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

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.

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 codewords 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 $\mathcal{E}:\mathcal{H}^{\otimes N}\to$

$\mathcal{H}_{\rm enc}$ ($\mathcal{D}:\mathcal{H}_{\rm enc}\to\mathcal{H}^{\otimes N}$) the encoding (decoding) channel, where $\mathcal{H}$ is the Hilbert space of a single qubit and $\mathcal{H}_{\rm enc}$ 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

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

$\|A\|$ 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

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

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

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

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 $\mathcal{E}\left(\rho_{\mathbf{n}}^{\otimes N}\right)=\bigoplus_j q_{j,N}\,\rho_{\mathbf{n},j}$. Such state has an exponentially smaller support, contained in the space $\mathcal{H}_N:=\bigoplus_{j=0}^{N/2}\mathcal{R}_j$, 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

$$\mathcal{D}(\rho):=\bigoplus_j\left(P_j\,\rho\,P_j\otimes\frac{I_{m_j}}{m_j}\right), \tag{4}$$

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 prob-

ability distribution $q_{j,N}$ in Eq. (3), given by

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

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 $\mathsf{S}_\epsilon := \left\{j_0 - \lfloor\sqrt{\ln(2/\epsilon)N}\rfloor, \dots, j_0 + \lfloor\sqrt{\ln(2/\epsilon)N}\rfloor\right\}$ and we compress the state $\rho_{\mathbf{n}}^{\otimes N}$ into the encoding space $\mathcal{H}_{\rm enc} = \bigoplus_{j\in\mathsf{S}_\epsilon}\mathcal{R}_j$, by applying the quantum channel

$$\mathcal{E}(\rho) := \bigoplus_{j\in\mathsf{S}_\epsilon} {\rm Tr}_{\mathcal{M}_j}\left[\,\Pi_j\,\rho\,\Pi_j\,\right] + \sum_{j\notin\mathsf{S}_\epsilon} {\rm Tr}\left[\Pi_j\,\rho\right]\,\rho_0\,, \qquad (6)$$

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

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

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 faithfully to 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 direction 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 $\gamma \ge 1/2$. 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 $e \ge 1/2$ 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 quantum circuit is approximate, we focus on approximate compression, so that the Schur transform error can be absorbed into the compres-

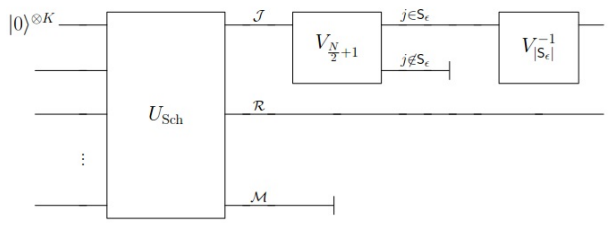

FIG. 1. **A quantum circuit for encoding.** The Schur transform turns the initial N qubits together with $K = O(\log N)$ ancillary qubits into three registers: the index register $\mathcal{J}$, the representation register $\mathcal{R}$, and the multiplicity register $\mathcal{M}$. The multiplicity register 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.

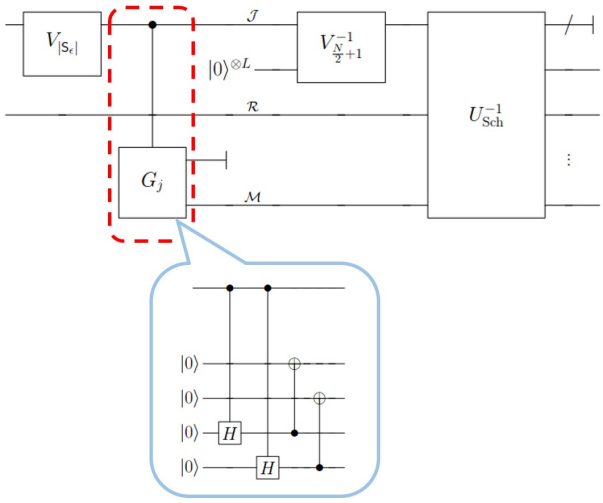

FIG. 2. **A quantum circuit for decoding.** The first operation is the position embedding $V_{|\mathsf{S}_\epsilon|}$, which produces $|\mathsf{S}_\epsilon|$ output qubits. The jth of these qubits controls the generation of a maximally mixed state of rank m_j (achieved by the controlled operation G_j, 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.

sion 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 $\left(2dr - r^2 - 1\right)/2 \log N$ 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.

Acknowledgments. 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 *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (2013) pp. 42–47.
[2] B. Julsgaard, J. Sherson, J. Cirac, J. Fiurasek, and E. Polzik, Nature **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, Nature Physics **5**, 95 (2009).
[4] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Nature **458**, 996 (2009).
[5] B. Schumacher, Physical Review A **51**, 2738 (1995).
[6] R. Jozsa and B. Schumacher, Journal of Modern Optics **41**, 2343 (1994).
[7] H.-K. Lo, Optics Communications **119**, 552 (1995).
[8] M. Horodecki, Physical Review A **57**, 3364 (1998).
[9] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, Physical Review Letters **81**, 1714 (1998).
[10] C. H. Bennett, A. W. Harrow, and S. Lloyd, Physical Review A **73**, 032336 (2006).
[11] N. Datta, J. Renes, R. Renner, and M. Wilde, Information Theory, IEEE Transactions on **59**, 8057 (2013).
[12] M. Plesch and V. Bužek, Physical Review A **81**, 032317 (2010).
[13] L. A. Rozema, D. H. Mahler, A. Hayat, P. S. Turner, and A. M. Steinberg, Physical Review Letters **113**, 160504 (2014).
[14] L. Vandersypen and I. Chuang, Reviews of Modern Physics **76**, 1037 (2005).
[15] D. Aharonov, A. Kitaev, and N. Nisan, in *Proceedings of the thirtieth annual ACM symposium on Theory of computing* (ACM, 1998) pp. 20–30.
[16] E. Knill and R. Laflamme, Physical Review Letters **81**, 5672 (1998).
[17] P. W. Shor and S. P. Jordan, Quantum Information & Computation **8**, 681 (2008).
[18] A. Datta, A. Shaji, and C. M. Caves, Physical Review Letters **100**, 050502 (2008).
[19] B. Lanyon, M. Barbieri, M. Almeida, and A. White, Physical Review Letters **101**, 200501 (2008).
[20] R. Demkowicz-Dobrzański, Physical Review A **71**, 062321 (2005).
[21] E. Bagan, M. Baig, A. Brey, R. Muñoz Tapia, and R. Tarrach, Physical Review Letters **85**, 5230 (2000).
[22] G. Gour and R. W. Spekkens, New Journal of Physics **10**, 033023 (2008).
[23] I. Marvian and R. W. Spekkens, New Journal of Physics **15**, 033001 (2013).
[24] I. Marvian and R. W. Spekkens, Physical Review A **90**, 062110 (2014).
[25] I. Marvian and R. W. Spekkens, arXiv preprint arXiv:1212.3378 (2012).
[26] M. Ahmadi, D. Jennings, and T. Rudolph, New Journal of Physics **15**, 013057 (2013).
[27] I. Marvian and R. W. Spekkens, Nature Communications **5** (2014).
[28] R. Blume-Kohout, S. Croke, and M. Zwolak, arXiv preprint arXiv:1201.6625 (2012).
[29] See Supplemental Material, which includes Refs. [34–42], for the explicit proof.
[30] W. Fulton and J. Harris, *Representation theory*, Vol. 129 (Springer Science & Business Media, 1991).
[31] M. Keyl and R. F. Werner, Physical Review A **64**, 052311 (2001).
[32] A. W. Harrow, Ph.D. thesis (2005), quant-ph/0512255.
[33] D. Bacon, I. Chuang, and A. Harrow, Physical Review Letters **97**, 170502 (2006).
[34] R. Blume-Kohout, H. K. Ng, D. Poulin, and L. Viola, Physical Review A **82**, 062306 (2010).
[35] A. S. Holevo, Problemy Peredachi Informatsii **9**, 3 (1973).
[36] R. Alicki and M. Fannes, Journal of Physics A: Mathematical and General **37**, L55 (2004).
[37] R. Alicki, S. Rudnicki, and S. Sadowski, Journal of Mathematical Physics **29**, 1158 (1988).
[38] C. Itzykson and M. Nauenberg, Reviews of Modern Physics **38**, 95 (1966).
[39] R. Goodman and N. R. Wallach, *Representations and invariants of the classical groups*, Vol. 68 (Cambridge University Press, 1998).
[40] R. O'Donnell and J. Wright, arXiv preprint (2015), arXiv 1501.05028.
[41] M. Christandl and G. Mitchison, Communications in Mathematical Physics **261**, 789 (2006).
[42] A. Kitaev, D. Mayers, and J. Preskill, Physical Review A **69**, 052326 (2004).

PROOF OF THEOREM 1

Here we show the optimality of our the error protocol in the main text. Specifically, we show that no zero-error protocol exists that compresses a complete ensemble of mixed states into less than $\lceil 2\log(N+2)-2\rceil$.

The zero error condition

The condition for zero-error compression requires that the average error defined as

$$e_N = \sum_{\mathbf{n}} p_{\mathbf{n}} \frac{\left\| \rho_{\mathbf{n}}^{\otimes N} - \mathcal{D} \circ \mathcal{E} \left(\rho_{\mathbf{n}}^{\otimes N} \right) \right\|}{2} = 0 \, . \tag{7}$$

This condition immediately implies $\| \mathcal{D} \circ \mathcal{E} (\rho_{\mathbf{n}}^{\otimes N}) - \rho_{\mathbf{n}}^{\otimes N} \| = 0$ for every $\mathbf{n}$ except for a zero-measure set. Since the Hermitian operator $\mathcal{D} \circ \mathcal{E} (\rho_{\mathbf{n}}^{\otimes N}) - \rho_{\mathbf{n}}^{\otimes N}$ has only zero eigenvalues, it must be a null operator. Hence, the channel $\mathcal{C} := \mathcal{D} \circ \mathcal{E}$ must fix $\rho_{\mathbf{n}}^{\otimes N}$, namely that

$$\mathcal{C} (\rho_{\mathbf{n}}^{\otimes N}) = \rho_{\mathbf{n}}^{\otimes N} \tag{8}$$

for every $\mathbf{n}$ except for a set of zero measure. Since $p_{\mathbf{n}}$ has full support on the Bloch sphere, the above condition holds for a dense set of points on the Bloch sphere. As a result, for every Bloch vector $\mathbf{n}$ there exists a sequence $\left\{ \rho_{\mathbf{n}_k}^{\otimes N} \right\}$ of Bloch vectors satisfying Eq. (8) such that $\lim_{k \to \infty} \mathbf{n}_k = \mathbf{n}$ and

$$\lim_{k \to \infty} \rho_{\mathbf{n}'_k}^{\otimes N} = \rho_{\mathbf{n}}^{\otimes N} \, .$$

Consequently, we have

$$\begin{aligned} \left\| \mathcal{D} \circ \mathcal{E} (\rho_{\mathbf{n}}^{\otimes N}) - \rho_{\mathbf{n}}^{\otimes N} \right\|_1 &= \left\| \mathcal{D} \circ \mathcal{E} \left(\lim_{k \to \infty} \rho_{\mathbf{n}'_k}^{\otimes N} \right) - \lim_{k \to \infty} \rho_{\mathbf{n}'_k}^{\otimes N} \right\| \\ &= \left\| \lim_{k \to \infty} \left[\mathcal{D} \circ \mathcal{E} (\rho_{\mathbf{n}'_k}^{\otimes N}) - \rho_{\mathbf{n}'_k}^{\otimes N} \right] \right\| \\ &= 0, \end{aligned}$$

which implies that $\mathcal{C} (\rho_{\mathbf{n}}^{\otimes N}) = \rho_{\mathbf{n}}^{\otimes N}$ for every vector $\mathbf{n}$ on the Bloch sphere.

The algebra associated to the fixed points of a channel

Here we develop a technique that generates fixed points of a given channel starting from an initial set of fixed points. Our technique is based on a result by Blume-Kohout *et al* [34] characterizes the fixed points. Specifically, Theorem 5 of Ref. [34] guarantees that one can find a decomposition of the Hilbert space as $\mathcal{H} = \bigoplus_k (\mathcal{L}_k \otimes \mathcal{M}_k)$, with the property that the fixed points of a given channel acting on $\mathcal{H}$ are all the operators of the form

$$A = \bigoplus_k \left(A^{(k)} \otimes \omega_0^{(k)} \right) \, , \tag{9}$$

where $A^{(k)}$ is an arbitrary matrix on $\mathcal{L}_k$ and $\omega_0^{(k)}$ is a fixed non-negative matrix on $\mathcal{M}_k$. Using this fact, we develop a technique that generates fixed points of a channel starting from an initial set of fixed points.

Proposition 1. *Let* $\mathrm{Fix}(\mathcal{C})$ *be the set of fixed points of channel* $\mathcal{C}$*, let* $\{A_x\}_{x \in \mathsf{X}} \subset \mathrm{Fix}(\mathcal{C})$ *be a subset of non-negative fixed points, and let* $\mu(\mathrm{d}\, x)$ *be a non-negative measure on* X*. Then, the set of operators*

$$\mathcal{A} = E^{-1/2} \, \mathrm{Fix}(\mathcal{C}) \, E^{-1/2} \, , \qquad E := \int \mu(\mathrm{d}\, x) \; A_x \, ,$$

is a matrix $*$*-algebra (i. e. a matrix algebra closed under adjoint). Moreover, one has* $E^{1/2} \mathcal{A} E^{1/2} \subseteq \mathsf{Fix}(\mathcal{C})$.

[Notation: for a non-invertible operator E, we define E^{-1} as the inverse on the support of E.]

Proof. Writing each operator A_x in the form (9), we obtain

$$E = \bigoplus_k \left(E^{(k)} \otimes \omega_0^{(k)} \right) \, , \qquad E^{(k)} = \int \mu(\mathrm{d}\, x) A_x^{(k)} \, .$$

Hence, for a generic fixed point $A \in \mathsf{Fix}(\mathcal{C})$, decomposed as in Eq. (9), we have

$$E^{-1/2} A E^{-1/2} = \bigoplus_k \left[\left(E^{(k)}\right)^{-1/2} A_k \left(E^{(k)}\right)^{-1/2} \otimes P_k \,, \right]$$

where P_k is the projector on the support of $\omega_0^{(k)}$. Since each A_k is a generic operator on $\mathcal{L}_k$, we have

$$E^{-1/2}\, \mathsf{Fix}(\mathcal{C})\, E^{-1/2} = \bigoplus_k \left[\mathsf{B}(\mathcal{S}_k) \otimes P_k\right] \,,$$

where $\mathsf{B}(\mathcal{S}_k)$ denotes the algebra of all linear operators on the subspace $\mathcal{S}_k = \mathsf{Supp}\left[E^{(k)}\right]$. Hence, $\mathcal{A} = E^{-1/2}\, \mathsf{Fix}(\mathcal{C})\, E^{-1/2}$ is an algebra and is closed under adjoint. On the other hand, we have

$$E^{1/2}\, \mathcal{A}\, E^{1/2} = \bigoplus_k \left[\mathsf{B}(\mathcal{S}_k) \otimes M_0^{(k)}\right] \,,$$

meaning that every operator in $E^{1/2}\, \mathcal{A}\, E^{1/2}$ is of the form (9)—that is, it is a fixed point. □

The minimal algebra required by the zero error condition

Let us apply Proposition 1 to the channel $\mathcal{C} = \mathcal{D} \circ \mathcal{E}$, resulting from the concatenation of the encoding and the decoding in a generic zero-error protocol. By the zero-error condition, all the states $\rho_{\mathbf{n}}^{\otimes N}$ are fixed points. The states can be decomposed as

$$\rho_{\mathbf{n}}^{\otimes N} = \bigoplus_{j=0}^{N/2} q_{j,N} \left(\rho_{\mathbf{n},j} \otimes \frac{I_{m_j}}{m_j}\right) . \tag{10}$$

A priori, this block decomposition could be completely unrelated with the block decomposition of Eq. (9). Proving that the two decompositions coincide will be the main part of our argument.

Choosing the measure $\mu(\mathrm{d}\, x)$ in Proposition 1 to be the invariant measure over $\mathbf{n}$, the average operator E is given by

$$E = \bigoplus_{j=0}^{N/2} q_{j,N} \left(\frac{I_j}{d_j} \otimes \frac{I_{m_j}}{m_j}\right) .$$

Hence, the algebra $\mathcal{A}$ defined in Proposition 1 must contain all the operators of the form

$$E^{-1/2}\, \rho_{\mathbf{n}}^{\otimes N}\, E^{-1/2} = \bigoplus_{j=0}^{N/2} \left(d_j\, \rho_{\mathbf{n},j} \otimes I_{m_j}\right) \,,$$

for every unit vector $\mathbf{n}$. Hence, $\mathcal{A}$ must contain the smallest algebra $\mathcal{A}_{\min}$ generated by the above operators. We will now characterize this algebra:

Proposition 2. *If the states in Eq. (10) are not maximally mixed, $\mathcal{A}_{\min}$ contains the matrix algebra of all operators on the symmetric subspace, corresponding to $j = N/2$ in the decomposition (10).*

Proof. Let us express the state $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ as $\rho = e^{-\beta Z}/\operatorname{Tr}[e^{-\beta Z}]$, $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$ for a suitable $\beta \geq 0$. By definition, for every unitary $U \in \mathsf{SU}(2)$, the algebra $\mathcal{A}_{\min}$ contains the operator

$$\begin{aligned} A_U &:= E^{-1/2} (U \rho U^\dagger)^{\otimes N} E^{-1/2} \\ &= \bigoplus_{j=0}^{N/2} \frac{d_j}{\operatorname{Tr}\left[e^{-\beta J_z^{(j)}}\right]} \left(U^{(j)}\, e^{-\beta J_z^{(j)}} U^{(j)\dagger} \otimes I_{m_j}\right) , \qquad J_z^{(j)} = \sum_{m=-j}^{j} m\, |j,m\rangle\langle j,m| \end{aligned} \tag{11}$$

where $U^{(j)}$ denotes the $(2j+1)$-dimensional irreducible representation of $\mathsf{SU}(2)$. Moreover, since the algebra $\mathcal{A}_{\min}$ is closed under linear combinations, $\mathcal{A}_{\min}$ must contain the operator

$$X_l = \int \mathrm{d}\, U\, \chi_U^{(l)}\ A_U\,,$$

where $\chi_U^{(l)}$ are the characters of the irreducible representations of $\mathsf{SU}(2)$ given by $\chi_U^{(l)} = \mathrm{Tr}[U^{(l)}]$. Let us set $l = N$. In this case, the orthogonality of $\mathsf{SU}(2)$ matrix elements eliminates all terms in the block decomposition of $\rho^{\otimes N}$, except for the term with $j = N/2$. Notice that in this case the multiplicity subspace is trivial. Hence, one has

$$X_N = \int \mathrm{d}\, U\, \chi_U^{(N)}\, d_{N/2}\, U^{(N/2)}\, \rho_{N/2}\, U^{(N/2)\dagger} \qquad\qquad \rho_{N/2} = \frac{e^{-\beta J_z^{(N/2)}}}{\mathrm{Tr}\left[e^{-\beta J_z^{(N/2)}}\right]}\,.$$

The matrix elements of X_N can be computed explicitly as

$$\begin{aligned}
\left\langle \frac{N}{2}, n\right| X_N \left|\frac{N}{2}, n'\right\rangle &= \frac{d_{N/2}}{\mathrm{Tr}\left[e^{-\beta J_z^{(N/2)}}\right]} \int \mathrm{d}\, U\, \chi_U^{(N)} \left[\sum_{m=-N/2}^{N/2} e^{-\beta m} \left\langle \frac{N}{2}, n\right| U^{(N/2)} \left|\frac{N}{2}, m\right\rangle \left\langle \frac{N}{2}, m\right| U^{(N/2)\dagger} \left|\frac{N}{2}, n'\right\rangle\right] \\
&= \delta_{n,n'}\,(-1)^n\, \frac{d_{N/2}\, \left\langle \frac{N}{2}, n, \frac{N}{2}, -n'|N, 0\right\rangle}{d_N\, \mathrm{Tr}\left[e^{-\beta J_z^{(N/2)}}\right]} \left[\sum_{m=-N/2}^{N/2} (-e^{-\beta})^m\, \overline{\left\langle \frac{N}{2}, m, \frac{N}{2}, -m\right| N, 0\right\rangle}\right] \\
&= \delta_{n,n'}\,(-1)^n\, \frac{d_{N/2}\, \left\langle \frac{N}{2}, n, \frac{N}{2}, -n'|N, 0\right\rangle}{d_N\, \mathrm{Tr}\left[e^{-\beta J_z^{(N/2)}}\right]} \left[\sum_{m=-N/2}^{N/2} \frac{(N!)^2(-e^{-\beta})^m}{(N/2-m)!(N/2+m)!\sqrt{(2N)!}}\right] \\
&= \delta_{n,n'}\,(-1)^{n+N/2}\, \frac{d_{N/2}(N!)e^{\beta N/2}(1-e^{-\beta})^N\, \left\langle \frac{N}{2}, n, \frac{N}{2}, -n'|N, 0\right\rangle}{d_N \sqrt{(2N)!}\, \mathrm{Tr}\left[e^{-\beta J_z^{(N/2)}}\right]}\,,
\end{aligned}$$

$\langle j_1, m_1, j_2, m_2|J, M\rangle$ denoting the Clebsch-Gordan coefficient. Note that the Clebsch-Gordan coefficient in the above expression is nonzero if and only if $n = n'$. As a consequence, the operator X_N has full support.

Now, since $\mathcal{A}_{\min}$ is an algebra, it must contain X_N as well as the whole Abelian algebra generated by it. In particular, it must contain the projector on the support of X_N—which is nothing but $P_{N/2}$, the projector on the symmetric subspace. Moreover, it must contain all the operators of the form

$$A_{U,N/2} = P_{N/2} A_U P_{N/2} \propto U^{(N/2)}\, e^{-\beta J_z^{(N/2)}}\, U^{(N/2)\dagger} \qquad \forall\, U \in \mathsf{SU}(2)\,.$$

Finally, for $\beta \neq 0$, it is easy to see that the smallest algebra $\mathcal{A}_{\min,N/2}$ containing the above operators is the algebra $\mathsf{B}(\mathcal{R}_{N/2})$. This can be easily seen by von Neumann's double commutant theorem: If an operator B commutes with the non-degenerate Hermitian operator $A_{U,N/2}$ for every U, then B must be proportional to the identity. Hence, the double commutant of $\mathcal{A}_{N/2}$—equal to $\mathcal{A}_{N/2}$ itself—is the whole $\mathsf{B}(\mathcal{R}_{N/2})$. In conclusion, we have the inclusion $\mathsf{B}(\mathcal{R}_{N/2}) \subseteq \mathcal{A}_{\min,N/2} \subseteq \mathcal{A}_{\min}$. □

Proposition 3. *If the states in Eq. (10) are neither pure nor maximally mixed, then $\mathcal{A}_{\min}$ is the full algebra generated by the N-fold tensor representation of $\mathsf{GL}(2)$, namely*

$$\mathcal{A}_{\min} = \bigoplus_{j=0}^{N/2} \left[\mathsf{B}(\mathcal{R}_j) \otimes I_{m_j}\right]\,,$$

$\mathsf{B}(\mathcal{R}_j)$ denoting the algebra of all linear operators on the representation space $\mathcal{R}_j$.

Proof. We prove that $\mathcal{A}_{\min}$ contains the algebra $\mathsf{B}(\mathcal{R}_j) \otimes I_{m_j}$ for every j. The proof is by induction, with j starting from $N/2$ and going down to 0. For $j = N/2$ we know that $\mathcal{A}_{\min}$ contains the algebra $\mathsf{B}(\mathcal{R}_{N/2})$ of all operators with support in the symmetric subspace. Let us assume that $\mathcal{A}_{\min}$ contains all the algebras $\mathsf{B}(\mathcal{R}_j) \otimes I_{m_j}$ with $j \geq j_* + 1$ and show that it must necessarily contain also the algebra $\mathsf{B}(\mathcal{R}_{j_*}) \otimes I_{m_{j_*}}$. By construction, we know that $\mathcal{A}_{\min}$ contains all the operators A_U of the form

$$A_U = \bigoplus_{j=0}^{N/2} \frac{d_j}{\mathrm{Tr}\left[e^{-\beta J_z^{(j)}}\right]} \left(U^{(j)}\, e^{-\beta J_z^{(j)}} U^{(j)\dagger} \otimes I_{m_j}\right)\,, \qquad\qquad J_z^{(j)} = \sum_{m=-j}^{j} m\, |j, m\rangle\langle j, m|\,.$$

Since the states in Eq. (10) are not pure, all the blocks in the sum are non-zero. Moreover, the induction hypothesis implies that $\mathcal{A}_{\min}$ should also contain the operators A'_U of the form

$$A'_U = \bigoplus_{j=0}^{j_*} \frac{d_j}{\mathrm{Tr}\left[e^{-\beta J_z^{(j)}}\right]} \left(U^{(j)}\, e^{-\beta J_z^{(j)}} U^{(j)\dagger} \otimes I_{m_j}\right), \qquad U \in SU(2)\,.$$

Now, we can repeat the argument used in the proof of Proposition 2: by linearity, $\mathcal{A}_{\min}$ must contain the operator

$$\begin{aligned} X_{2j_*} &= \int \mathrm{d}\, U\, \chi_U^{(2j_*)}\; A'_U \\ &= \frac{d_{j_*}}{\mathrm{Tr}\left[e^{-\beta J_z^{(j_*)}}\right]} \int \mathrm{d}\, U\, \chi_U^{(2j_*)} \;\; \left(U^{(j_*)}\, e^{-\beta J_z^{(j_*)}}\, U^{(j_*)\dagger} \otimes I_{m_{j_*}}\right). \end{aligned}$$

Explicit calculation (same as in Proposition 2) shows that X_{2j_*} has full rank. Hence, the projector on the support of X_{2j_*} is $P_{j_*} = I_{j_*} \otimes I_{m_{j_*}}$. Since $\mathcal{A}_{\min}$ should contain this projector, it must also contain all operators of the form

$$\begin{aligned} A'_{U,j_*} &= P_{j_*} A'_U P_{j_*} \\ &\propto U^{(j_*)} e^{-\beta J_z^{(j_*)}} U^{(j_*)\dagger} \otimes I_{m_{j_*}}\,, \qquad U \in SU(2)\,. \end{aligned}$$

Again, using von Neumann's double commutant theorem, it is easy to show that the smallest algebra containing all the above operators is $\mathsf{B}(\mathcal{R}_{j^*}) \otimes I_{m_{j_*}}$. In conclusion we proved that $\mathcal{A}_{\min}$ must contain $\mathsf{B}(\mathcal{R}_{j^*}) \otimes I_{m_{j_*}}$. By induction, this proves the inclusion

$$\mathcal{A}_{\min} \supseteq \bigoplus_{j=0}^{N/2} \left[\mathsf{B}(\mathcal{R}_j) \otimes I_{m_j}\right].$$

In the other hand, the definition of $\mathcal{A}_{\min}$ implies the opposite inclusion. Hence, one must have the equality. □

Zero-error compression of a complete ensemble implies zero error compression for every ensemble of permutationally invariant states

Propositions 1 and 3 imply the following

Corollary 1. *If the states (10) are neither pure nor maximally mixed, every channel $\mathcal{C}$ preserving them must preserve all permutationally invariant states.*

Proof. By Propositions 1 and 3, the channel $\mathcal{C}$ must satisfy

$$\mathsf{Fix}(\mathcal{C}) \supseteq \mathcal{A}_{\min} = \bigoplus_{j=0}^{N/2} \left[\mathsf{B}(\mathcal{R}_j) \otimes I_{m_j}\right],$$

meaning that the full algebra generated by the tensor representation of $\mathsf{GL}(2)$ is contained in the set of fixed points. □

We are now in position to prove Theorem 1 in the main text:

Proof of Theorem 1. Suppose that a compression protocol has zero error on a complete ensemble of mixed states. Then, Corollary 1 implies that the protocol should have zero error on all permutationally invariant states. In particular, the protocol should be able to transmit without error the following ensemble of orthogonal pure states

$$S := \left\{ \rho_{j,m} = |j,m\rangle\langle j,m| \otimes \frac{I_{m_j}}{m_j}, p_{j,m} = \frac{1}{D} \;\middle|\; j = 0,\dots,N/2\,, m = -j,\dots,j,\; D := \sum_j d_j \right\}.$$

A lower bound on the dimension d_{enc} of the encoding space $\mathcal{H}_{\mathrm{enc}}$ is then obtained by considering the amount of classical information carried by S. In detail, the lower bound can be calculated using the monotonicity of Holevo's chi quantity in quantum data processing. Holevo's chi quantity of S [35] is defined as follows

$$\chi\left(S\right) := H\left(\sum_{j,m} p_{j,m}\rho_{j,m}\right) - \sum_{j,m} p_{j,m} H\left(\rho_{j,m}\right)$$

with $H(\rho)$ being the von Neumann entropy of the state ρ. Since the chi quantity is non-increasing under quantum evolutions, in the zero-error scenario we have

$$\chi\left(S\right)=\chi\left(S_{\mathrm{enc}}\right) \tag{12}$$

where S_{enc} is the encoded ensemble $S_{\mathrm{enc}}:=\{\mathcal{E}(\rho_{j,m}),p_{j,m}\}$. On the other hand, the dimension of the encoding subspace is lower bounded by the chi quantity [8]

$$\log d_{\mathrm{enc}}\geq\chi\left(S_{\mathrm{enc}}\right). \tag{13}$$

The chi quantity for the ensemble S can be computed as $\chi\left(S\right)=\log D$. Combining this equality with Eqs. (12) and (13) we get

$$d_{\mathrm{enc}}\geq D=\left(\frac{N}{2}+1\right)^{2},$$

which concludes the optimality proof. The protocol showed in the main text saturates the bound. □

PROOF OF THEOREM 2

As stated in the main text, we assume $p>\frac{1}{2}$, because for $p=1/2$ the ensemble is trivial, consisting only of the maximally mixed state.

We first notice that the error of the compression protocol is upper bounded as

$$\begin{aligned}e_N&=\frac{1}{2}\left\|\rho_{\mathbf{n}}^{\otimes N}-\mathcal{D}\circ\mathcal{E}\left(\rho_{\mathbf{n}}^{\otimes N}\right)\right\|,\qquad\forall\mathbf{n}\in\mathbb{S}^2\\&=\frac{1}{2}\left\|\sum_{j\notin\mathsf{S}_\epsilon}q_{j,N}\left[\rho_{\mathbf{n},j}\otimes\frac{I_{m_j}}{m_j}-\mathcal{D}(\rho_0)\right]\right\|\\&\leq\sum_{j\notin\mathsf{S}_\epsilon}q_{j,N}\,,\end{aligned} \tag{14}$$

the last step following from the triangle inequality and from the fact that the trace distance of two states is upper bounded by 2. Note that the upper bound is independent of $\mathbf{n}$, meaning that the protocol works equally well for all states with the same spectrum (or equivalently, for all states with the same purity).

At this point, it is enough to prove that the upper bound vanishes in the large N limit. To this purpose, we use the expression for $q_{j,N}$ [Eq. (5) in the main text] and observe that one has

$$1-e_N\geq\sum_{j\in\mathsf{S}_\epsilon}\frac{2(2j+1)}{j_0}B\left(N+1,p,\frac{N}{2}+j+1\right)-\sum_{j\in\mathsf{S}_\epsilon}\frac{2(2j+1)}{j_0}B\left(N+1,p,\frac{N}{2}-j\right) \tag{15}$$

where $j_0=(2p-1)(N+1)/2$. The second summand in the r.h.s. of Eq. (15) is negligible in the large N limit: precisely, it can be bounded as

$$\begin{aligned}\sum_{j\in\mathsf{S}_\epsilon}\frac{2(2j+1)}{j_0}B\left(N+1,p,\frac{N}{2}-j\right)&\leq\sum_{j=0}^{\frac{N}{2}}\frac{2(2j+1)}{j_0}B\left(N+1,p,\frac{N}{2}-j\right)\\&\leq\frac{1}{2p-1}\sum_{j=0}^{\frac{N}{2}}B\left(N+1,p,\frac{N}{2}-j\right)\\&\leq\frac{1}{2p-1}\exp\left[-\frac{2(2p-1)^2N^2}{N+1}\right]\end{aligned} \tag{16}$$

having used the Hoeffding's inequality in the last step. Hence, this term goes to zero exponentially fast with N,

Now, recall that we chose S_ϵ to be the interval

$$\mathsf{S}_\epsilon=\left[j_0-1/2-\sqrt{N\ln(2/\epsilon)},j_0-1/2+\sqrt{N\ln(2/\epsilon)}\right]. \tag{17}$$

Setting $j_0 - j - 1/2 = x$, we then obtain

$$\begin{aligned} e_N &\leq 1 - \sum_{x=-\sqrt{N\ln(2/\epsilon)}}^{\sqrt{N\ln(2/\epsilon)}} \left(1 - \frac{x}{j_0}\right) B\left(N+1, p, p(N+1) - x\right) + \frac{1}{2p-1} \exp\left[-\frac{2(2p-1)^2 N^2}{N+1}\right] \\ &= 1 - \sum_{x=-\sqrt{N\ln(2/\epsilon)}}^{\sqrt{N\ln(2/\epsilon)}} B\left(N+1, p, p(N+1) - x\right) + \frac{1}{2p-1} \exp\left[-\frac{2(2p-1)^2 N^2}{N+1}\right] \\ &\leq 2\exp\left[\frac{2N}{N+1}\ln\frac{\epsilon}{2}\right] + \frac{1}{2p-1}\exp\left[-\frac{2(2p-1)^2N^2}{N+1}\right] \\ &\leq \epsilon^{\frac{2N}{N+1}} + \frac{1}{2p-1}\exp\left[-\frac{2(2p-1)^2N^2}{N+1}\right] \end{aligned}$$

In the second last step we have used the Hoeffding's inequality. Now it can be seen that the right hand side of the bound vanishes exponentially fast with N, and we can always find a N_0 such that $e_N \leq \epsilon^{3/2} < \epsilon$ for any $N > N_0$. The dimension of the encoded system is now

$$\begin{aligned} d_{\rm enc} &= \sum_{j \in \mathsf{S}_\epsilon} (2j+1) \\ &= 2(2p-1)\sqrt{N\ln(2/\epsilon)}(N+1) \end{aligned}$$

An upper bound on the number of required qubits is given by

$$\begin{aligned} \log d_{\rm enc} &= \log\left[2(2p-1)N\sqrt{N\ln\frac{2}{\epsilon}}\right] + \log\left(1 + \frac{1}{N}\right) \\ &\leq \frac{3}{2}\log N + \log\left[2(2p-1)\sqrt{\ln\frac{2}{\epsilon}}\right] + 1 \end{aligned}$$

□

THE PURE STATE CASE: NO DISCONTINUOUS GAP BETWEEN ZERO-ERROR AND APPROXIMATE COMPRESSION

Here we prove that the type of discontinuity highlighted by our Theorems 1 and 2 is specific to mixed states. Consider the pure state ensemble $\left\{(|\mathbf{n}\rangle\langle\mathbf{n}|)^{\otimes N}\,, \mathrm{d}^2\,\mathbf{n}\right\}$, where $|\mathbf{n}\rangle$ is the pure qubit state with Bloch vector $\mathbf{n}$ and $\mathrm{d}^2\,\mathbf{n}$ is the invariant measure on the Bloch sphere. Suppose that the state $(|\mathbf{n}\rangle\langle\mathbf{n}|)^{\otimes N}$ is encoded into a state $\rho_{\mathbf{n},\rm enc}$ on a Hilbert space of dimension $d_{\rm enc}$. Assuming that the compression error is bounded by ϵ, an argument by Horodecki [8] gives a lower bound on $d_{\rm enc}$. The argument is based on the following lemma, based on the Alicki-Fannes inequality

Lemma 1 ([36]). *Let $\{\rho_x\,, p_x\}$ be an ensemble of states and let $\{\rho_{x,\rm enc}\,, p_x\}$ be the ensemble of the encoded states. If the compression protocol has error bounded by ϵ, then the following inequality holds*

$$|\chi\left(\{\rho_x\,, p_x\}\right) - \chi\left(\{\rho_{x,\rm enc}\,, p_x\}\right)| \leq 2\left[\epsilon \log d_{\rm in} + \eta(\epsilon)\right], \tag{18}$$

where $d_{\rm in}$ is the rank of the average state $\rho = \sum_x p_x\rho_x$ and $\eta(x) = -x\ln x$.

In our case, $d_{\rm in}$ is the dimension of the symmetric subspace, namely

$$d_{\rm in} = d_{\frac{N}{2}} = N + 1\,. \tag{19}$$

Moreover, we have

$$\chi\left(\left\{(|\mathbf{n}\rangle\langle\mathbf{n}|)^{\otimes N}\,, \mathrm{d}^2\,\mathbf{n}\right\}\right) = H\left(I_{\frac{N}{2}}/d_{\frac{N}{2}}\right) = \log(N+1). \tag{20}$$

and, by the Holevo's bound [35],

$$\chi\left(\left\{\rho_{\mathbf{n},\mathrm{enc}}\,,\mathrm{d}^2\,\mathbf{n}\right\}\right)\leq\log d_{\mathrm{enc}}\,. \tag{21}$$

In our case, we have $d_{\mathrm{in}}=d_{\frac{N}{2}}=N+1$. Hence, combining Eqs. (18), (19), (20), and (21) we obtain the bound

$$\log d_{\mathrm{enc}}\geq(1-2\epsilon)\log(N+1)-2\eta(\epsilon).$$

Now, note that the r.h.s. is continuous in ϵ and tends to $\log(N+1)$ when ϵ tends to zero. The value $\log(N+1)$ is exactly the minimum number of qubits needed to encode a generic state in the symmetric subspace with zero error. Hence, as ϵ tends to zero, the number of qubits needed for approximate compression tends to the number of qubits needed for zero-error compression.

PROOF OF THEOREM 3

Here we prove the optimality of our protocol among all compression protocols where the encoding is covariant and the decoding preserves the magnitude of the total angular momentum. Precisely, we assume that

1. the encoding space $\mathcal{H}_{\mathrm{enc}}$ supports a unitary representation of the group $\mathsf{SU}(2)$, denoted by $\{V_g\mid g\in\mathsf{SU}(2)\}$
2. the encoding channel satisfies the covariance condition

$$\mathcal{E}\circ\mathcal{U}_g=\mathcal{V}_g\circ\mathcal{E}\,,\qquad\forall g\in\mathsf{SU}(2)\,, \tag{22}$$

   where $\mathcal{U}_g$ and $\mathcal{V}_g$ are the unitary channels defined by $\mathcal{U}_g(\cdot):=U_g\cdot U_g^\dagger$ and $\mathcal{V}_g=V_g\cdot V_g^\dagger$.
3. the decoding channel $\mathcal{D}$ preserve the magnitude of the total angular momentum, in the sense that, for every input state ρ, one has

$$\mathrm{Tr}\left[\mathbf{K}^2\,\mathcal{D}(\rho)\right]=\mathrm{Tr}\left[\mathbf{J}^2\,\rho\right]\,, \tag{23}$$

   where $\mathbf{K}=(K_x,K_y,K_z)$ are the generators of the representation $\{V_g\,,g\in\mathsf{SU}(2)\}$ and $\mathbf{J}=(J_x,J_y,J_z)$ are the generators of the representation $\{U_g^{\otimes N},g\in\mathsf{SU}(2)\}$.

Under these conditions, we can prove the optimality of the protocol presented in Theorem 3 of the main text.

Proof of Theorem 3. For the purpose of this proof, it is convenient to parametrize the mixed states $\rho_{\mathbf{n}}$ as $\rho_g=U_g\rho U_g^\dagger$, where ρ is a fixed state and g is a generic element of $\mathsf{SU}(2)$. Let us decompose the encoding space as

$$\mathcal{H}_{\mathrm{enc}}=\bigoplus_j\left(\mathcal{R}_j\otimes\widetilde{\mathcal{M}}_j\right)\,, \tag{24}$$

where j is the quantum number of the angular momentum, $\mathcal{R}_j$ is the corresponding representation space, and $\widetilde{\mathcal{M}}_j$ is a suitable multiplicity space. By definition, one has

$$\begin{aligned}\mathcal{H}_{\mathrm{enc}}&\supseteq\mathsf{Span}\left\{\mathsf{Supp}\left[\mathcal{E}\left(\rho_g^{\otimes N}\right)\right],g\in\mathsf{SU}(2)\right\}\\&=\mathsf{Span}\left[\mathsf{Supp}\left(\Omega\right)\right]\,,\qquad\Omega:=\int\mathrm{d}\,g\,\mathcal{E}\left(\rho_g^{\otimes N}\right)\,.\end{aligned} \tag{25}$$

Since $\mathcal{E}$ is covariant, the state Ω satisfies the relation $V_g\Omega V_g^\dagger=\Omega\,,\forall g\in\mathsf{SU}(2)$. Hence, Ω can be written in the block diagonal form

$$\Omega=\bigoplus_{j\in\mathsf{S}}\left(\frac{I_j}{d_j}\otimes\omega_j\right)\,,$$

where ω_j is a suitable state on the multiplicity space and S is a suitable set of values of the angular momentum number. Combining the above decomposition with Eq. (25), we obtain the bound

$$d_{\mathrm{enc}}\geq\mathsf{rank}\,\Omega\geq\sum_{j\in\mathsf{S}}d_j\,. \tag{26}$$

On the other hand, since the decoding preserves the magnitude of the angular momentum, one has

$$\mathrm{Tr}[\Pi_j \, \mathcal{D} \circ \mathcal{E} \left(\rho_g^{\otimes N}\right)] = \mathrm{Tr}[\widetilde{\Pi}_j \mathcal{E} \left(\rho_g^{\otimes N}\right)] \,, \qquad \forall j = 0, \dots, N/2 \,, \forall g \in \mathsf{SU}(2) \,,$$

where Π_j is the projector on $\mathcal{R}_j \otimes \mathcal{M}_j$ while $\widetilde{\Pi}_j$ is the projector on $\mathcal{R}_j \otimes \widetilde{\mathcal{M}}_j$. Hence, we have

$$\sum_{j \in \mathsf{S}} \mathrm{Tr}[\Pi_j \mathcal{D} \circ \mathcal{E} \left(\rho_g^{\otimes N}\right)] = 1 \,, \qquad \forall g \in \mathsf{SU}(2) \,, \tag{27}$$

meaning that all the output states $\mathcal{D} \circ \mathcal{E} \left(\rho_g^{\otimes N}\right)$ are contained in the subspace $\mathcal{H}_N := \bigoplus_{j \in \mathsf{S}} (\mathcal{R}_j \otimes \mathcal{M}_j)$. Hence, we have

$$\begin{aligned} e_N &= \frac{1}{2} \left\| \rho_g^{\otimes N} - \ \mathcal{D} \circ \mathcal{E} \left(\rho_g^{\otimes N}\right) \right\| \qquad \forall g \in \mathsf{SU}(2) \\ &\geq \frac{1}{2} \left\| P_N \left[\rho_g^{\otimes N} - \ \mathcal{D} \circ \mathcal{E} \left(\rho_g^{\otimes N}\right) P_N\right] \right\| + \frac{1}{2} \left\| (I^{\otimes N} - P_N) \left[\rho_g^{\otimes N} - \ \mathcal{D} \circ \mathcal{E} \left(\rho_g^{\otimes N}\right)\right] (I^{\otimes N} - P_N) \right\| \\ &= \frac{1}{2} \left\| (I^{\otimes N} - P_N) \rho_g^{\otimes N} (I^{\otimes N} - P_N) \right\| \\ &\geq \sum_{j \notin \mathsf{S}} \frac{q_{j,N}}{2} \end{aligned} \tag{28}$$

where P_N is the projector on $\mathcal{H}_N$. Now we prove that any protocol with $d_{\mathrm{enc}} = O\left(N^{3/2-\delta}\right)$, $\delta > 0$, will have a non-vanishing error. Recall from the main text that the probability distribution $q_{j,N}$ can be expressed as

$$q_{j,N} = \frac{2j+1}{2j_0} \left[B\left(N+1, p, \frac{N}{2} + j + 1\right) - B\left(N+1, p, \frac{N}{2} - j\right) \right] \tag{29}$$

where $B(n, p, k)$ is the binomial distribution with n trials and with probability p and

$$j_0 = (p - 1/2)(N+1) \,.$$

Combing Eq. (28) with Eq. (29), we have

$$e_N \geq \frac{1}{2} - \frac{1}{2} \sum_{j \in \mathsf{S}} \frac{2j+1}{2j_0} B\left(N+1, p, \frac{N}{2} + j + 1\right) .$$

We split the set S into two subsets S_1 and S_2, defined as

$$\begin{aligned} \mathsf{S}_1 &= \mathsf{S} \cap \left[j_0 - \frac{\sqrt{cN}+1}{2}, j_0 + \frac{\sqrt{cN}+1}{2} \right] \\ \mathsf{S}_2 &= \mathsf{S} \setminus \mathsf{S}_1 \end{aligned}$$

where c is an arbitrary constant. The error is then bounded as

$$e_N \geq \frac{1}{2} (1 - s_1 - s_2) \qquad s_k := \sum_{j \in \mathsf{S}_k} \frac{2j+1}{2j_0} B\left(N+1, p, \frac{N}{2} + j + 1\right) \,, \ k = 1, 2 \,. \tag{30}$$

We now bound s_1 and s_2. Let us start from s_1: by definition, we have

$$\begin{aligned} s_1 &\leq \frac{\max_{j \in \mathsf{S}_1} (2j+1)}{2j_0} \sum_{j \in \mathsf{S}_1} B\left(N+1, p, \frac{N}{2} + j + 1\right) \\ &= O(1) \sum_{j \in \mathsf{S}_1} B\left(N+1, p, \frac{N}{2} + j + 1\right) \\ &\leq O(1) \, |\mathsf{S}_1| B\left(N+1, p, \frac{N}{2} + j_0 + 1\right) \\ &= O\left(N^{-1/2}\right) \, |\mathsf{S}_1| \,. \end{aligned} \tag{31}$$

In turn, S_1 can be bounded from the relation

$$\begin{aligned} |\mathsf{S}_1| \left(\min_{j \in \mathsf{S}_1} 2j+1 \right) &\leq \sum_{j \in \mathsf{S}_1} (2j+1) \\ &\leq d_{\text{enc}} \\ &= O\left(N^{3/2-\delta} \right) , \end{aligned} \tag{32}$$

which implies $|\mathsf{S}_1| \leq O(N^{1/2-\delta})$. Inserting this relation into Eq. (31), we finally obtain

$$s_1 \leq O\left(N^{-\delta} \right) . \tag{33}$$

Regarding s_2, we have the bound

$$\begin{aligned} s_2 &\leq \frac{N+1}{j_0} \left[\sum_{j \leq j_0 - \frac{\sqrt{cN}+1}{2}} B\left(N+1, p, \frac{N}{2}+j+1 \right) \right] \\ &= \frac{1}{p-1/2} \left[\sum_{j \leq j_0 - \frac{\sqrt{cN}+1}{2}} B\left(N+1, p, \frac{N}{2}+j+1 \right) \right] \\ &\leq \frac{e^{-c/2}}{p-1/2} , \end{aligned} \tag{34}$$

the last inequality coming from Hoeffding's bound.

Finally, combining the inequalities (30), (33), and (34), we obtain the lower bound

$$e_N \geq \frac{1}{2} \left[1 - O\left(N^{-\delta} \right) - \frac{e^{-c/2}}{p-1/2} \right] ,$$

Since the constant c is arbitrary, the bound becomes $e_N \geq 1/2 - O\left(N^{-\delta} \right)$. □

UPPER BOUND ON THE COMPLEXITY OF GENERATING APPROXIMATE MAXIMALLY MIXED STATES

The decoding requires the preparation of maximally mixed states to be placed in the multiplicity register. For a given value of j, this is accomplished by generating a maximally entangled state of rank m_j. In the following we present a three-step protocol for this purpose.

1. Choose an integer $n = O(N)$ such that $m_j \in (2^{n-1}, 2^n]$. Prepare n maximally entangled qubit states. The resulting the state is $\rho = [|\Phi^+\rangle\langle\Phi^+|]^{\otimes n}$, with $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and lies in a space of dimension 2^{2n}.
2. Perform the measurement in the computational basis on one qubit of each entangled pair. The measurement outcomes of the individual qubit measurements are saved in a sequence of n binary digits, let us denote it by $\underline{y}$.
3. Compare the string $\underline{y}$ with the binary expression of m_j. If $\underline{y}$, as a number, is larger than m_j, the protocol fails and we have to restart by preparing again n maximally entangled qubits. Otherwise, we keep the remaining qubits, which, on average, will be in a maximally entangled mixed state of rank m_j.

The last step can be seen by noting down the quantum operation $\mathcal{C}_{\text{yes}}$ corresponding to the successful outcomes of the projective measurement, given by

$$\mathcal{C}_{\text{yes}}(\sigma) = \sum_{y \leq m_j} |\underline{y}\rangle\langle\underline{y}|\sigma|\underline{y}\rangle\langle\underline{y}| .$$

The protocol is successful in more than half of the cases. For that reason, the probability of failure vanishes exponentially in the number of repetitions l as $p_{\text{no}} \leq 2^{-l}$. To ensure that the error is vanishing fast enough with the number of state copies N, we repeat the protocol N times. Then, the complexity of the protocol is comprised of preparing the qubit states, which takes $O(N)$ steps, and from comparing the n digit binary strings on a classical computer, which also takes $O(N)$ steps. By repeating the protocol N times, the overall complexity yields $O(N^2)$. It is safe to run the protocol N times to assure for an exponentially vanishing error, because the complexity of the decoding is still dominated by the Schur transform.

ZERO-ERROR COMPRESSION FOR QUANTUM SYSTEMS OF DIMENSION $d > 2$

In this and the following sections, we generalize our results to quantum systems of arbitrary finite dimension $d < \infty$.

Upper bound on the number of encoding qubits

Theorem 4. *In dimension d, every ensemble of N identically prepared mixed states of rank r can be encoded without error into less than $\left(2dr - r^2 + r - 2\right)/2 \log(N + d - 1)$ qubits.*

The proof is based on the Schur-Weyl duality, which allows one to decompose the N-copy Hilbert space as

$$\mathcal{H}^{\otimes N} \simeq \bigoplus_{\lambda \in \mathcal{Y}_{N,d}} \left(\mathcal{R}_\lambda \otimes \mathcal{M}_\lambda\right) ,$$

where $\mathcal{R}_\lambda$ is a representation space, $\mathcal{M}_\lambda$ is a multiplicity space, and the sum runs over the set $\mathcal{Y}_{N,d}$ of all Young diagrams of N boxes arranged in d rows, parametrized as $\lambda = (\lambda_1, \ldots, \lambda_d)$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, $\sum_{i=1}^d \lambda_i = N$. We use the notations

$$d_\lambda = \dim \mathcal{R}_\lambda$$

and

$$m_\lambda = \dim \mathcal{M}_\lambda .$$

Relative to this decomposition, every state of the form $\rho^{\otimes N}$ where ρ has rank r can be cast into the form

$$\rho^{\otimes N} = \bigoplus_{\lambda \in \mathcal{Y}_{N,r}} q_{\lambda,N} \left(\rho_\lambda \otimes \frac{I_{m_\lambda}}{m_\lambda}\right) ,$$

where ρ_λ is a quantum state on $\mathcal{R}_\lambda$, I_{m_λ} is the identity on $\mathcal{M}_\lambda$, and $q_{\lambda,N}$ is a suitable probability distribution. Note that only the Young diagrams with r rows or less are present here (for this fact, see e.g. [37]).

The proof of Theorem 4 makes use of the following lemmas:

Lemma 2. *For every $\lambda \in \mathcal{Y}_{N,r}$, one has $d_\lambda \leq (N + d - 1)^{(2dr - r^2 - r)/2}$.*

Proof. The dimension can be expressed as

$$d_\lambda = \frac{\prod_{1 \leq i < j \leq d} (\lambda_i - \lambda_j - i + j)}{\prod_{k=1}^{d-1} k!} , \tag{35}$$

cf. Eq. (III.10) of [38]. Since $\lambda_i = 0$ for $i > r$, we have the following chain of (in)equalities

$$\begin{aligned} d_\lambda &= \frac{\prod_{1 \leq i < j \leq r} (\lambda_i - \lambda_j - i + j) \cdot \prod_{1 \leq i \leq r < j \leq d} (\lambda_i - i + j) \cdot \prod_{r < i < j \leq d} (j - i)}{\prod_{k=1}^{d-1} k!} \\ &\leq \frac{(N + r - 1)^{\binom{r}{2}} \cdot (N + d - 1)^{r(d-r)} \cdot \prod_{l=1}^{d-r-1} l!}{\prod_{k=1}^{d} k!} \\ &\leq \frac{(N + d - 1)^{(2dr - r^2 - r)/2}}{\prod_{k=d-r}^{d-1} k!} . \end{aligned}$$

□

Lemma 3. *The total dimension of all the representation spaces corresponding to Young diagrams with no more than r rows is upper bounded as*

$$\sum_{\lambda \in \mathcal{Y}_{N,r}} d_\lambda < (N + d - 1)^{\frac{2dr - r^2 + r - 2}{2}} .$$

Proof. By Lemma 2 one has

$$\begin{aligned}\sum_{\lambda\in\mathcal{Y}_{N,r}} d_\lambda &\leq (N+d-1)^{\frac{2dr-r^2-r}{2}}\ |\mathcal{Y}_{N,r}| \\ &< (N+d-1)^{\frac{2dr-r^2+r-2}{2}}\,,\end{aligned}$$

having used the equality $|\mathcal{Y}_{N,r}| = \binom{N+r-1}{r-1}$ [39] and the elementary bound $\binom{N+r-1}{r-1} < (N+1)^{r-1} \leq (N+d-1)^{r-1}$. □

Proof of Theorem 4. A zero-error compression protocol is given by the following encoding and decoding channels:

$$\begin{aligned}\mathcal{E}(\rho) &= \bigoplus_{\lambda\in\mathcal{Y}_{N,r}} \mathrm{Tr}_{\mathcal{M}_\lambda}[\Pi_\lambda\rho\Pi_\lambda] \\ \mathcal{D}(\rho') &= \bigoplus_{\lambda\in\mathcal{Y}_{N,r}} P_\lambda\rho' P_\lambda\otimes\frac{I_{m_\lambda}}{m_\lambda}\,,\end{aligned}$$

where Π_λ is the projector on $\mathcal{R}_\lambda\otimes\mathcal{M}_\lambda$ and P_λ is the projector on $\mathcal{R}_\lambda$. The encoding space is $\mathcal{H}_{\mathrm{enc}} = \bigoplus_{\lambda\in\mathcal{Y}_{N,r}}\mathcal{R}_\lambda$ and has dimension $d_{\mathrm{enc}} = \sum_{\lambda\in\mathcal{Y}_{N,r}} d_\lambda$, which we bound as

$$\begin{aligned}d_{\mathrm{enc}} &= \sum_{\lambda\in\mathcal{Y}_{N,r}} d_\lambda \\ &< (N+d-1)^{\frac{2dr-r^2+r-2}{2}}\,,\end{aligned}$$

having used Lemma 3. □

Lower bound on the number of encoding qubits used by the zero-error protocol

Here we give a lower bound on the dimension of the encoding space in the zero-error protocol described in the proof of Theorem 4. Precisely, we have the following

Lemma 4. *The total dimension of all the representation spaces corresponding to Young diagrams with no more than r rows is lower bounded as*

$$\sum_{\lambda\in\mathcal{Y}_{N,r}} d_\lambda \geq c(r,d)\, N^{\frac{2dr-r^2+r-2}{2}}\,, \tag{36}$$

where c is a suitable function.

Proof. For simplicity, we use the notation $f(N,r,d) \gtrsim g(N,r,d)$ to mean that there exists a function $c(r,d)$ such that $f(N,r,d) \geq c(r,d)g(N,r,d)$ for every N. If $f(N,r,d) \gtrsim g(N,r,d)$ and $g(N,r,d) \gtrsim f(N,r,d)$, then we write $f(N,r,d) \approx g(N,r,d)$. With this notation, we have

$$d_\lambda \gtrsim \prod_{1\leq i<j\leq d}(\lambda_i-\lambda_j)\,,$$

having used Eq. (35). Consider the case when N is a multiple of $r(r+1)/2$ and define $s = 2N/r(r+1)$. Define the subset of Yang diagrams

$$\mathsf{S}_{\mathrm{core}} = \left\{\lambda\in\mathcal{Y}_{N,r}\ |\ \lambda_i\in\left[(r-i+1)s-\frac{s}{2r},(r-i+1)s+\frac{s}{2r}\right]\,,\qquad \forall i=1,\ldots,r-1\right\}$$

For every diagram in $\mathsf{S}_{\mathrm{core}}$ we have the lower bound

$$\begin{aligned}d_\lambda &\gtrsim \left[\prod_{1\leq i<j\leq r-1}(\lambda_i-\lambda_j)\right]\left[\prod_{1\leq i\leq r-1}(\lambda_i-\lambda_r)\right]\left[\prod_{1\leq i<r<j\leq d}\lambda_i\right]\left[\prod_{r<j\leq d}\lambda_r\right] \\ &\geq \left\{\prod_{1\leq i<j\leq r}\left[(j-i)s-\frac{s}{r}\right]\right\}\left\{\prod_{1\leq i\leq r-1}\left[(r-i)s-\frac{s}{2}\right]\right\}\left\{\prod_{1\leq i\leq r<j\leq d}(r-i)s\right\}\left\{\prod_{r<j\leq d}\frac{s}{2}\right\} \\ &\approx s^{\frac{2dr-r^2-r}{2}} \\ &\approx N^{\frac{2dr-r^2-r}{2}}\,.\end{aligned} \tag{37}$$

Now, the total dimension of the subspaces with Young diagrams in $\mathsf{S}_{\rm core}$ an be lower bounded as

$$\begin{aligned}\sum_{\lambda\in\mathsf{S}_{\rm core}} d_\lambda &\gtrsim N^{\frac{2dr-r^2-r}{2}}\,|\mathsf{S}_{\rm core}|\\ &= N^{\frac{2dr-r^2-r}{2}}\,\left(\frac{s}{r}\right)^{r-1}\\ &\approx N^{\frac{2dr-r^2-r}{2}}\,N^{r-1}\\ &= N^{\frac{2rd-r^2+r-2}{2}}\,.\end{aligned}$$

Since $\mathsf{S}_{\rm core}$ is a subset of $\mathcal{Y}_{N,r}$, we obtain Eq. (36). □

Following the steps adopted in the $d = 2$ case, it is also possible to show that the upper bound of Lemma 4 is actually an upper bound for *every* zero-error protocol that works for a *complete* ensemble of mixed states—i. e. for an ensemble of the form $\{\rho_g^{\otimes N}\,, p_g\}$ where the state ρ_g is non-degenerate and the probability distribution p_g is dense on $\mathsf{SU}(d)$. Essentially, the argument is based on the use of Proposition 3, which can be applied here to all the $\mathsf{SU}(2)$ subgroups of $\mathsf{SU}(d)$.

APPROXIMATE COMPRESSION FOR QUANTUM SYSTEMS OF DIMENSION $d > 2$

Compression protocol

Here we consider ensembles of N identically prepared mixed states, each of them having the same spectrum. Every such ensemble can be written in the form $\{\rho_g^{\otimes N}, p_g\}$, where ρ_g is a density matrix of the form

$$\rho_g = U_g\rho_0 U_g^\dagger\,,\qquad g\in\mathsf{SU}(d)\,,$$

ρ_0 is a rank-r density matrix with non-degenerate positive eigenvalues, and p_g is a probability distribution over the group $\mathsf{SU}(d)$. For ensembles of this form, we have the following

Theorem 5. *For every $\epsilon > 0$ there exists an integer N_0 such that for every $N \geq N_0$ the ensemble $\{\rho_g^{\otimes N}\,, p_g\}$ can be compressed with error less than ϵ into $N_{\rm enc}$ qubits, with*

$$N_{\rm enc} \leq \frac{2dr-r^2-1-m}{2}\log(N+d-1) + \frac{m+r-1}{2}\log\left[4d(d+1)\ln(N+1) + 8\ln\left(\frac{1}{\epsilon}\right) + O\left(\frac{1}{\sqrt{N}}\right)\right]$$

and $m := \sum_{i=1}^r \mu_i$, where μ_i be the cardinality of the set $\{j : j > i\,, p_j = p_i\}$. We notice that $m = 0$ when the spectrum is non-degenerate.

The proof of the theorem is based on the Schur-Weyl decomposition

$$\rho_g^{\otimes N} = \bigoplus_{\lambda\in\mathcal{Y}_{N,r}} q_{\lambda,N}\left(U_g^{(\lambda)}\,\rho_{0,\lambda}\,U_g^{(\lambda)\,\dagger}\otimes\frac{I_{m_\lambda}}{m_\lambda}\right)\,, \tag{38}$$

where $\rho_{0,\lambda}$ is a fixed density matrix on $\mathcal{R}_\lambda$ and $U_g^{(\lambda)}$ is the irreducible representation of $\mathsf{SU}(d)$ acting on $\mathcal{R}_\lambda$. The key point is that the probability distribution $q_{\lambda,N}$ is concentrated on the Young diagrams such that the vector

$$p_\lambda := \left(\frac{\lambda_1}{N},\ldots,\frac{\lambda_d}{N}\right) \tag{39}$$

is close to the vector of the eigenvalues of ρ_0 [40, 41], listed as

$$p = (p_1,\ldots,p_d)\,,\qquad p_1\geq p_2\geq\cdots\geq p_r > p_{r+1} = \cdots = p_d = 0\,. \tag{40}$$

Precisely, we will use the following

Lemma 5 ([40, 41])**.** *Let p_λ and p be the vectors defined in Eqs. (39) and (40), respectively, and let $d(a,b) := \frac{1}{2}\sum_i |a_i - b_i|$ be the total variation distance between two vectors. Then, one has*

$$\mathsf{Prob}\left[\lambda : d(p_\lambda, p) > x\right] \leq (N+1)^{d(d+1)/2} \cdot e^{-2Nx^2} ,$$

with $\mathsf{Prob}\left[\lambda : d(p_\lambda, p) > x\right] := \sum_{\lambda:\, d(p_\lambda,p)>x} q_{\lambda,N}$, $q_{N,\lambda}$ being the probability distribution in Eq. (38).

The idea of the proof is to discard all Young diagrams whose probability vector p_λ falls outside in a ball of size $O(1/\sqrt{N})$ around the vector p. The dimensions of the subspaces associated to the remaining diagrams can be bounded with the following

Lemma 6. *The maximum dimension of a subspace $\mathcal{R}_\lambda$ satisfying $d(p_\lambda, p) \leq x$ is upper bounded as*

$$d_\lambda \leq (4Nx + r)^m \, (N + d - 1)^{\frac{2dr - r(r+1)}{2} - m} . \tag{41}$$

Proof. The dimension can be bounded as

$$\begin{aligned}
d_\lambda &= \frac{\prod_{1\leq i<j\leq d}(\lambda_i - \lambda_j - i + j)}{\prod_{k=1}^{d-1} k!} \\
&\leq \frac{\prod_{1\leq i\leq r}\left\{\left[\prod_{i<j\leq i+\mu_i}(\lambda_i - \lambda_j - i + j)\right]\left[\prod_{i+\mu_i<j\leq d}(\lambda_i - \lambda_j - i + j)\right]\right\}}{\prod_{k=1}^{d-1} k!} \\
&\leq \frac{\prod_{1\leq i\leq r}\left\{\left[\prod_{i<j\leq i+\mu_i}(4Nx + \mu_i)\right]\left[\prod_{i+\mu_i<j\leq d}(N + d - 1)\right]\right\}}{\prod_{k=1}^{d-1} k!} \\
&\leq \frac{\prod_{1\leq i\leq r}(4Nx + \mu_i)^{\mu_i}\,(N + d - 1)^{d - i - \mu_i}}{\prod_{k=1}^{d-1} k!} \\
&\leq \frac{(4Nx + r)^m\,(N + d - 1)^{\frac{2dr - r(r+1)}{2} - m}}{\prod_{k=1}^{d-1} k!} ,
\end{aligned}$$

having used the fact that the ball $\mathsf{S} = \{\lambda \in \mathcal{Y}_{N,r} : d(p_\lambda, p) \leq x\}$ is contained in the hypercube $\mathsf{S}' = \{\lambda \in \mathcal{Y}_{N,r} : |\lambda_i/N - p_i| \leq 2x\,, \forall i = 1, \ldots, r-1\}$, so that, for $p_i = p_j$, $i < j$, one has $\lambda_i - \lambda_j \leq 4Nx$. □

Lemma 7. *The total dimension of the subspaces satisfying $d(p_\lambda, p) \leq x$ satisfies*

$$\sum_{\lambda\in\mathcal{Y}_{N,r}:\, d(p_\lambda,p)\leq x} d_\lambda \leq (N + d - 1)^{\frac{2dr - r(r+1)}{2} - m}(4Nx + r)^{m+r-1} .$$

Proof. Immediate from Lemma 6 and from the fact that the ball $\mathsf{S} = \{\lambda \in \mathcal{Y}_{N,r} : d(p_\lambda, p) \leq x\}$ is contained in the hypercube $\mathsf{S}' = \{\lambda \in \mathcal{Y}_{N,r} : |\lambda_i/N - p_i| \leq 2x\,, \forall i = 1, \ldots, r-1\}$, yielding the bound

$$|\mathsf{S}| \leq |\mathsf{S}'| \leq (4Nx)^{r-1} .$$

□

Proof of Theorem 5. To compress within an error ϵ, we choose the encoding and decoding channels

$$\begin{aligned}
\mathcal{E}(\rho) &= \bigoplus_{\lambda\in\mathsf{S}_\epsilon} \mathrm{Tr}_{\mathcal{M}_\lambda}[\Pi_\lambda \rho \Pi_\lambda] \;\oplus\; \mathrm{Tr}\left[\rho\left(I^{\otimes N} - \Pi_\epsilon\right)\right]\,\rho_{\mathrm{fail}} \\
\mathcal{D}(\rho') &= \bigoplus_{\lambda\in\mathsf{S}_\epsilon} \left(P_\lambda \rho' P_\lambda \otimes \frac{I_{m_\lambda}}{m_\lambda}\right) ,
\end{aligned}$$

with $\Pi_\epsilon = \bigoplus_{\lambda\in\mathsf{S}_\epsilon} \Pi_\lambda$, $\mathsf{Supp}(\rho_{\mathrm{fail}}) \subseteq \mathcal{H}_{\mathrm{enc}} = \bigoplus_{\lambda\in\mathsf{S}_\epsilon} \mathcal{R}_\lambda$, and

$$\mathsf{S}_\epsilon := \{\lambda \in \mathcal{Y}_{N,r} \mid d(p_\lambda, p) \leq x_\epsilon\} , \qquad x_\epsilon = \sqrt{\frac{d(d+1)/2\ln(N+1) + \ln(1/\epsilon)}{2N}} .$$

The value of x_ϵ is chosen in order to bound the compression error as

$$\begin{aligned}
e_N &= \frac{1}{2}\left\|\mathcal{D}\circ\mathcal{E}\left(\rho_g^{\otimes N}\right)-\rho_g^{\otimes N}\right\| \qquad \forall g\in\mathsf{SU}(d)\\
&\leq \frac{1}{2}\,\mathrm{Tr}\left[\rho\left(I^{\otimes N}-\Pi_\epsilon\right)\right]\,\left\|\mathcal{D}(\rho_{\rm fail})-\rho_{g,{\rm fail}}\right\|\,, \qquad \rho_{g,{\rm fail}}:=\bigoplus_{\lambda\not\in\mathsf{S}_\epsilon}\frac{q_{\lambda,N}}{\mathrm{Tr}\left[\rho\left(I^{\otimes N}-\Pi_\epsilon\right)\right]}\left(U_g^{(\lambda)}\,\rho_{0,\lambda}\,U_g^{(\lambda)\,\dagger}\otimes\frac{I_{m_\lambda}}{m_\lambda}\right)\\
&\leq \mathrm{Tr}\left[\rho\left(I^{\otimes N}-\Pi_\epsilon\right)\right]\\
&= \sum_{\lambda\not\in\mathsf{S}_\epsilon} q_{\lambda,N}\\
&\leq (N+1)^{d(d+1)/2}\cdot e^{-2Nx^2}\\
&= \epsilon\,,
\end{aligned}$$

the last inequality coming from Lemma 5. On the other hand, the encoding subspace has dimension

$$\begin{aligned}
d_{\rm enc} &= \sum_{\lambda\in\mathsf{S}_\epsilon} d_\lambda\\
&\leq (N+d-1)^{dr-\frac{r(r+1)}{2}-m}(4Nx+r)^{m+r-1}\\
&\leq (N+d-1)^{dr-\frac{r(r+1)}{2}-m}\,N^{\frac{m+r-1}{2}}\,\left[4d(d+1)\ln(N+1)+8\ln\left(\frac{1}{\epsilon}\right)+O\left(\frac{1}{\sqrt{N}}\right)\right]^{\frac{m+r-1}{2}}\\
&\leq (N+d-1)^{\frac{2dr-r^2-1-m}{2}}\,\left[4d(d+1)\ln(N+1)+8\ln\left(\frac{1}{\epsilon}\right)+O\left(\frac{1}{\sqrt{N}}\right)\right]^{\frac{m+r-1}{2}}\,,
\end{aligned}$$

having used Lemma 7 and the definition of x_ϵ. Hence, the number of encoding qubits satisfies

$$\begin{aligned}
N_{\rm enc} &\leq \log d_{\rm enc}\\
&\leq \frac{2rd-r^2-1-m}{2}\log(N+d-1)+\frac{m+r-1}{2}\log\left[4d(d+1)\ln(N+1)+8\ln\left(\frac{1}{\epsilon}\right)+O\left(\frac{1}{\sqrt{N}}\right)\right]\,.
\end{aligned}$$

□

Optimality proof in the presence of symmetry

Here we prove the converse of Theorem 5. Our proof is valid for protocols where the encoding is covariant and the decoding preserves the *nonabelian charges* [42] identified by the Young diagrams. Precisely, we assume that

1. the encoding space $\mathcal{H}_{\rm enc}$ supports a unitary representation of the group $\mathsf{SU}(d)$, denoted by $\{V_g \mid g\in\mathsf{SU}(d)\}$.
2. the encoding channel satisfies the covariance condition $\mathcal{E}\circ\mathcal{U}_g=\mathcal{V}_g\circ\mathcal{E}$, $\forall g\in\mathsf{SU}(d)$.
3. the decoding channel $\mathcal{D}$ preserves the nonabelian charges associated to $\mathsf{SU}(d)$, namely, for every input state ρ, one has

$$\mathrm{Tr}\left[\Pi_\lambda\,\mathcal{D}(\rho)\right]=\mathrm{Tr}\left[\widetilde{\Pi}_\lambda\,\rho\right] \qquad \forall\lambda\in\mathcal{Y}_{N,d}\,, \tag{42}$$

where $\widetilde{\Pi}_\lambda$ is the projector on the direct sum of all the invariant subspaces of $\mathcal{H}_{\rm enc}$ with Young diagram λ.

By the same argument as in the qubit case, the error of the compression protocol satisfying the above assumption can be lower bounded as $e_N\geq(1/2)\sum_{\lambda\in\mathsf{S}}q_{\lambda,N}$, with S being a subset of $\mathcal{Y}_{N,r}$ specified by the protocol. The encoding dimension is given by $d_{\rm enc}=\sum_{\lambda\in\mathsf{S}}q_{\lambda,N}$. We have the following theorem.

Theorem 6. *Every compression protocol that encodes a complete N-qubit ensemble into*

$$\left(\frac{2dr-r^2-1-m}{2}-\delta\right)\log N\,, \qquad \delta>0\,,$$

qubits with covariant encoding and a decoding that preserves the nonabelian charges will necessarily have error $e\geq 1/2$ in the asymptotic limit. Here $m:=\sum_{i=1}^r\mu_i$, where μ_i be the cardinality of the set $\{j:j>i\,,p_j=p_i\}$. We notice that $m=0$ when the spectrum is non-degenerate.

To prove the theorem, we first define the cubic lattice

$$\mathsf{H}_\epsilon = \left\{ \lambda \in \mathcal{Y}_{N,r} \;\middle|\; \lambda_i \in \left[p_i N - \frac{\sqrt{N^{1+\epsilon}}}{2}, p_i N + \frac{\sqrt{N^{1+\epsilon}}}{2} \right], \quad \forall\, i = 1, \ldots, r-1 \right\} \tag{43}$$

for any constant $\epsilon \in (0,1)$. With this definition, the sum of the probability $q_{\lambda,N}$ when $\lambda \notin \mathsf{H}_\epsilon$ vanishes exponentially in N. Precisely, we have the following lemma.

Lemma 8. *For the set H_ϵ defined by Eq. (43), the following bound holds.*

$$\sum_{\lambda \notin \mathsf{H}_\epsilon} q_{\lambda,N} \leq (N+1)^{\frac{d(d+1)}{2}} e^{-\frac{N^\epsilon}{8}}.$$

Proof. For any Young diagram λ not in the set H_ϵ, there exist at least one j such that $|\lambda_j - p_i N| \geq \sqrt{N^{1+\epsilon}}/2$. Thus we have

$$d(p_\lambda, p) \geq \frac{1}{2} \left| \frac{\lambda_j}{N} - p_j \right| \geq \frac{1}{4\sqrt{N^{1-\epsilon}}}.$$

Substituting this fact into Lemma 5, we immediately get the following lemma. □

Now we start to bound the probability distribution $q_{\lambda,N}$ within the set H_ϵ. Notice that the exact expression of $q_{\lambda,N}$ is given as [37]

$$q_{\lambda,N} = \frac{\det \Delta}{\det \Sigma} \cdot m_\lambda \tag{44}$$

where the matrix Σ is independent of N (and thus its expression is not relevant to bounding the probability) and the matrix Δ is a rank r square matrix defined as the following.

$$\Delta_{ij} = \left[\prod_{\beta=0}^{\mu_j - 1} (\lambda_i + r - i - \beta) \right] p_j^{\lambda_i + r - i - \mu_j}, \tag{45}$$

with μ_i defined in Theorem 6. Notice that we follow the convention $\prod_{i=0}^{-1} f(i) = 1$. We first prove the following bound of $\det \Delta$.

Lemma 9. *For any λ in the set H_ϵ defined by Eq. (43), the following bound holds asymptotically for large N:*

$$\det \Delta \lesssim N^{\frac{(1+\epsilon)m}{2}} \left(\prod_{i=1}^{r} p_i^{\lambda_i} \right), \qquad m = \sum_{i=1}^{r} \mu_i \,.$$

Proof. Suppose that there are k distinct positive values in the spectrum, and the i-th biggest value has degeneracy r_i. We can then divide the set $\{1, \ldots, r\}$ into k subsets $\mathsf{L}_1 \cup \cdots \cup \mathsf{L}_k$, corresponding to the distinct eigenvalues, so that L_i is the set of indices corresponding to the i-th biggest eigenvalue. Recalling that r_j is the degeneracy of the j-th eigenvalue, we have

$$\mathsf{L}_i = \left\{ \sum_{j=1}^{i-1} r_j + 1, \ldots, \sum_{j=1}^{i} r_j \right\}.$$

Notice that, by definition, one has

$$p_l = p_k \qquad \forall\, l, k \in \mathsf{L}_i. \tag{46}$$

With the above definition, the spectrum now reads

$$\underbrace{p_1 = \cdots = p_{r_1}}_{\mathsf{L}_1} > \underbrace{p_{r_1+1} = \cdots = p_{r_1+r_2}}_{\mathsf{L}_2} > \cdots > \underbrace{p_{\sum_{i=1}^{k-1} r_i + 1} = \cdots = p_r}_{\mathsf{L}_k} > p_{r+1} = \cdots = p_d = 0.$$

Correspondingly, we define a subgroup P_r of the group S_r, consisting of the product of permutations that act within the subsets $\{\mathsf{L}_i\}$. Precisely,

$$\mathsf{P}_r := \left\{ \sigma^{(1)} \times \sigma^{(2)} \times \cdots \times \sigma^{(k)} \,|\, \sigma^{(i)} \in \mathsf{S}_{r_i}; i = 1, \ldots, k \right\}.$$

With the above definition, we divide $\det \Delta$ into two terms

$$\begin{aligned} \det \Delta &= t_1 + t_2 \\ t_1 &= \sum_{\sigma \in \mathsf{P}_r} \operatorname{sgn}(\sigma) \left(\prod_{i=1}^{r} \Delta_{i\,\sigma_i} \right) \\ t_2 &= \sum_{\sigma \notin \mathsf{P}_r} \operatorname{sgn}(\sigma) \left(\prod_{i=1}^{r} \Delta_{i\,\sigma_i} \right), \end{aligned} \tag{47}$$

denoting by σ_i the index that comes from applying σ to i.

Let us bound t_1. By definition, P_r contains every permutation σ such that $p_i = p_{\sigma_i}$ for every i. Therefore, we have

$$\begin{aligned} t_1 &= \sum_{\sigma \in \mathsf{P}_r} \operatorname{sgn}(\sigma) \left\{ \prod_{i=1}^{r} \left[\prod_{\beta=0}^{\mu_{\sigma_i}-1} (\lambda_i + r - i - \beta) \right] p_{\sigma_i}^{\lambda_i + r - i - \mu_{\sigma_i}} \right\} \\ &= \sum_{\sigma \in \mathsf{P}_r} \operatorname{sgn}(\sigma) \left\{ \prod_{i=1}^{r} \left[\prod_{\beta=0}^{\mu_{\sigma_i}-1} (\lambda_i + r - i - \beta) \right] p_{i}^{\lambda_i + r - i} \right\} \left(\prod_{i=1}^{r} p_{\sigma_i}^{-\mu_{\sigma_i}} \right) \\ &= \sum_{\sigma \in \mathsf{P}_r} \operatorname{sgn}(\sigma) \left\{ \prod_{i=1}^{r} \left[\prod_{\beta=0}^{\mu_{\sigma_i}-1} (\lambda_i + r - i - \beta) \right] p_{i}^{\lambda_i + r - i} \right\} \left(\prod_{i=1}^{r} p_{i}^{-\mu_i} \right) \\ &= \left(\prod_{i=1}^{r} p_i^{\lambda_i + r - i - \mu_i} \right) \sum_{\sigma \in \mathsf{P}_r} \operatorname{sgn}(\sigma) \left[\prod_{i=1}^{r} \prod_{\beta=0}^{\mu_{\sigma_i}-1} (\lambda_i + r - i - \beta) \right]. \end{aligned}$$

Since i and σ_i are always in the same subset L_l (for suitable l), we can rewrite the term $\prod_{i=1}^{r} \prod_{\beta=0}^{\mu_{\sigma_i}-1} (\lambda_i + r - i - \beta)$ as $\prod_{l=1}^{k} \prod_{i \in \mathsf{L}_l} \prod_{\beta=0}^{\mu_{\sigma_i}-1} (\lambda_i + r - i - \beta)$. We then have

$$\begin{aligned} t_1 &= \left(\prod_{i=1}^{r} p_i^{\lambda_i + r - i - \mu_i} \right) \sum_{\sigma \in \mathsf{P}_r} \operatorname{sgn}(\sigma) \left\{ \prod_{l=1}^{k} \left[\prod_{i \in \mathsf{L}_l} \prod_{\beta=0}^{\mu_{\sigma_i}-1} (\lambda_i + r - i - \beta) \right] \right\} \\ &= \left(\prod_{i=1}^{r} p_i^{\lambda_i + r - i - \mu_i} \right) \prod_{l=1}^{k} \left\{ \sum_{\sigma^{(l)} \in \mathsf{S}_{r_l}} \operatorname{sgn}\left(\sigma^{(l)}\right) \left[\prod_{i \in \mathsf{L}_l} \prod_{\beta=0}^{\mu_{\sigma_i^{(l)}}-1} (\lambda_i + r - i - \beta) \right] \right\} \\ &= \left(\prod_{i=1}^{r} p_i^{\lambda_i + r - i - \mu_i} \right) \prod_{l=1}^{k} \left\{ \sum_{\sigma^{(l)} \in \mathsf{S}_{r_l}} \operatorname{sgn}\left(\sigma^{(l)}\right) \left[\prod_{i \in \mathsf{L}_l} (\Delta_l)_{i \sigma_i^{(l)}} \right] \right\} \\ &= \left(\prod_{i=1}^{r} p_i^{\lambda_i + r - i - \mu_i} \right) \left(\prod_{l=1}^{k} \det \Delta_l \right). \end{aligned}$$

Here Δ_l is a rank r_l square matrix defined as

$$(\Delta_l)_{ij} = \prod_{\beta=0}^{r_l - j - 1} (\lambda_i + r - i - \beta),$$

observing that μ_j assumes the values $r_l - 1, r_l - 2, \ldots, 1, 0$ for the indices in L_l. The determinant of Δ_l equals to

$\prod_{1\le i<j\le r_l}(\lambda_i-\lambda_j+j-i)$. Combining this with the definition of H_ϵ (43), we have

$$\begin{aligned}
t_1 &= \left[\prod_{l=1}^{k}\prod_{1\le i<j\le r_l}(\lambda_i-\lambda_j+j-i)\right]\left(\prod_{i=1}^{r}p_i^{\lambda_i+r-i-\mu_i}\right)\\
&\lesssim \left[\prod_{l=1}^{k}\left(\sqrt{N^{1+\epsilon}}\right)^{\frac{r_l(r_l-1)}{2}}\right]\left(\prod_{i=1}^{r}p_i^{\lambda_i+r-i-\mu_i}\right)\\
&= N^{\frac{(1+\epsilon)m}{2}}\left(\prod_{i=1}^{r}p_i^{\lambda_i+r-i-\mu_i}\right)\\
&\approx N^{\frac{(1+\epsilon)m}{2}}\left(\prod_{i=1}^{r}p_i^{\lambda_i}\right). \qquad (48)
\end{aligned}$$

The last step follows from the fact that

$$m=\sum_{i=1}^{r}\mu_i=\sum_{i=1}^{k}\sum_{j=1}^{r_i}(r_i-j).$$

Next, we bound the second term t_2 in Eq. (47) as

$$\begin{aligned}
t_2 &\le \sum_{\sigma\notin\mathsf{P}_r}\left(\prod_{i=1}^{r}\Delta_{i\,\sigma_i}\right)\\
&= \sum_{\sigma\notin\mathsf{P}_r}\left\{\prod_{i=1}^{r}\left[\prod_{j=0}^{\mu_{\sigma_i}-1}(\lambda_i+r-i-j)\right]p_{\sigma_i}^{\lambda_i+r-i-\mu_{\sigma_i}}\right\}\\
&\le \sum_{\sigma\notin\mathsf{P}_r}\left[\prod_{i=1}^{r}(N+r-1)^{\mu_{\sigma_i}}p_{\sigma_i}^{\lambda_i+r-i-\mu_{\sigma_i}}\right]\\
&= (N+r-1)^m\sum_{\sigma\notin\mathsf{P}_r}\left[\prod_{i=1}^{r}p_{\sigma_i}^{\lambda_i+r-i-\mu_{\sigma_i}}\right]\\
&= (N+r-1)^m\sum_{\sigma\notin\mathsf{P}_r}\left[\prod_{i=1}^{r}\left(\frac{p_{\sigma_i}}{p_i}\right)^{\lambda_i+r-j-\mu_{\sigma_j}}\right]\left[\prod_{j=1}^{r}p_j^{\lambda_j+r-j-\mu_{\sigma_j}}\right]\\
&= (N+r-1)^m\sum_{\sigma\notin\mathsf{P}_r}\left[\prod_{i=1}^{r}\left(\frac{p_{\sigma_i}}{p_i}\right)^{Np_i+O(\sqrt{N^{1+\epsilon}})}\right]\left[\prod_{j=1}^{r}p_j^{\lambda_j+r-j-\mu_{\sigma_j}}\right]\\
&\approx (N+r-1)^m\sum_{\sigma\notin\mathsf{P}_r}\exp\left[-ND(p||\sigma_p)\right]\left[\prod_{i=1}^{r}p_i^{\lambda_i+r-i-\mu_{\sigma_i}}\right],
\end{aligned}$$

where $D(p||q):=\sum_i p_i\ln(p_i/q_i)$ is the Kullback-Leibler divergence and $\sigma_p:=(\sigma_{p_1},\ldots,\sigma_{p_r})$. Now, since $\sigma\notin\mathsf{P}_r$, we always have $D(p||\sigma_p)>0$. Therefore, the second term in Eq. (47) vanishes exponential in N. Combining this fact with Eq. (47) and Eq. (48) we get the desired bound on $\det\Delta$. □

Lemma 10. *For any λ in the set H_ϵ defined by Eq. (43), the following bound holds asymptotically for large N.*

$$\frac{q_{\lambda,N}}{d_\lambda}\lesssim N^{-\frac{2dr-r^2-1-(1+\epsilon)m}{2}}.$$

Proof. The dimension of $\mathcal{M}_\lambda$ is given by

$$m_\lambda=\frac{N!}{\prod_{i=1}^{d}(\lambda_i+d-i)!}\prod_{1\le i<j\le d}(\lambda_i-\lambda_j+j-i)$$

(see e. g. [37]) and can be bounded as

$$\begin{aligned} m_\lambda &\leq \frac{1}{\lambda_1^{d-1}\,\lambda_2^{d-2}\,\dots\,\lambda_r^{d-r}} \binom{N}{\lambda} \prod_{1\leq i<j\leq d} (\lambda_i - \lambda_j + j - i) \\ &\lesssim N^{-\frac{2dr-r^2-r}{2}} \binom{N}{\lambda} \prod_{1\leq i<j\leq d} (\lambda_i - \lambda_j + j - i) \end{aligned}$$

for any $\lambda \in \mathsf{H}_\epsilon$. Substituting the above bound and the bound in Lemma 9 into Eq. (44), we have

$$\begin{aligned} q_{\lambda,N} &\lesssim \frac{N^{\frac{(1+\epsilon)m}{2}}}{\det\Sigma} \left(\prod_{i=1}^r p_i^{\lambda_i}\right) \cdot N^{-\frac{2dr-r^2-r}{2}} \binom{N}{\lambda} \prod_{1\leq i<j\leq d} (\lambda_i - \lambda_j + j - i) \\ &\lesssim N^{-\frac{2dr-r^2-r-(1+\epsilon)m}{2}} m(N,p,\lambda) \prod_{1\leq i<j\leq d} (\lambda_i - \lambda_j + j - i) \\ &\lesssim N^{-\frac{2dr-r^2-1-(1+\epsilon)m}{2}} \prod_{1\leq i<j\leq d} (\lambda_i - \lambda_j + j - i) \end{aligned}$$

which holds for any $\lambda \in \mathsf{H}_\epsilon$. The last inequality comes from the upper bound of the multinomial $m(N,p,\lambda)$. Finally, we get the desired bound of $q_{\lambda,N}/d_\lambda$ by combining the above bound with the expression of d_λ

$$d_\lambda = \frac{\prod_{1\leq i<j\leq d}(\lambda_i - \lambda_j - i + j)}{\prod_{k=1}^{d-1} k!}.$$

□

Finally, we can bound the error of any compression protocol with an encoding set S and with the encoding dimension $d_{\text{enc}} = O\left(N^{\frac{2dr-r^2-1-m}{2}-\delta}\right)$ as

$$\begin{aligned} e_N &\geq \frac{1}{2}\sum_{\lambda\in\mathsf{S}} q_{\lambda,N} \\ &= \frac{1}{2}\left(1 - \sum_{\lambda\notin\mathsf{S}} q_{\lambda,N}\right) \\ &\geq \frac{1}{2}\left(1 - \sum_{\lambda\notin\mathsf{H}_{\delta/m}} q_{\lambda,N} - \sum_{\lambda\in\mathsf{H}_{\delta/m}\cap\mathsf{S}} q_{\lambda,N}\right) \\ &\geq \frac{1}{2}\left[1 - \sum_{\lambda\notin\mathsf{H}_{\delta/m}} q_{\lambda,N} - \max_{\lambda\in\mathsf{H}_{\delta/m}}\left(\frac{q_{\lambda,N}}{d_\lambda}\right)\sum_{\lambda\in\mathsf{S}} d_\lambda\right] \\ &\geq \frac{1}{2}\left[1 - \sum_{\lambda\notin\mathsf{H}_{\delta/m}} q_{\lambda,N} - \max_{\lambda\in\mathsf{H}_{\delta/m}}\left(\frac{q_{\lambda,N}}{d_\lambda}\right)\cdot d_{\text{enc}}\right] \\ &\gtrsim \frac{1}{2}\left[1 - (N+1)^{\frac{d(d+1)}{2}} e^{-\frac{1}{8}N^{\frac{\delta}{m}}} - N^{-\frac{\delta}{2}}\right] \\ &= \frac{1}{2}\left(1 - N^{-\frac{\delta}{2}}\right). \end{aligned}$$