Efficient Quantum Compression for Ensembles of Identically Prepared Mixed States

Yuxiang Yang, Giulio Chiribella, and Daniel Ebler
Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong

(Received 26 June 2015; revised manuscript received 27 January 2016; published 23 February 2016)

We present one-shot compression protocols that optimally encode ensembles of N identically prepared mixed states into $O(\log N)$ qubits. In contrast to the case of pure-state ensembles, we find that the number of encoding qubits drops down discontinuously as soon as a nonzero error is tolerated and the spectrum of the states is known with sufficient precision. For qubit ensembles, this feature leads to a 25% saving of memory space. Our compression protocols can be implemented efficiently on a quantum computer.

DOI: 10.1103/PhysRevLett.116.080501

Storing data into the smallest possible space is of crucial importance in present-day digital technology, especially when dealing with large amounts of information and with limited memory space [1]. The need for saving space is even more pressing in the quantum domain, where storing data is an expensive task that requires sophisticated error correction techniques [2–4].

For quantum data, Schumacher's compression [5] and its extensions [6–10] provide optimal ways to store information in the asymptotic limit of many identical and independent uses of the same source. However, in many situations there may be correlations from one use of the source to the next. In such situations, it is convenient to regard N uses of the original source as a single use of a new source, which emits messages of length N. This scenario is an instance of one-shot quantum data compression [11]. An important example of one-shot compression is when the states emitted at N subsequent moments of time are perfectly correlated, resulting in code words of the form $\rho_x^{\otimes N}$ for some density matrix ρ_x and some random parameter x. This situation arises when the original source is an uncharacterized preparation device, which generates the same quantum state at every use. For quantum bits (qubits), Plesch and Bužek [12] observed that every ensemble of identically prepared pure states can be stored without any error into $\log(N+1)$ qubits, thus allowing for an exponential saving of memory space. Recently, Rozema *et al.* [13] brought this idea into the realm of experiment, demonstrating a prototype of one-shot compression in a photonic setup.

The possibility of implementing one-shot compression in the lab opens new questions that require one to go beyond the ideal case of pure states and no errors. First, due to the presence of noise, real-life implementations typically involve mixed states—think, e. g., of quantum information processing with NMR [14], where the standard is to have thermal states at a given temperature, or, more generally, of mixed-state quantum computing [15–19]. For mixed states, the basic principle of pure-state compression does not work: in the qubit case, for example, projecting the quantum state into the smallest subspace containing the code words does not lead to any compression if the states $\rho_x^{\otimes N}$ are mixed, because in that case the smallest subspace is the whole Hilbert space. As a result, it is natural to search for compression protocols that work for mixed states and to ask which protocols achieve the best compression performance. An even more important question is how the number of qubits needed to store data depends on the errors in the decoding. Tolerating a nonzero error is natural in real-life implementations, which typically suffer from noise and imperfections.

In this Letter we answer the above questions, proposing compression protocols for ensembles of identically prepared mixed states. We first analyze the zero-error scenario, showing that the storage of N mixed qubits with known purity and unknown Bloch vector requires a quantum memory of at least $2\log N$ qubits. The size of the required memory is twice that of the required memory for pure states, but it is still exponentially smaller that the initial data size. The maximum compression is achieved by a protocol that does not require knowledge of the purity. We then investigate the more realistic case of protocols with an error tolerance. When the purity is known with sufficient precision, we find out that tolerating an error, no matter how small, allows one to encode the initial data into only $3/2\log N$ qubits, plus a small correction independent of N. Remarkably, the discontinuity in the error parameter takes place as soon as the prior knowledge of the purity is more precise than the knowledge that could be gained by measuring the N input qubits. The existence of a discontinuity is a striking deviation from the pure-state case, for which we prove that there is no significant advantage in introducing an error tolerance. Furthermore, we show that our compression protocol can be implemented efficiently and that the compression rate is optimal under the requirements that the encoding be rotationally covariant and the decoding preserve the magnitude of the total angular momentum. These assumptions are relevant in physical situations where the mixed states are used as indicators of spatial directions [20,21] and the decoding operations are limited by conservation laws [22–27]. All our results can be generalized to quantum systems of arbitrary finite dimension, where we quantify how the presence of degeneracy in the spectrum affects the compression rates.

Let us start from the qubit case, assuming N to be even for the sake of concreteness. We denote by $\mathcal{E}:\mathcal{H}^{\otimes N}\to\mathcal{H}_{\mathrm{enc}}$ ($\mathcal{D}:\mathcal{H}_{\mathrm{enc}}\to\mathcal{H}^{\otimes N}$) the encoding (decoding) channel, where $\mathcal{H}$ is the Hilbert space of a single qubit and $\mathcal{H}_{\mathrm{enc}}$ is the Hilbert space of the encoding system. For an ensemble of identically prepared qubit states $\{\rho_x^{\otimes N}, p_x\}$ the average error of the compression protocol is

$$e_N=\sum_x p_x\frac{\|\rho_x^{\otimes N}-\mathcal{D}\circ\mathcal{E}(\rho_x^{\otimes N})\|}{2}, \qquad (1)$$

$\|A\|$ denoting the trace norm. We consider ensembles where all the states ρ_x have the same purity, which is assumed to be perfectly known (this assumption will be lifted later). Let us write ρ_x as $\rho_{\mathbf{n}}=p|\mathbf{n}\rangle\langle\mathbf{n}|+(1-p)|-\mathbf{n}\rangle\langle-\mathbf{n}|$, where $|\mathbf{n}\rangle$ denotes the two-dimensional pure state with Bloch vector $\mathbf{n}=(n_x,n_y,n_z)$ and $p\geq 1/2$ is the maximum eigenvalue. We focus on mixed states ($p\neq 1$), excluding the trivial case $p=1/2$, in which the ensemble consists of just one state. For $p\notin\{1,1/2\}$, we call the ensemble $\{\rho_{\mathbf{n}}^{\otimes N}, p_{\mathbf{n}}\}$ complete if the probability distribution $p_{\mathbf{n}}$ is dense in the unit sphere. The typical example is an ensemble of mixed states with known purity and completely unknown Bloch vector. For every complete ensemble we demonstrate a sharp contrast between two types of compression: (i) zero-error compression, wherein the decoded state is equal to the initial state, and (ii) approximate compression, wherein small errors are tolerated. In the zero-error case we have the following.

Theorem 1: The minimum number of logical qubits needed to compress a complete N-qubit ensemble is $\lceil 2\log(N+2)-2\rceil$. Every compression protocol that has zero error on a complete ensemble must have zero error on every ensemble of identically prepared mixed states and on every ensemble of permutationally invariant N-qubit states.

Intuitively, the reason for the exponential reduction of the number of qubits is that the states in the ensemble are invariant under permutations and, therefore, they do not carry all the information that could be encoded into N qubits. This observation was anticipated by Blume-Kohout *et al.* in the context of state discrimination and tomography [28]. The key point of Theorem 1 is the optimality proof, which establishes that if a mixed-state ensemble is complete, then compressing it is as hard as compressing any arbitrary ensemble of permutationally invariant states [29].

In preparation of our analysis of approximate compression, it is instructive to look into an optimal protocol achieving zero-error compression. The starting point is the Schur-Weyl duality [30], stating that there exists a basis in which the N-fold tensor action of the group $\mathsf{GL}(2)$ and the natural action of the permutation group S_N are both block diagonal. In this basis, the Hilbert space of the N qubits can be decomposed as

$$\mathcal{H}^{\otimes N}\simeq\bigoplus_{j=0}^{N/2}(\mathcal{R}_j\otimes\mathcal{M}_j), \qquad (2)$$

where j is the quantum number of the total angular momentum, $\mathcal{R}_j$ is a representation space, in which the group $\mathsf{GL}(2)$ acts irreducibly, and $\mathcal{M}_j$ is a multiplicity space, in which the group acts trivially. Now, since the state $\rho_{\mathbf{n}}^{\otimes N}$ is invariant under permutations of the N qubits, one has

$$\rho_{\mathbf{n}}^{\otimes N}=\bigoplus_{j=0}^{N/2}q_{j,N}\left(\rho_{\mathbf{n},j}\otimes\frac{I_{m_j}}{m_j}\right), \qquad (3)$$

where $q_{j,N}$ is a suitable probability distribution in j, $\rho_{\mathbf{n},j}$ is a quantum state on $\mathcal{R}_j$, I_{m_j} is the identity on $\mathcal{M}_j$, and m_j is the dimension of $\mathcal{M}_j$. From Eq. (3) it is obvious that all information about the input state lies in the representation spaces. Hence, $\rho_{\mathbf{n}}^{\otimes N}$ can be encoded faithfully into the state $\mathcal{E}(\rho_{\mathbf{n}}^{\otimes N})=\bigoplus_j q_{j,N}\rho_{\mathbf{n},j}$. Such state has an exponentially smaller support, contained in the space $\mathcal{H}_N\coloneqq\bigoplus_{j=0}^{N/2}\mathcal{R}_j$, whose dimension is $\dim\mathcal{H}_N=(N/2+1)^2$. Hence, the initial state can be encoded into $\lceil\log\dim\mathcal{H}_N\rceil$ qubits—the amount declared in Theorem 1. A perfect decoding is achieved by the channel

$$\mathcal{D}(\rho)\coloneqq\bigoplus_j\left(P_j\rho P_j\otimes\frac{I_{m_j}}{m_j}\right), \qquad (4)$$

where P_j is the projector on the representation space $\mathcal{R}_j$.

Considering that qubits are a costly resource, it is worth pointing out a slight modification of the above protocol, which uses approximately $\log N$ qubits and $\log N$ classical bits. The modified protocol consists in (i) measuring the value of j, thus projecting N qubits into the state $\rho_{\mathbf{n},j}\otimes I_{m_j}/m_j$, (ii) discarding the multiplicity part, (iii) encoding the state $\rho_{\mathbf{n},j}$ into $\lceil\log(N+1)\rceil$ qubits, and (iv) transmitting the encoded state to the receiver, along with a classical message specifying the value of j. Knowing the value of j, the receiver can append an additional system in the state I_{m_j}/m_j and embed the state $\rho_{\mathbf{n},j}\otimes I_{m_j}/m_j$ into the right subspace.

Let us consider now the more realistic case of approximate compression. Here, the number of encoding qubits drops down discontinuously.

Theorem 2: For every allowed error rate $\epsilon>0$ and for every complete qubit ensemble, there exists a number $N_0>0$ such that for any $N\geq N_0$ the ensemble can be encoded into $3/2\log N+\log[4(2p-1)\sqrt{\ln(2/\epsilon)}]$ qubits with error smaller than ϵ.

The idea is to work out the explicit form of the probability distribution $q_{j,N}$ in Eq. (3), given by

$$q_{j,N} = \frac{2j+1}{2j_0}\left[B\left(N+1, p, \frac{N}{2}+j+1\right) - B\left(N+1, p, \frac{N}{2}-j\right)\right], \tag{5}$$

where $B(n, p, k)$ is the binomial distribution with n trials and with probability p, and $j_0 = (p - 1/2)(N+1)$. For large N, the distribution $q_{j,N}$ is approximately the product of a linear function with the normal distribution of variance $(N+1)p(1-p)$ centered around j_0. In order to compress, we get rid of the tails: for every $\epsilon > 0$, we select a set $\mathsf{S}_\epsilon \coloneqq \{j_0 - \lfloor\sqrt{\ln(2/\epsilon)N}\rfloor, \ldots, j_0 + \lfloor\sqrt{\ln(2/\epsilon)N}\rfloor\}$ and we compress the state $\rho_{\mathbf{n}}^{\otimes N}$ into the encoding space $\mathcal{H}_{\mathrm{enc}} = \bigoplus_{j\in\mathsf{S}_\epsilon} \mathcal{R}_j$, by applying the quantum channel

$$\mathcal{E}(\rho) \coloneqq \bigoplus_{j\in\mathsf{S}_\epsilon} \mathrm{Tr}_{\mathcal{M}_j}[\Pi_j\rho\Pi_j] + \sum_{j\notin\mathsf{S}_\epsilon} \mathrm{Tr}[\Pi_j\rho]\rho_0, \tag{6}$$

where Π_j is the projector on $\mathcal{R}_j \otimes \mathcal{M}_j$, $\mathrm{Tr}_{\mathcal{M}_j}$ is the partial trace over $\mathcal{M}_j$, and ρ_0 is a fixed state with support inside $\mathcal{H}_{\mathrm{enc}}$. The encoding space has dimension

$$\dim \mathcal{H}_{\mathrm{enc}} = \sum_{j\in\mathsf{S}_\epsilon}(2j+1) \leq (2j_0+1)\left(2\sqrt{N\ln\frac{2}{\epsilon}}+1\right),$$

growing as $N^{3/2}$. The initial state can be recovered, up to error ϵ, by a suitable decoding channel [29].

Theorem 2 guarantees that N identical copies of a mixed state with known purity can be stored (up to an error ϵ) into $3/2 \log N$ qubits, plus an overhead that is doubly logarithmic in $1/\epsilon$. This result is good news for future implementations, because the overhead grows slowly with the required accuracy. For example, when $p = 0.6$, $N = 20$ identically prepared qubits with Bloch vectors pointing in arbitrary directions can be compressed into 8 qubits with an error smaller than 1%. In addition to the fully quantum version of the protocol, one can construct a hybrid version where the initial state is stored partly into qubits and partly into classical bits, as discussed in the zero-error case. In the hybrid version, the discontinuity between zero-error and approximate compression pertains to the number of classical bits needed to communicate the value of j, which decreases from $\log N$ to $1/2 \log N$ as soon as a nonzero error is tolerated.

Our result highlights a radical difference between mixed and pure states: for mixed states, every finite error tolerance $\epsilon > 0$ allows one to reduce the size of the compression space from the original $2 \log N$ qubits to $3/2 \log N$ qubits. Such a discontinuity does not take place for pure states: for pure states with completely unknown Bloch vector, every compression protocol with tolerance ϵ requires at least $(1-2\epsilon)\log N$ qubits [29].

It is worth commenting on the importance of knowing the purity. Our approximate protocol requires the purity to be perfectly known, so that one can encode only the subspaces where the quantum number j is in a strip around the most likely value. If the purity is only partially known, the protocol can be adapted by broadening the size of the strip, i.e., by changing the set S_ϵ. Specifically, suppose that the eigenvalues of $\rho_{\mathbf{n}}$ are known up to an error $\Delta p = O(N^{-\gamma})$, with $\gamma \geq 1/2$. In this case, the number of encoding qubits can be reduced to $3/2 \log N + g(\epsilon, \gamma)$ where g is a function depending on ϵ and γ, but not on N. Hence, the discontinuity between zero-error and approximate compression persists. However, the situation is different if the eigenvalues are known with less precision: if the error in the specification of the eigenvalues scales as $N^{-\gamma}$ with $\gamma < 1/2$, then the number of encoding qubits becomes $(2-\gamma)\log N$. Quite intriguingly, the separation between the two regimes takes place exactly when the knowledge of the eigenvalues becomes more precise than the knowledge that could be extracted through spectrum estimation [31]. Note that our protocol can be combined for free with spectrum estimation, which only requires measuring the value of j. However, the *a posteriori* knowledge of the measurement outcome cannot replace the *a priori* knowledge of the spectrum: indeed, finding the outcome j leads to estimating the maximum eigenvalue as $\hat{p} = 1/2 + j/(N+1)$ [31] and then to encoding the state $\rho_{\mathbf{n},j}$ into $\lceil \log(2j+1) \rceil$ qubits. In order to decode, the receiver needs a classical message communicating the value of j, which requires $\lceil \log(N/2+1) \rceil$ bits in the one-shot scenario. This leads to the same resource scaling as in the zero-error case, i.e., approximately $\log N$ qubits to send the encoded state and $\log N$ bits to communicate j.

The protocol of Theorem 2 is optimal within the physically relevant class of protocols constrained by covariance under rotations and by the preservation of the magnitude of the angular momentum. More precisely, we have the following [29].

Theorem 3: Every compression protocol that encodes a complete N-qubit ensemble into $(3/2 - \delta)\log N$ qubits with covariant encoding and a decoding that preserves the magnitude of the total angular momentum will necessarily have error $e \geq 1/2$ in the asymptotic limit.

Let us now discuss the complexity of the compression protocol. To operate on the input state we use the Schur transform [12,32,33], which transforms the initial N qubits together with $O(\log N)$ ancillary qubits into three registers: (i) the index register, where the value of j is stored into the state of $\log(N/2+1)$ qubits, (ii) the representation register, which uses $\log(N+1)$ qubits to encode the representation spaces, and (iii) the multiplicity register, where the multiplicity spaces are encoded into $O(N)$ qubits (see Fig. 1). Since the implementation of the Schur transform in a

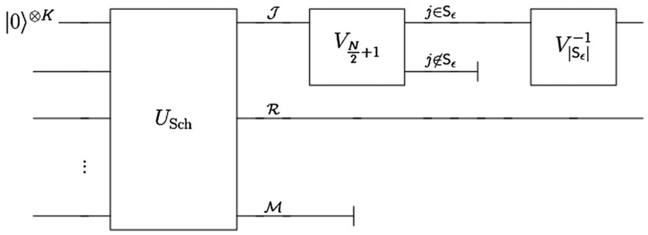

FIG. 1. A quantum circuit for encoding. The Schur transform turns the initial N qubits together with $K = O(\log N)$ ancillary qubits into three registers: the index register $\mathcal{J}$, the representation register $\mathcal{R}$, and the multiplicity register $\mathcal{M}$. The multiplicity register is discarded. The index register is encoded into $N/2 + 1$ qubits by the position embedding $V_{N/2+1}$. The qubits in positions outside S_ϵ are discarded and the remaining qubits are reencoded into $\lceil \log |\mathsf{S}_\epsilon| \rceil$ qubits.

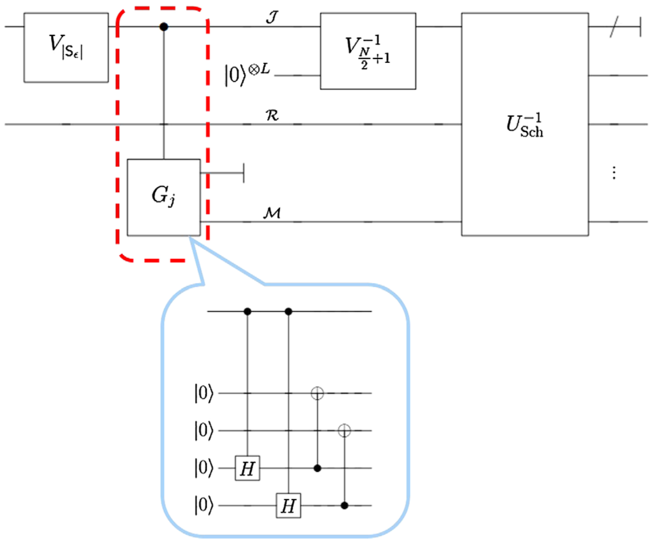

FIG. 2. A quantum circuit for decoding. The first operation is the position embedding $V_{|\mathsf{S}_\epsilon|}$, which produces $|\mathsf{S}_\epsilon|$ output qubits. The jth of these qubits controls the generation of a maximally mixed state of rank m_j (achieved by the controlled operation G_j, represented explicitly in the blue inset for $m_j = 4$). The third step is the initialization of $L = N/2 + 1 - |\mathsf{S}_\epsilon|$ qubits which are put in positions corresponding to values of j outside S_ϵ. After a total of $N/2 + 1$ qubits are in place, the inverse of the position embedding is performed, followed by the inverse of the Schur transform. The output of the circuit is a state on N qubits and $K = O(\log N)$ ancillas, which are finally discarded.

quantum circuit is approximate, we focus on approximate compression, so that the Schur transform error can be absorbed into the compression error. Let us analyze first the encoding. The first step is the approximate Schur transform, whose complexity is $\mathrm{poly}(N, \log 1/\epsilon')$, ϵ' being the approximation error [32,33]. We set ϵ' to be vanishing exponentially in N, resulting in a complexity $\mathrm{poly}(N)$ for the implementation of the Schur transform. After the Schur transform has been performed, the encoding circuit embeds the index register into an exponentially larger register of $N/2 + 1$ qubits, transforming the state $|j\rangle$ into the state where the jth qubit is set to $|1\rangle$ and the rest of the qubits are set to $|0\rangle$ [12]. We refer to this transformation as position embedding and denote it by V_D, where D is the dimension of the register that is being embedded (in this case $D = N/2 + 1$). The point of position embedding is to physically encode the value of j in a form that makes it easy to check whether or not j belongs to the set S_ϵ. In fact, such a check can be equivalently implemented on a classical computer. After this step, the circuit discards the qubits in positions outside the set S_ϵ and transforms the remaining qubits into $\log |S_\epsilon|$ qubits by applying $V^{-1}_{|S_\epsilon|}$. Now, the complexity of position embedding is upper bounded by $D(\log D)^2$ [12]. Since j ranges from 0 to $N/2$, the total complexity of the position embedding and of its inverse scales as $N(\log N)^2$. From the above reasoning, it is clear that the bottleneck of the encoding is the implementation of the Schur transform, which leads to an overall complexity of $\mathrm{poly}(N)$ for the encoding circuit. The situation is similar for the decoding, which also uses position embedding to perform operations depending on j (see Fig. 2). The only new parts are the initialization of $N/2 + 1 - |\mathsf{S}_\epsilon|$ qubits in the index register and the preparation of maximally mixed states of rank m_j in the multiplicity register, which can be approximately generated with exponential precision in $O(N^2)$ operations [29]. Summing over the values of j in S_ϵ, we then obtain a number of operations upper bounded by $O(N^2)|\mathsf{S}_\epsilon| = O(N^{5/2})$. From the above count it is clear that the overall complexity is polynomial in N. In addition to the computational complexity, it is worth discussing the size of the ancillary systems needed in our compression protocol. Since the multiplicity register is discarded, the Schur transform in our protocol needs only an ancilla of $O(\log N)$ qubits [28]. The position embeddings require ancillas of size $O(N)$, but, as mentioned earlier, they can be implemented on a classical computer. Hence, the total number of qubits that need to be kept coherent throughout our protocol scales only as $O(\log N)$.

Our compression protocol, presented for qubits, can be generalized to quantum systems of arbitrary dimension d. In this case, an ensemble of N identically prepared rank-r states with known spectrum can be compressed with error less than ϵ into approximately $(2dr - r^2 - 1)/2 \log N$ qubits. In addition, one can take advantage of the presence of degeneracies and further reduce the number of qubits: every time the same eigenvalue appears in the spectrum, the number of qubits is reduced by at least $1/2 \log N$ (see [29] for the exact value). Again, the protocol can be implemented efficiently and is optimal under suitable symmetry assumptions [29].

In this Letter we showed how to efficiently store ensembles of identically prepared quantum systems into an exponentially smaller memory space. For mixed states we discovered that, whenever a nonzero error is allowed, the size of the memory is cut down in a discontinuous way, provided that the spectrum of the state is known with sufficient precision. Intriguingly, the dropoff in the memory size takes place as soon as the prior information about the eigenvalues is more than the information that could be extracted by a measurement on the input copies. Our approximate compression protocols can be implemented efficiently on a quantum computer.

We thank M. Ozols and the referees of this Letter for a number of comments that stimulated substantial improvements of the original manuscript. This work is supported by the National Natural Science Foundation of China through Grant No. 11450110096, by the Foundational Questions Institute (Grant No. FQXi-RFP3-1325), by the 1000 Youth Fellowship Program of China, and by the HKU Seed Funding for Basic Research.

[1] S. Sagiroglu and D. Sinanc, in *2013 International Conference on Collaboration Technologies and Systems (CTS)* (2013), pp. 42–47.

[2] B. Julsgaard, J. Sherson, J. Cirac, J. Fiurasek, and E. Polzik, Nature (London) **432**, 482 (2004).

[3] B. Zhao, Y.-A. Chen, X.-H. Bao, T. Strassel, C.-S. Chuu, X.-M. Jin, J. Schmiedmayer, Z.-S. Yuan, S. Chen, and J.-W. Pan, Nat. Phys. **5**, 95 (2009).

[4] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Nature (London) **458**, 996 (2009).

[5] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[6] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).

[7] H.-K. Lo, Opt. Commun. **119**, 552 (1995).

[8] M. Horodecki, Phys. Rev. A **57**, 3364 (1998).

[9] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **81**, 1714 (1998).

[10] C. H. Bennett, A. W. Harrow, and S. Lloyd, Phys. Rev. A **73**, 032336 (2006).

[11] N. Datta, J. Renes, R. Renner, and M. Wilde, IEEE Trans. Inf. Theory **59**, 8057 (2013).

[12] M. Plesch and V. Bužek, Phys. Rev. A **81**, 032317 (2010).

[13] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, Phys. Rev. Lett. **113**, 160504 (2014).

[14] L. Vandersypen and I. Chuang, Rev. Mod. Phys. **76**, 1037 (2005).

[15] D. Aharonov, A. Kitaev, and N. Nisan, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing* (ACM, New York, 1998), pp. 20–30.

[16] E. Knill and R. Laflamme, Phys. Rev. Lett. **81**, 5672 (1998).

[17] P. W. Shor and S. P. Jordan, Quantum Inf. Comput. **8**, 681 (2008).

[18] A. Datta, A. Shaji, and C. M. Caves, Phys. Rev. Lett. **100**, 050502 (2008).

[19] B. P. Lanyon, M. Barbieri, M. P. Almeida, and A. G. White, Phys. Rev. Lett. **101**, 200501 (2008).

[20] R. Demkowicz-Dobrzański, Phys. Rev. A **71**, 062321 (2005).

[21] E. Bagan, M. Baig, A. Brey, R. Muñoz-Tapia, and R. Tarrach, Phys. Rev. Lett. **85**, 5230 (2000).

[22] G. Gour and R. W. Spekkens, New J. Phys. **10**, 033023 (2008).

[23] I. Marvian and R. W. Spekkens, New J. Phys. **15**, 033001 (2013).

[24] I. Marvian and R. W. Spekkens, Phys. Rev. A **90**, 062110 (2014).

[25] I. Marvian and R. W. Spekkens, arXiv:1212.3378.

[26] M. Ahmadi, D. Jennings, and T. Rudolph, New J. Phys. **15**, 013057 (2013).

[27] I. Marvian and R. W. Spekkens, Nat. Commun. **5** (2014).

[28] R. Blume-Kohout, S. Croke, and M. Zwolak, arXiv: 1201.6625.

[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.116.080501, which includes Refs. [34–42], for the explicit proof.

[30] W. Fulton and J. Harris, *Representation Theory*, Graduate Texts in Mathematics Vol. 129 (Springer Science and Business Media, New York, 1991).

[31] M. Keyl and R. F. Werner, Phys. Rev. A **64**, 052311 (2001).

[32] A. W. Harrow, Ph.D. thesis, Massachusetts Institute of Technology, 2005, arXiv:quant-ph/0512255.

[33] D. Bacon, I. L. Chuang, and A. W. Harrow, Phys. Rev. Lett. **97**, 170502 (2006).

[34] R. Blume-Kohout, H. K. Ng, D. Poulin, and L. Viola, Phys. Rev. A **82**, 062306 (2010).

[35] A. S. Holevo, Prob. Peredachi Inf. **9**, 3 (1973).

[36] R. Alicki and M. Fannes, J. Phys. A **37**, L55 (2004).

[37] R. Alicki, S. Rudnicki, and S. Sadowski, J. Math. Phys. **29**, 1158 (1988).

[38] C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).

[39] R. Goodman and N. R. Wallach, *Representations and Invariants of the Classical Groups*, Encyclopedia of Mathematics and its Applications Vol. 68 (Cambridge University Press, Cambridge, England, 1998).

[40] R. O'Donnell and J. Wright, arXiv:1501.05028.

[41] M. Christandl and G. Mitchison, Commun. Math. Phys. **261**, 789 (2006).

[42] A. Kitaev, D. Mayers, and J. Preskill, Phys. Rev. A **69**, 052326 (2004).