}^{C_A}\}\big) + \chi\big(\{p_{i^n}, \rho_{i^n}^{C_B}\}\big) &\geq \chi\big(\{p_{i^n}, \rho_{i^n}^{C_A} \otimes \rho_{i^n}^{C_B}\}\big) \\ &\geq \chi\big(\{p_{i^n} \\ &\qquad D(\rho_{i^n}^{C_A} \otimes \rho_{i^n}^{C_B})\}\big) \\ &\geq S(\mathcal{E}_{AB}^{\otimes n}) - nh(\epsilon) \end{aligned}$$

where $h(\epsilon) \to 0$ as $\epsilon \to 0$. The three inequalities follow, in order, from the superadditivity of χ for ensembles of product states [19], the Lindblad–Uhlmann monotonicity of χ under quantum channels, and the Fannes inequality [20]. Again using $n(R_A+\delta) \geq S(\rho^{C_A})$ and $n(R_B+\delta) \geq S(\rho^{C_B})$, this inequality implies that

$$\begin{aligned} &\sum_{i^n} p_{i^n}\big(S(\rho_{i^n}^{C_A}) + S(\rho_{i^n}^{C_B})\big) \\ &\qquad\qquad \leq n\big(R_A + R_B - S(\mathcal{E}_{AB}) + 2\delta + h(\epsilon)\big). \end{aligned}$$

This, in turn combined with Inequality (12) and its counterpart for R_B, yields, by invoking Proposition III.2

$$\begin{aligned} 2(R_A + R_B) \geq{}& S(\mathcal{E}_A) + S(\mathcal{E}_B) \\ &+ S(\mathcal{E}_{AB}) - 4\delta - 2g(\epsilon) - h(\epsilon) \end{aligned}$$

and we are done. □

IV. Optimal Compression for Sources of Bell States

The result of the previous section, that distributed compression of irreducible ensembles of product states generically cannot take full advantage of classical correlations, may be somewhat discouraging. Fortunately, this is not quite the end of the story. In this section we consider mixtures of Bell states. The quantum correlations present in the ensemble allow us to use a variation on the hashing protocol for purifying EPR pairs

[8], combined with a type of superdense coding. This protocol is fully efficient, in the sense that the total number of qubits communicated matches the Schumacher bound for the joint ensemble. We will show the following.

Theorem IV.1: Let

$$\mathcal{E}_{AB} = \begin{Bmatrix} p_1, & |\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ p_2, & |\phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\ p_3, & |\psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\ p_4, & |\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \end{Bmatrix} \tag{13}$$

be an ensemble of Bell pairs, and let $H = H(p_1, p_2, p_3, p_4)$. Then the rate pair (R_A, R_B) can be achieved by distributed compression if and only if

$$R_A \geq H/2 \quad \text{and} \quad R_B \geq H/2. \tag{14}$$

Proof of Achievability

While the states in the ensemble are highly entangled, they are also mutually orthogonal. So, while the ensemble $\mathcal{E}_{AB}$ is highly quantum mechanical from the points of view of Alice and Bob, it is classical from the point of view of the decoder, whose operations are not encumbered by any locality constraints. Our protocol makes use of this observation in an essential way: Alice and Bob will perform a series of local unitary operations before sending some fraction of their Bell pairs to Charlie, who will then perform a measurement to establish the identity of the states he has received. By appropriate choices of the local operations, all the information about the input can be hashed into the identity of the state sent to Charlie.

A Bell pair can be labelled by a pair of bits. We will follow the convention of [8], in which the Bell pair state $|0\rangle|y_1\rangle + (-1)^{y_2}|1\rangle|1-y_1\rangle$ is represented by the label (y_1, y_2). This labeling has the property that given two Bell pairs described by (y_1, y_2) and (z_1, z_2), local unitary operations suffice to add z_1 or z_2 to either of y_1 or y_2. For example, a bilateral CNOT can be used to implement the transformation

$$(z_1, z_2), (y_1, y_2) \mapsto (z_1 + y_1, z_2), (y_1, y_2 + z_2). \tag{15}$$

(Note, however, that although the operation succeeds in adding z_2 to y_2, there is an unavoidable "backaction" on y_1.) With this convention, a sequence of n Bell pairs can be described by a $2n$-bit string, which we shall denote by x^n. This string, in turn, can be considered as a concatenation of two strings x^C and x^W that are $2m$ and $2(n-m)$ bits long, respectively. x^C will represent the bits that Alice and Bob send through the channel to the decoder, and x^W will represent the bits that are thrown away.

We will use a protocol in which Alice and Bob share $2m$ random $2(n-m)$-bit-long strings $s(k)$, where k ranges from 1 to $2(n-m)$; the necessity of sharing randomness can be removed from the final protocol by observing that the average fidelity of the protocol is the probability expectation (over the shared randomness) of the average fidelities of schemes with the value of the shared radomness fixed.

The protocol is much like hashing and consists of $2m$ rounds of the following procedure. In the kth round, given the random strings above, Alice and Bob replace x_k^C with $x_k^{C'} = x_k^C + s(k) \cdot x^W$ using local operations as discussed above. The effect of these operations will be to perform $2m$ random "bit masks" on the string that is ultimately measured by Charlie, who therefore extracts the parity of a random subset of bits. After every two rounds $2j$ and $2j+1$ (where j ranges from 1 to m), Alice and Bob put the Bell pair described by the bits $x_{2j}^{C'}$ and $x_{2j+1}^{C'}$ aside. Finally, they send all m pairs to Charlie, who measures it in the Bell basis to ascertain $x_{2j}^{C'}$ and $x_{2j+1}^{C'}$.

We wish to determine the minimal m such that Charlie can decode the original n pairs with near-vanishing error probability. Consider two strings x^n and y^n, where x^n is the true initial string. We will evaluate the probability that x^n and y^n are different but nonetheless result in the same $2m$ decoder outcomes, i.e., the decoder cannot uniquely decode the state. Denote the event in which all the decoder measurements agree for x^n and y^n by E. Then

$$\begin{aligned} &\Pr(x^n \neq y^n, E) \\ &\quad = \Pr(x^n \neq y^n) \Pr(E|x^n \neq y^n) \\ &\quad = \Pr(x^n \neq y^n) \\ &\qquad \times [\Pr(x^W \neq y^W) \Pr(E|x^W \neq y^W, x^n \neq y^n) \\ &\qquad + \Pr(x^W = y^W) \Pr(E|x^W = y^W, x^n \neq y^n)] \\ &\quad = \Pr(x^n \neq y^n)[\Pr(x^W \neq y^W) 2^{-2m} \\ &\qquad + \Pr(x^W = y^W) \Pr(E|x^C \neq y^C, x^W = y^W)] \end{aligned} \tag{16}$$

where the last equality follows from multiplying by a factor of 1/2 for every subsequent random bit mask $s(k) \cdot x^W$ done by Alice and Bob.

Now, we argue that the second term in the last equality is zero. Consider the first number j such that the bit x_j^C is not equal to y_j^C. The information that actually gets sent to the decoder is in fact more complicated than x_j^C because of the random bit masks. In each case, the bit that gets sent is

$$\begin{aligned} x_j^{C'} &= x_j^C + f(x_1^W, x_2^W, \ldots, x_m^W) + g(x_1^C, \ldots, x_{j-1}^C) \\ y_j^{C'} &= y_j^C + f(y_1^W, y_2^W, \ldots, y_m^W) + g(y_1^C, \ldots, y_{j-1}^C) \end{aligned} \tag{17}$$

where f takes into account the bit masks, and g takes into account the backaction due to previous bit masks. But since $x^W = y^W$ for that term and $x_k^C = y_k^C$ for $k < j$ by hypothesis, the f and g functions are equal, and thus $x_j^{C'} \neq y_j^{C'}$. Then $\Pr(E|x^C \neq y^C, x^W = y^W) = 0$, as we wished to show. This yields

$$\Pr(x^n \neq y^n, E) \leq 2^{-2m}. \tag{18}$$

Additionally, we know that a typical set of candidates for the initial sequence of size $2^{n(H+\delta)}$ members will with probability greater than $1 - O(\exp(-\delta^2 n))$ contain the true initial sequence x [12]. The decoding will then fail only for two reasons: the true initial sequence is outside the typical set or it was impossible to uniquely decode based on the measurement outcome. Therefore,

$$\Pr(\text{failure}) \leq 2^{n(H+\delta)-2m} + O(\exp(-\delta^2 n)). \tag{19}$$

We can see that if $2m = n(H + 2\delta)$, the error probability approaches zero. The number of Bell pairs m that must be sent is just n times the rate at which Alice and Bob must send their qubits

$$R_A = R_B = H/2. \tag{20}$$

Proof of Optimality

The rate pair $(H/2, H/2)$ is also optimal: neither rate can be reduced below $H/2$. The total number of qubits sent from Alice and Bob to Charlie must be at least H by the optimality of Schumacher compression. On the other hand, Alice and Bob's local density operators are independent of the input. Intuitively, all information about the identity of the state exists in the correlations between their systems. As a result, it is impossible to do better than splitting the total rate equally between them. For comparison's sake, observe that the Slepian–Wolf bounds in this case are

$$R_A, R_B \geq H - 1 \quad \text{and} \quad R_A + R_B \geq H. \tag{21}$$

These inequalities do not ensure $R_A, R_B \geq H/2$, so we see that even here, where it is possible to fully exploit the correlations, the Slepian–Wolf bounds are insufficient to describe the achievable rate region. On the other hand, while it is not applicable in this case, Theorem III.3 would have given the stronger bound $R_A + R_B \geq \frac{1}{2}(2 + H) > H$, which is in fact violated by our coding theorem.

In order to prove optimality of the given rate pair, it is sufficient to show that $R_A \geq H/2$ regardless of the size of R_B. In what follows, we can therefore assume that Bob noiselessly transmits all of his source qubits to Charlie. We can augment any high-fidelity compression scheme by a state preparation scheme. Imagine a state preparer, Peter, who prepares Bell states according to the given distribution before giving one qubit of each pair to Alice and the other to Bob. Alice and Bob compress these Bell states as before. If the average fidelity of the compression scheme is $1 - \epsilon$, we can think of this augmented state-preparation/compression scheme as classical communication from Peter to Charlie with average error probability ϵ. The Fannes inequality [20] ensures that there exists a function $f(\epsilon)$ that approaches zero as ϵ approaches zero such that the classical communication rate from Peter to Charlie, measured in bits, is $H - f(\epsilon)$.

Let us define Peter's state preparation more precisely: for each Bell state, he can prepare a singlet, give one of the qubits to Bob, and then act on the other qubit with an appropriate Pauli rotation before handing it to Alice. Since Bob will give all his qubits to Charlie perfectly anyway, we can eliminate Bob from consideration and consider an equivalent picture in which Peter initially shares singlets with Charlie and encodes his classical information by acting with Paulis according to a distribution of entropy H. In this communication channel from Peter to Charlie, Alice is the bottleneck: she sends qubits at rate R_A. This rate assisted by entanglement can simply be thought of as superdense coding; it can result in a classical transmission rate from Peter to Charlie of at most $2R_A$. Combining this rate with our other expression for this classical transmission rate gives

$$2nR_A \geq n(H - f(\epsilon)). \tag{22}$$

Letting $\epsilon \to 0$ proves that $R_A \geq H/2$. Switching the roles of Alice and Bob completes the proof.

V. Further Examples

In this section we present a pair of examples that are designed to illustrate the range of compression strategies available to encoders. In each case, as with the optimal Bell pair strategy, the key is to make make use of orthogonality in the ensemble even though it is not directly accessible to the encoders.

A. Hidden Orthogonality

Based on the results of Section III, one might imagine that since Alice and Bob must act locally, a system in which both Alice and Bob's ensembles are *locally* irreducible (and consisting of pure states) would suffice for Alice and Bob not to be able to take full advantage of correlations. However, this is not the case, as we will show in an example that demonstrates that compressing correlated reducible product sources can involve quite subtle strategies. This example demonstrates the necessity of *global* irreducibility in Theorem III.3.

Let $\mathcal{E}_{AB} = \{1/3, |\varphi_i\rangle_A \otimes |\psi_i\rangle_B\}$ where

$$|\varphi_1\rangle = |0\rangle \tag{23}$$
$$|\varphi_2\rangle = \sqrt{\alpha}|0\rangle + \sqrt{1-\alpha}|1\rangle \tag{24}$$
$$|\varphi_3\rangle = |1\rangle \tag{25}$$
$$|\psi_1\rangle = \sqrt{1-\beta}|0\rangle + \sqrt{\beta}|1\rangle \tag{26}$$
$$|\psi_2\rangle = |1\rangle \tag{27}$$
$$|\psi_3\rangle = \sqrt{1-\beta}|2\rangle + \sqrt{\beta}|0\rangle \tag{28}$$

and both α and β are assumed to be small but nonzero. This ensemble is irreducible from the points of view of A and B individually but is reducible for AB. That is, $\mathcal{E}_A$ and $\mathcal{E}_B$ are irreducible but $\mathcal{E}_{AB}$ is not, since $|\varphi_1\rangle \otimes |\psi_1\rangle \perp |\varphi_3\rangle \otimes |\psi_3\rangle$ and $|\varphi_2\rangle \otimes |\psi_2\rangle \perp |\varphi_3\rangle \otimes |\psi_3\rangle$.

The encoder at A simply performs Schumacher compression at the rate $S(\mathcal{E}_A) \approx H(2/3)$. The encoder at B begins by projecting onto $|2\rangle$ and the subspace of states orthogonal to $|2\rangle$, which we write as $|2\rangle^\perp$. If the outcome is $|2\rangle$, he sets the state to $|0\rangle$. This operation has the effect $|\psi_i\rangle \mapsto |\psi_i'\rangle$ where $|\psi_1'\rangle = |\psi_1\rangle$, $|\psi_2'\rangle = |\psi_2\rangle$ and $|\psi_3'\rangle = |0\rangle$. The effect of the operation is shown in Fig. 4. The encoder then performs Schumacher compression on the ensemble $\{1/3, |\psi_i'\rangle\}$ at rate $H(2/3) + f(\beta)$ where $f(\beta) \to 0$ as $\beta \to 0$.

The decoder first Schumacher-decompresses the outputs of Alice and Bob's channels individually. Next, he projects onto $|10\rangle$ and $|10\rangle^\perp$. Notice that

$$\text{Span}(|\varphi_1\rangle|\psi_1\rangle, |\varphi_2\rangle|\psi_2\rangle) \subset |10\rangle^\perp \tag{29}$$

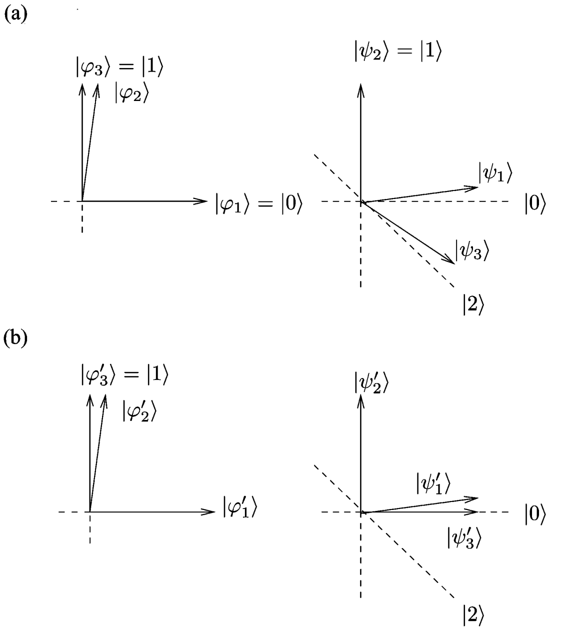

Fig. 4. Hidden orthogonality: (a) depicts the ensemble states $\{|\varphi_i\rangle|\psi_i\rangle\}$ while (b) shows the ensemble after Bob has performed the first half of his compression operation.

and that $|\varphi_3\rangle|\psi'_3\rangle = |10\rangle$. Therefore, if the outcome is $|10\rangle$, he sets the state to $|\varphi_3\rangle|\psi_3\rangle$. Otherwise, he does nothing.

In this way, $\mathcal{E}_{AB}$ can be compressed to approximately $(H(2/3), H(2/3))$ qubits per signal. Emphasizing the need for global, not just local, irreducibility in Theorem III.3, we can calculate that for this scheme, $R_A + R_B \approx 2H(2/3) \approx 1.8366$. On the other hand, the lower bound from the theorem is

$$\begin{aligned}\frac{1}{2}\big(S(\mathcal{E}_A) + S(\mathcal{E}_B) + S(\mathcal{E}_{AB})\big) &\approx \frac{1}{2}H(2/3) + \log 3\\ &\approx 2.0441 \qquad (30)\end{aligned}$$

a rate which is clearly bettered by this example.

Summarizing, Bob performs a locally dissipative operation that can only be reversed by combining his output with the output of Alice's channel. This regime, in which the ensembles are locally irreducible but globally reducible, seems to provide the greatest variety of effects and would consequently seem to be the hardest to solve in general. Indeed, the Bell state example of the previous section also falls into this category. These types of semiclassical strategies promise to frequently beat the bounds that apply to fully irreducible ensembles, but the optimal rates in the general case are completely unknown.

B. A Hybrid Strategy

In this example, we return to the realm of orthogonal entangled states but without requiring that the states be maximally entangled. The compression strategy will combine ideas from the hidden orthogonality example of Section V-A, specifically the locally irreversible measurement, and the protocol for compressing Bell states in Section IV, in which local unitary transformations were used to "piggyback" extra information onto the fraction of states sent to the decoder.

Let $\mathcal{E}_{AB}$ be an ensemble consisting of two orthogonal states, $|\varphi_0\rangle$ and $|\varphi_1\rangle$, in $\mathbb{C}^2 \otimes \mathbb{C}^2$ occurring with probabilities p_0 and p_1, respectively. By a result of Walgate *et al.* [21] we may assume without loss of generality that

$$|\varphi_0\rangle = \alpha_0|00\rangle + \beta_0|11\rangle \qquad (31)$$

and

$$|\varphi_1\rangle = \alpha_1|01\rangle + \beta_1|10\rangle \qquad (32)$$

since any other ensemble will be locally equivalent to one of this type.

As we said, the idea behind this example is to combine two different strategies. Suppose, given a state $|\varphi_i\rangle$ drawn from $\mathcal{E}_{AB}$, that Alice performs a projective measurement in the standard $\{|0\rangle, |1\rangle\}$ basis, whose outcome is $|j\rangle$. First, observe that if she sends the outcome on to Charlie and Bob also sends his state to Charlie, then Charlie can uniquely identify i, the identity of the input state. (i is a function of the parity of the outcomes of local measurements in the standard basis.) Whenever Alice's measurement outcome is not independent of the (classical) post-measurement state on Bob's system, compression of Alice's communication below the rate $H(j)$ will be possible, according to the Slepian–Wolf theorem. Up to this point, the strategy is effectively classical. To go beyond Slepian–Wolf, given a state $|\varphi_{i^n}\rangle$ drawn from $\mathcal{E}_{AB}^{\otimes n}$, we will have Alice measure only $n-m$ states, encoding information about the outcome on the remaining m, which will be sent to Charlie.

Let us estimate the rate achievable using this procedure. Denote by q_j the probability that Alice gets outcome j, by ω_j^B Bob's state given that Alice has measured j and by $\mathcal{E}'$ the ensemble $\{q_j, \omega_j^B\}$. Then

$$\omega_0^B = \frac{1}{q_0}\sum_i p_i|\alpha_i|^2|i\rangle\langle i|$$

and

$$\omega_1^B = \frac{1}{q_1}\sum_i p_i|\beta_i|^2|\neg i\rangle\langle\neg i|. \qquad (33)$$

Alice will perform the measurement on the product register $W = A_1 \ldots A_{|W|}$, where $|W| = n-m$. The number of typical j^W strings will be roughly $\exp(|W|H(\mathcal{E}'))$. Moreover, that set will partition into subsets of size roughly $\exp(|W|\chi(\mathcal{E}'))$ (and a low-probability remainder) for which Bob's density operators can be distinguished with negligible probability of error. Hence, Alice will only need to send $|W|(H(\mathcal{E}') - \chi(\mathcal{E}'))$ bits. This she will do by applying unitary encodings on her unmeasured states. Denote by $\mathcal{E}''$ the ensemble of states $(U \otimes I)\rho^{AB}(U^\dagger \otimes I)$, for a set of unitaries satisfying $\mathbb{E}\, U\psi U^\dagger = I/\dim(A)$ for all states ψ and $\rho^{AB} = \sum_i p_i\varphi_i^{AB}$. (For qubits, applying a random Pauli operator will do.) By the Holevo–Schumacher–Westmoreland (HSW) theorem [22], [23] this encoding of classical information in quantum states can achieve the communication rate

$$\begin{aligned}\chi(\mathcal{E}'') &= S(\mathcal{E}'') - \mathbb{E}S((U \otimes I)\rho^{AB}(U^\dagger \otimes I)) \qquad (34)\\ &= \log\dim(A) + S(\rho^B) - S(\rho^{AB}). \qquad (35)\end{aligned}$$

Thus, requiring that

$$m\chi(\mathcal{E}'') = (n-m)\left(H(\mathcal{E}') - \chi(\mathcal{E}')\right) \qquad (36)$$

yields a rate m/n for Alice of

$$R_A = \frac{H(\mathcal{E}') - \chi(\mathcal{E}')}{H(\mathcal{E}') - \chi(\mathcal{E}') + \chi(\mathcal{E}'')}. \tag{37}$$

As strange as this formula looks, it is important to observe that if $p_i = 1/2$ and $\alpha_i = \beta_i = 1/\sqrt{2}$, we recover the optimal rate $R_A = 1/2$ from our study of the compression of Bell states. In our proposal, however, Bob must always send at a rate $R_B = S(\mathcal{E}_B)$, which is not optimal in this case.

We will now show more carefully that this procedure actually works. The argument will essentially just require patching together known results. The versions we present here are all from [13]. First, we will need the Holevo–Schumacher–Westmoreland theorem.

Theorem V.1 (HSW [22], [23]): Consider the ensemble $\mathcal{E} = \{p_i, \rho_i\}$. For $0 < \tau < 1$, $\lambda < 1$, and sufficiently large n, there is some $\delta < K'/\sqrt{n}$ such that the following holds: For a subset A of the ensemble $\mathcal{E}^{\otimes n}$ such that the total probability of the states in A is greater than or equal to τ, and a classical alphabet $M = \{1, \ldots, 2^{n\mu}\}$, there exists a code (composed of a function f that maps elements of M to codestates $\gamma_k := \rho_{i_1} \otimes \cdots \otimes \rho_{i_n} \in A$ and an observable E on the Hilbert space of the codewords) such that the maximum error probability (defined as $\max_k\{1 - \mathrm{tr}(\gamma_k E_k) : k \in M\}$) is λ, and $\mu \geq \chi(\mathcal{E}) - \delta$.

This, in turn, implies the code partition theorem, which we will also use.

Theorem V.2: Again, consider the ensemble $\mathcal{E} = \{p_i, \rho_i\}$ and the n-block version, $\mathcal{E}^{\otimes n}$. For any $\lambda, \delta, \eta > 0$ and for sufficiently large n, there exist $m \leq 2^{n(H(\mathcal{E}) - \chi(\mathcal{E}) + 3\delta)}$ many n-block codes (as in HSW) with maximum error probability λ and pairwise disjoint "large" codebooks C_i: $|C_i| \geq 2^{n(\chi(\mathcal{E}) - 2\delta)}$ such that $\Pr\{\text{state from } \mathcal{E}^{\otimes n} \text{not in } \bigcup_{i=1}^m C_i\} < \eta$.

Finally, the gentle measurement lemma will also be useful. This result ensures that if Charlie can ascertain Alice and Bob's states with near-zero chance of error, then he can do so without causing any significant disturbance. (In this lemma, $\|\cdot\|_1$ denotes the trace norm.)

Lemma V.3: Let $\{\rho_a\}$, $a \in A$ be a family of states, and E an observable indexed by $b \in B$. Let $\phi : A \to B$ be a map and let there be $\lambda > 0$ such that for every $a \in A$, $1 - \mathrm{tr}(\rho_a E_{\phi(a)}) \leq \lambda$, i.e., the observable identifies $\phi(a)$ from ρ_a with maximal error probability λ. Then the measurement disturbs the states ρ_a very little: for every $a \in A$, $\|\rho_a - \sum_{b \in B} \sqrt{E_b} \rho_a \sqrt{E_b}\|_1 \leq \sqrt{8\lambda} + \lambda$.

According to the code partition theorem, for any $\lambda, \delta, \eta > 0$ and sufficiently large $|W|$, the ensemble $\mathcal{E}'^{\otimes |W|}$ "partitions" into at most $\exp(|W|(H(\mathcal{E}') - \chi(\mathcal{E}') + 3\delta))$ codes, each with probability of error at most λ and containing at least $\exp(|W|(\chi(\mathcal{E}') - 2\delta))$ codewords, such that the probability of any state in $\mathcal{E}'^{\otimes n}$ not lying in any of the codes is less than η. By the HSW theorem, for any $\lambda', \delta' > 0$, Alice can find a second code based on $\mathcal{E}''$ with maximum error probability λ' containing at least $\exp(m(\chi(\mathcal{E}'') - 2\delta''))$ codewords. Therefore, she will be able to send the identity of the code from the code partition theorem this way provided

$$m(\chi(\mathcal{E}'') - 2\delta') \geq (n - m)\,(H(\mathcal{E}') - \chi(\mathcal{E}') + 3\delta) \tag{38}$$

which gives the same rate we found earlier in our rough estimate. It remains to show that Charlie can still recover the original state once he has decoded the piggy-backed information about the code identity. The probability of error in identifying the code is bounded above by λ'. Let D be the complement of W (the system sent from Alice to Charlie) and recall that the identity of the code, call it k, is encoded by applying a unitary operator $U_k \otimes I_B$ to the state ρ^D. In reality, however, ρ^D is an average over input states: $\rho^D = \sum_{i^D} p_{i^D} \varphi_{i^D}$. Let $|\varphi_{i^D,k}\rangle = (U_k \otimes I_B)|\varphi_{i^D}\rangle$ and let $\tau_{i^D,k} = \sum_{k'} \sqrt{E_{k'}} \varphi_{i^D,j} \sqrt{E_{k'}}$ be Charlie's postmeasurement state. By the gentle measurement lemma

$$\begin{aligned}
\sum_{i^D} & p_{i^D} \|\varphi_{i^D,k} - \tau_{i^D,k}\|_1 \\
&= \Big\| \sum_{i^D} p_{i^D} |i^D\rangle\langle i^D| \otimes \varphi_{i^D,k} \\
&\qquad - \sum_{i^D} p_{i^D} |i^D\rangle\langle i^D| \otimes \tau_{i^D,k} \Big\|_1 \\
&\leq \sqrt{8\lambda'} + \lambda'.
\end{aligned} \tag{39}$$

Now, the total probability of error on the first $n - m$ states is bounded above by $\lambda' + \lambda + \eta$. On the rest, the decoding consists of applying $U_{k'}^\dagger \otimes I_B$, where k' is the measured code. Noting that $1 - F(\rho, \sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1$ for any states ρ and σ [24], we find that the average fidelity goes to one as λ' goes to zero. Thus, the overall average fidelity goes to one as $\lambda' + \lambda + \eta$ goes to zero.

VI. Discussion

We have studied the problem of performing distributed compression on a source of correlated quantum states. For some sources, namely sources of an irreducible set of product states, we find that it is much harder to exploit correlations in a compression protocol than would be suggested by the classical Slepian–Wolf theorem. We did not attempt to find a coding strategy matching the bound of our Theorem III.3. Indeed, since its first formulation in [14], we found the lower bound so odd that none of us even suspected that it might be tight. (It did lead us to develop some unwarranted pessimism about the problem, however. We included in an earlier preprint version of this paper the erroneous assertion that local Schumacher compression is optimal for compression of irreducible product state sources. That is true if only unitary decoding operations are permitted but not in general.) In any case, the very recent solution of the quantum Slepian–Wolf problem with free classical side-communication [25], which occurred roughly a year after initial posting of the present paper, and coding results obtained thereafter for our model without classical communication [26], show that the rate pair $\big(S(\mathcal{E}_A), \frac{1}{2}\big(S(\mathcal{E}_B) + S(\mathcal{E}_{AB}) - S(\mathcal{E}_A)\big)\big)$ is indeed universally achievable. In other words, quite surprisingly, our bound of Theorem III.3 is tight in the sense that it gives the complete rate region for irreducible ensembles of product states.

For sources of Bell states, on the other hand, we demonstrated an optimal method of compression based on the hashing protocol for entanglement distillation that fully exploits the quantum correlations between the two encoders. Nonetheless, the optimal rate region is not captured by the direct quantum analog of the classical result due to Slepian and Wolf, nor by our tighter bound. We also provided some other examples to illustrate the types of protocols that might occur in an intermediate regime, where it appears possible to exploit some of the correlations between the local sources but not all.

Thus, as compared to the classical version of the problem, we find a bewildering array of different strategies and achievable rates that are not easily synthesized into a single formula. Finding such a formula and a uniform approach to the problem integrating all possible ensembles remains an important open problem.

A Postscript: We note that the more recent studies [25], [26] change the model slightly: the source is described not by an ensemble but by a density operator. Compression has to succeed for all possible decompositions of that density operator into pure state ensembles, which is equivalently described by saying that the purification of the source density operator has to be preserved with high (entanglement) fidelity. It turns out that in this model one can show, regardless of the source, that $R_B \geq \frac{1}{2}\big(S(\mathcal{E}_B) + S(\mathcal{E}_{AB}) - S(\mathcal{E}_A)\big)$, and analogously for R_A [27]. Hence, even in this related but different model we are rather close to understanding the full rate region.

Acknowledgment

The authors would like to thank Dominic Berry, Sumit Daftuar, Igor Devetak, Michal Horodecki, Debbie Leung, Jonathan Oppenheim, Hideo Mabuchi, John Smolin, and Jon Thomas Yard for enjoyable and helpful conversations.

References

[1] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, pp. 2738–2747, 1995.

[2] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Mod. Opt.*, vol. 41, pp. 2343–2349, 1994.

[3] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Phys. Rev. A*, vol. 54, p. 1098, 1996.

[4] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, Los Alamitos, CA, 1994.

[5] M. Ohya, M. Ohya, and D. Petz, *Quantum Entropy and its use*, ser. Texts and monographs in physics. Berlin, Germany: Springer-Verlag, 1993.

[6] J. I. Cirac, P. Zoller, H. J. Kimble, and H. Mabuchi, "Quantum state transfer and entanglement distribution among distant nodes in a quantum network," *Phys. Rev. Lett.*, vol. 78, no. 16, pp. 3221–3224, 1997.

[7] M. Sasura and V. Bužek, "Multiparticle entanglement with quantum logic networks: Application to cold trapped ions," *Phys. Rev. A*, vol. 64, pp. 012305–012305, 2001.

[8] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," *Phys. Rev. A*, vol. 54, pp. 3824–3851, 1996.

[9] M. Koashi and N. Imoto, "Operations that do not disturb partially known quantum states," *Phys. Rev. A*, vol. 66, p. 022318, 2002.

[10] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum channel," *Probl. Inf. Transm.*, vol. 9, no. 3, pp. 177–183, 1973.

[11] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 471–480, 1973.

[12] T. M. Cover and J. A. Thomas*, Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[13] A. Winter, Coding Theorems of Quantum Information Theory Universität Bielefeld, 1999, Ph.D. dissertation.

[14] P. Hayden, "Distributing Quantum Information," Ph.D. dissertation, Balliol College, University of Oxford, Oxford, U.K., 2001.

[15] I. Devetak and A. Winter, "Classical data compression with quantum side information," *Phys. Rev. A*, vol. 68, p. 042301, 2003.

[16] M. Grassl, T. Beth, and T. Pellizzari, "Codes for the quantum erasure channel," *Phys. Rev. A*, vol. 56, pp. 33–39, 1997.

[17] L. Vaidman, L. Goldenberg, and S. Wiesner, "Error prevention scheme with four particles," *Phys. Rev. A*, vol. 54, pp. 1745–1748, 1996.

[18] H. Barnum, P. Hayden, R. Jozsa, and A. Winter, "On the reversible extraction of classical information from a quantum source," *Proc. R. Soc. (Lond.) A*, vol. 457, pp. 2019–2039, 2001.

[19] A. S. Holevo, "Capacity of a quantum communications channel," *Probl. Inf. Transm.*, vol. 5, no. 4, pp. 247–253, 1979.

[20] M. Fannes, "A continuity property of the entropy density for spin lattice systems," *Commun. Math. Phys.*, vol. 31, pp. 291–294, 1973.

[21] J. Walgate, A. J. Short, L. Hardy, and V. Vedral, "Local distinguishability of multipartite orthogonal quantum states," *Phys. Rev. Lett.*, vol. 85, pp. 4972–4975, 2000.

[22] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, pp. 269–273, 1998.

[23] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, pp. 131–138, 1997.

[24] C. A. Fuchs and J. v. d. Graaf, "Cryptographic distinguishability measures for quantum mechanical states," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1216–1227, 1999.

[25] M. Horodecki, M. Oppenheim, and A. Winter, "Partial quantum information," *Nature*, vol. 436, pp. 673–676, 2005.

[26] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, The Mother of all Protocols: Restructuring Quantum Information's Family Tree 2006, quant-ph/06006225.

[27] J. Oppenheim, 2005, personal communication.