Universal quantum data compression via nondestructive tomography

Charles H. Bennett,[1,*] Aram W. Harrow,[2,†] and Seth Lloyd[3,‡]
[1]*IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, New York 10598, USA*
[2]*Department of Computer Science, University of Bristol, Bristol, BS8 1UB, United Kingdom*
[3]*Department of Mechanical Engineering, MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
(Received 11 March 2004; revised manuscript received 5 December 2005; published 24 March 2006)

Quantum-state tomography—the practice of estimating a quantum state by performing measurements on it—is useful in a variety of contexts. We introduce "gentle tomography" as a version of tomography that preserves the measured quantum data. As an application of gentle tomography, we describe a polynomial-time method for universal source coding.

DOI: 10.1103/PhysRevA.73.032336 PACS number(s): 03.67.Lx, 03.65.Wj

I. INTRODUCTION

Suppose that we have a sequence of quantum states, each drawn from an ensemble with known density matrix ρ. Schumacher compression then allows the sequence to be efficiently encoded so that $S(\rho)=-\mathrm{tr}\,\rho\log\rho$ qubits are required to encode each state in the limit that the length of the sequence goes to infinity [1] (in this paper log and exp are base 2). This resembles classical source coding, in which a source can be compressed to a rate asymptotically approaching its Shannon entropy. However, classical compression can be performed by algorithms that are *universal* (do not depend on a description of the source) and *efficient* (have a running time polynomial in the length of the input). In contrast, most existing quantum compression algorithms either rely on knowing the basis in which ρ is diagonal [1] or have no known polynomial time implementations [2,3].

This paper presents an efficient, universal, quantum data compression algorithm; that is, knowing only the dimension d and the number of copies n, it can compress an unknown idenpendent and identically distributed (i.d.d.) quantum source $\rho^{\otimes n}$ in poly(n,d) time to a rate converging to its von Neumann entropy $S(\rho)$ and with error approaching zero as n increases. Another efficient universal quantum data compression algorithm was presented in [4], but our algorithm has the advantages of simplicity and a better rate-disturbance trade-off.

Our algorithm consists of two parts: a weak measurement of $\rho^{\otimes n}$ that estimates ρ accurately without causing very much damage to the state, followed by compressing $\rho^{\otimes n}$ based on this estimate. Conceptually, this resembles classical methods of compression which determine the empirical distribution of their input in their first pass over the data and perform the compression in the second pass. The only new difficulties we will encounter in the quantum case involve performing state tomography on ρ without causing very much damage and compressing ρ based on an imperfect estimate.

*Electronic address: bennetc@watson.ibm.com
†Electronic address: a.harrow@bris.ac.uk
‡Electronic address: slloyd@mit.edu

II. GENTLE TOMOGRAPHY

The problem of weakly measuring states of the form $\rho^{\otimes n}$ was introduced in [5] and further developed in [3,6]. While it is impossible to measure a single state ρ without causing disturbance, we expect ordinary classical logic to apply to $\rho^{\otimes n}$ when n is large, so that it is possible to measure even noncommuting observables precisely with little disturbance. For example, in nuclear magnetic resonance, the total x magnetization of $n=O(10^{20})$ nuclear spins is continuously measured without causing decoherence by a probe consisting of a coil of wire around the sample. This is possible because the measurement does not precisely determine the number of nuclear spins pointing in the x direction, but only gives a crude estimate of the quantity. In this section, we will introduce a procedure for state tomography on $\rho^{\otimes n}$ and then show how to modify it so that its disturbance vanishes for large n while at the same time it yields an asymptotically accurate estimate of ρ.

Let $\{\sigma_k\}_{k=1}^{d^2-1}$ be an orthonormal $(\mathrm{tr}\,\sigma_j\sigma_k=\delta_{jk})$ basis of traceless Hermitian $d\times d$ matrices, and write the density matrix ρ as $\rho=I/d+\Sigma_k(\mathrm{tr}\,\rho\sigma_k)\sigma_k$. Estimating ρ reduces to estimating the d^2-1 quantities $\mathrm{tr}\,\rho\sigma_k$. If we now diagonalize σ_k as $\sigma_k=\Sigma_{i=1}^{d}\lambda_i|v_i\rangle\langle v_i|$, then $\mathrm{tr}\,\rho\sigma_k=\Sigma_i\lambda_i\langle v_i|\rho|v_i\rangle$, so state tomography reduces to estimating $d(d^2-1)$ quantites of the form $\langle\phi|\rho|\phi\rangle$ and then performing a classical computation.

If we did not mind damaging the state, then one method of estimating $\alpha:=\langle\phi|\rho|\phi\rangle$ would be to apply the projective measurement $\{|\phi\rangle\langle\phi|, I-|\phi\rangle\langle\phi|\}$ to each copy of ρ. The number of occurrences of $|\phi\rangle\langle\phi|$ would be binomially distributed with mean $n\alpha$ and variance $n\alpha(1-\alpha)\leqslant n/4$, so we could reliably estimate α to an accuracy of $O(n^{-1/2})$. Of course, this measurement would drastically damage some states, such as $(1/\sqrt{2})(|\phi\rangle+|\phi^{\perp}\rangle)$.

Instead of measuring each state individually, we can also express this measurement as a collective operation on all n states simultaneously. It is given by the operators

$$M_k=\sum_{\substack{x\in\{0,1\}^n\\|x|=k}}\bigotimes_{i=1}^{n}x_i|\phi\rangle\langle\phi|+(1-x_i)(\mathbb{1}-|\phi\rangle\langle\phi|), \qquad (1)$$

where k ranges from 0 to n and $|x|$ denotes the number of 1's in the n-bit string x. Clearly, measuring $\{M_k\}$ yields the same

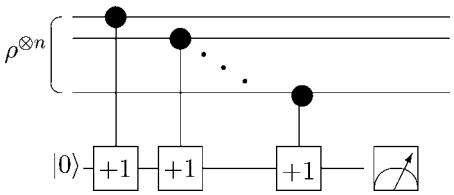

FIG. 1. Circuit for performing the measurement in Eq. (1). The controlled-(+1) operations map $|\phi\rangle|x\rangle$ to $|\phi\rangle|x+1\rangle$ for any value x of the target and leave other states unchanged.

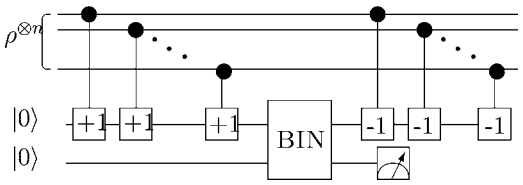

FIG. 2. Circuit for performing the gentle measurement in Eq. (2). The controlled-(+1) and controlled-(−1) operations act on the target only when the control is in the state $|\phi\rangle$. The gate BIN classically computes which bin contains the top register and stores it in the bottom register.

statistics as measuring each state individually and counting the $|\phi\rangle\langle\phi|$ outcomes. The measurement can also be constructed efficiently: we unitarily count the number of occurences of $|\phi\rangle$ in the n states in an ancilla register and then measure the ancilla (see Fig. 1).

Unfortunately, even the collective measurement in Eq. (1) causes substantial damage to the state. For example, if the measurement $\{M_k\}$ is repeated, then the distribution of k will have a variance of $O(n)$ the first time and 0 on subsequent measurements.

In [6] this problem was solved by initalizing the ancilla in Fig. 1 to the state $\Sigma_k e^{-k^2/2\Delta^2}|k\rangle$ instead of $|0\rangle$. The measurement of k then has variance $\Delta^2+O(n)$ and it can be shown [7] that the damage to $\rho^{\otimes n}$ is $O(n/(\Delta^2+n))$. Reference [3] proposed a method which causes more damage to the state, but is easier to analyze for our purposes.

To implement the gentle measurement of [3], we will divide up the range from 0 to n into m bins, with boundaries $0=b_0\leqslant b_1\cdots\leqslant b_m=n+1$. Define the function $\mathrm{BIN}(k) := \min\{j: k<b_j\}$ to be the index of the bin containing k. Then we will modify the collective measurement of Eq. (1) to measure only $\mathrm{BIN}(k)$ instead of determining k exactly. The new measurement $\{M'_j\}$ is given in terms of the M_k of Eq. (1) by

$$M'_j = \sum_{b_{j-1}\leqslant k<b_j} M_k, \tag{2}$$

where j ranges from 1 to m.

If the average bin size $(n+1)/m$ is much larger than the $O(\sqrt{n})$ width of $\rho^{\otimes n}$, then we expect to project onto a measurement outcome that contains almost all of the support of $\rho^{\otimes n}$, thereby causing little disturbance. Since we want to avoid having a bin boundary within $O(\sqrt{n})$ of the state, *for any choice of* ρ, we will choose the b_j uniformly at random from between 0 and n.

The choice of m now defines a trade-off between disturbance caused to $\rho^{\otimes n}$ and information gained about ρ. Choosing a smaller m means that each bin is larger, so that a measurement outcome lets us infer less about ρ, but we have a smaller probability of damaging $\rho^{\otimes n}$ by projecting onto only part of its support.

Proposition 1. The measurement $\{M'_j\}$ described above can be implemented in $O(n)$ gates. If we choose $m=n^s$ for $0<s<1/2$, then the measurement will fail with probability $O(n^{s-1/2}\log n)$. Upon success, the measurement outcome is within $O(n^{1-s}\log n)$ of $n\alpha$ and the disturbance (in the sense of entanglement fidelity) is less than $\exp[-O(\log^2 n)] \leqslant O(n^{-p})$ for any constant p.

Proof of proposition 1. We begin by describing how to implement $\{M'_j\}$. First we count the number of times $|\phi\rangle$ occurs in $\rho^{\otimes n}$ and store the result $k\in\{0,\ldots,n\}$ in an ancilla register. Then we coherently calculate $\mathrm{BIN}(k)$, mapping $|k\rangle|0\rangle$ to $|k\rangle|\mathrm{BIN}(k)\rangle$. Then we reverse our calculation of k, leaving only the register $|\mathrm{BIN}(k)\rangle$. If we measure this register and call the outcome j, we end up implementing the projective measurement M'_j. This is demonstrated in Fig. 2.

We define three possible causes of failure: (i) some b_i will be too close to $n\alpha$ (within $n^{1/2}\log n$), (ii) there will not be any b_i on either side of $n\alpha$ within $n^{1-s}\log n$, and (iii) measuring M'_j will yield a bin that does not contain $n\alpha$. Failure event (i) is the union of m different events ($|b_j-n\alpha| < n^{1/2}\log n$), each with probability $P_j\leqslant(2n^{1/2}\log n)/n$, so by the union bound the total probability of (i) is $\leqslant\Sigma_j P_j \leqslant m(2n^{-1/2}\log n)=2n^{s-1/2}\log n$. Next, the probability that no b_i is in $[n\alpha-n^{1-s}\log n, n\alpha]$ is $\leqslant(1-n^{-s}\log n)^m\leqslant e^{-\log n} =n^{-1}$ and likewise for the interval $[n\alpha, n\alpha+n^{1-s}\log n]$, so the probability of (ii) is $\leqslant 2/n$. Finally, suppose there is no bin within $n^{1/2}\log n$ of $n\alpha$ [i.e., (i) has not occurred]. Thus, the probability of outcome M'_j after performing the measurement in Eq. (2) is at least as high as the probability that $|k-n\alpha|<n^{1/2}\log n$ after performing the measurement in Eq. (1). Since k is the sum of n independent 0-1 random variables with expectation α, we can use a Chernoff bound [8] to show that the probability of (iii) is less than $\exp[-O(\log^2 n)]$. Thus, the possibility of failure is dominated by the probability of (i), which is $O(n^{s-1/2}\log n)$.

We say that the gentle measurement is successful if none of (i), (ii), or (iii) occurs. In this case, we can take as our estimate for α an arbitrary value within the bin we have measured and by (ii) will err by no more than $2n^{-s}\log n$. Finally, let M'_j be the measurement outcome we obtain, let $|\varphi\rangle_{AB}$ be a purification of $\rho_A^{\otimes n}$, and define $\pi := M'_j\otimes I_B$. Then the post-measurement state is $|\varphi'\rangle=\pi|\varphi\rangle/\sqrt{\langle\varphi|\pi|\varphi\rangle}$ and the entanglement fidelity is $F_e=\langle\varphi|\varphi'\rangle=\langle\varphi|\pi|\varphi\rangle/\sqrt{\langle\varphi|\pi|\varphi\rangle} =\sqrt{\langle\varphi|\pi|\varphi\rangle}$. From (iii) we have $\langle\varphi|\pi|\varphi\rangle\geqslant 1-\epsilon$ where $\epsilon=\exp[-O(\log^2 n)]$, so $F_e\geqslant\sqrt{1-\epsilon}=1-\exp[-O(\log^2 n)]$.[1] □

To perform gentle tomography we simply divide the n states into $d(d^2-1)$ blocks of length $l=\lfloor n/d(d^2-1)\rfloor$ and gently measure each block. If $\{|v_i^{(k)}\rangle\}_{i=1}^d$ is the basis for σ_k, then

[1] A similar result was proved in lemma 9 of [9].

we can index the blocks by $i=1,\ldots,d$ and $k=1,\ldots,d^2-1$ and measure $|v_i^{(k)}\rangle$ on block (i,k).

Proposition 2 (gentle tomography). For any $0<s<1/2$ and fixed Hilbert space dimension, applying the procedure described above to $\rho^{\otimes n}$ requires poly(n) time and fails with probability $O(n^{s-1/2}\log n)$. Upon success, the disturbance is less than $O(n^{-2})$ and the estimate $\tilde{\rho}$ satisfies $\|\rho-\tilde{\rho}\|_1 \leq O(n^{-s}\log n)$.

Proof. We say that tomography succeeds when each of the $d(d^2-1)$ measurements succeeds individually. Since the dimension d is a constant, we can use proposition 1 to bound the failure probability by $O(d^{3(3/2-s)}n^{s-1/2}\log n) \sim O(n^{s-1/2}\log n)$ and the state disturbance by $O(d^9 n^{-2}) \sim O(n^{-2})$.

We still need to describe how to form an accurate estimate $\tilde{\rho}$. Assume that each gentle measurement has succeeded. Then the $d(d^2-1)$ gentle measurements output, not state estimates, but bins $(b_1,b_2,|\phi\rangle)$, guaranteeing only that $b_1 \leq n\langle\phi|\rho|\phi\rangle \leq b_2$. We will try to find a state $\tilde{\rho}$ that is consistent with each bin. Since ρ is consistent with each bin, we know that some such $\tilde{\rho}$ exists. It satisfies constraints corresponding to a semidefinite program: $\tilde{\rho}\geq 0$, $\mathrm{tr}\,\tilde{\rho}=1$, and $b_1 \leq \langle\phi|\tilde{\rho}|\phi\rangle \leq b_2$ for each bin $(b_1,b_2,|\phi\rangle)$. Thus, if all of the gentle measurements succeed, we can find a valid $\tilde{\rho}$ using standard methods for solving semidefinite programs [10] in time poly$(d,\log n)$.[2]

Given such a $\tilde{\rho}$, the condition on bins implies that for each gentle measurement $|\langle\phi|(\rho-\tilde{\rho})|\phi\rangle|<\epsilon$, where $\epsilon=O(n^{-s}\log n)$. Then, if $\sigma_k=\Sigma_i\lambda_i|v_i\rangle\langle v_i|$, $|\mathrm{tr}(\rho-\tilde{\rho})\sigma_k| = |\Sigma_i\lambda_i\langle v_i|(\rho-\tilde{\rho})|v_i\rangle| \leq \epsilon\Sigma_i|\lambda_i| \leq \sqrt{d}\epsilon$. Thus, by the Cauchy-Schwartz inequality,

$$\|\rho-\tilde{\rho}\|_1 \leq d\|\rho-\tilde{\rho}\|_2 = d\sqrt{\sum_k [\mathrm{tr}(\rho-\tilde{\rho})\sigma_k]^2} \leq d^{5/2}\epsilon\,. \;\square$$

This extends our trade-off curve for gentle measurements to full gentle state tomography.

III. UNIVERSAL COMPRESSION

Now look more closely at the quantum coding. Schumacher compression works by identifying the eigenvalues and eigenvectors of ρ, then coherently performing classical Shannon compression on sequences of those eigenvectors with probabilities given by the corresponding eigenvalues. However, we are forced to operate with only an estimate $\tilde{\rho}\approx\rho$, so we will need to use a data compression scheme that deals well with small inaccuracies in the state estimate.

This case has been analyzed in [4], which found that compressing ρ in the basis $\{|i\rangle\}$ with any classical algorithm gives an asymptotic rate of $R=-\Sigma_i\langle i|\rho|i\rangle\log\langle i|\rho|i\rangle$. This is because compressing ρ faithfully reduces to compressing the diagonal entries of ρ in an arbitrary basis $\{|i\rangle\}$. Due to the non-negativity of the relative entropy $[S(\rho\|\sigma) = \mathrm{tr}\,\rho(\log\rho-\log\sigma)\geq 0]$, we have $R\leq -\mathrm{tr}\,\rho\log\sigma = S(\rho)+S(\rho\|\sigma)$ for any density matrix σ that can be diagonalized as $\sigma=\Sigma_i p_i|i\rangle\langle i|$. Thus, for any density matrix σ, we can encode ρ by diagonalizing it in the basis of σ and then using a classical reversible algorithm. This will achieve a rate $R\leq S(\rho)+S(\rho\|\sigma)$.

Unfortunately, there is no simple bound for $S(\rho\|\tilde{\rho})$ in terms of $\|\rho-\tilde{\rho}\|_1$; in fact, the relative entropy can be infinite if the support of ρ is not contained within the support of $\tilde{\rho}$. This problem corresponds to the situation when our state estimate has led the encoder to believe that certain vectors will never appear, so that when it encounters them in ρ, it has made no provision to deal with them. The solution to this is simple: assume that any input vector has a small, but nonzero, chance of occurring. This means that instead of encoding according to $\tilde{\rho}$, we will use $\tilde{\rho}_\delta := (1-\delta)\tilde{\rho}+\delta I/d$ as our state estimate, for some small $\delta>0$.

Suppose that after performing gentle tomography, $\|\tilde{\rho}-\rho\|_1<\epsilon$. Then, if we choose $\epsilon,\delta=O(n^{-s}\log n)$, we can bound the rate by

$$R \leq -\mathrm{tr}\rho\log\tilde{\rho}_\delta \leq S(\tilde{\rho}_\delta)+O(n^{-s}\log^2 n) \leq S(\rho) + O(n^{-s}\log^2 n)$$

The second inequality follows from the operator inequality $\tilde{\rho}_\delta \geq \delta I/d$ [implying $\|-\log\tilde{\rho}_\delta\|_\infty \leq \log(d/\delta)=O(\log n)$] and the bound $\mathrm{tr}\,AB \leq \|A\|_1\|B\|_\infty$ applied to $\mathrm{tr}(\tilde{\rho}_\delta-\rho)\log\tilde{\rho}_\delta$. The last inequality is due to Fannes' inequality[12]. We have neglected the inefficiency of the classical coding, since we can choose it to be $O(n^{-s})$ and it will incur only exponentially small damage for $s<\frac{1}{2}$.

To analyze the errors, note that since we usually cannot tell when tomography has failed, we ought to consider failure to be another form of disturbance. Thus, the $O(n^{s-1/2})$ probability of failure dominates the state disturbance and the errors from classical coding. This is consistent with the observation in [3] that universal compression schemes have yet to achieve better than a polynomially vanishing error.

Since our compression algorithm outputs a variable number of qubits, damage to the encoded state is not the only possible form of error. Upon failure, our algorithm risks producing a string length well above the $n[S(\rho)+n^{-s}\log^2 n]$ qubits we expect; in fact, the only absolute bound we can establish is $n\log d$ qubits. Fortunately, the probability that $\rho^{\otimes n}$ is compressed to nR qubits for $R>S(\rho)$ decreases as $O(\exp(-nK))$ for some constant K depending only on ρ and R. Following [3], we define this *overflow exponent* as

$$K = \lim_{n\to\infty}\frac{-1}{n}\log[\text{probability that } \rho^{\otimes n} \text{ yields} \geqslant nR \text{ qubits}]. \tag{3}$$

The codes described in [3] achieve the optimal value of K: $\inf_{\sigma:H(\sigma)\geqslant R} S(\sigma\|\rho)$. In contrast, our algorithm[3] achieves

[2] If one of the gentle measurements fails, this semidefinite program may fail or it may report a totally erroneous answer.

[3] It is possible to gently measure $\mathrm{tr}\,\rho\sigma_k$ directly, instead of inferring it from d gentle measurements of σ_k's eigenvectors. Using this for gentle tomography results in a compression scheme with an overflow exponent d times higher, though still not optimal.

$$K = \inf_{\sigma: H(\sigma) \geq R} \frac{1}{d(d^2-1)} \sum_{k=1}^{d^2-1} S(M_k(\sigma) \| M_k(\rho)), \qquad (4)$$

where M_k denotes the operation of measuring in the eigenbasis of σ_k [i.e., $M_k(\rho) = \Sigma_i |v_i^{(k)}\rangle\langle v_i^{(k)}| \rho |v_i^{(k)}\rangle\langle v_i^{(k)}|$].

To review, our encoding procedure is the following.

(i) Perform gentle tomography on $\rho^{\otimes n}$ using n^s bins, yielding an estimate $\tilde{\rho}$.

(ii) Construct a modified estimate $\tilde{\rho}_\delta = (1-\delta)\tilde{\rho} + \delta I/d$ for $\delta = O(n^{-s})$.

(iii) Coherently encode $\rho^{\otimes n}$ with an efficient classical algorithm (such as arithmetic coding [11]) using the basis of $\tilde{\rho}_\delta$ as the computational basis.

(iv) Attach an approximate classical description of $\tilde{\rho}_\delta$ with $O(d^2\sqrt{n})$ bits of precision and a $\lceil \log(n \log d) \rceil$-bit register indicating the length of the compressed data.

The decoding procedure is simply to extract the description of $\tilde{\rho}_\delta$ and use it as the basis for a classical decoding algorithm.

IV. CONCLUSION

We have described a polynomial time algorithm for compressing $\rho^{\otimes n}$ into $nS(\rho) + O(n^{-s} \log^2 n)$ qubits with error rate $O(n^{s-1/2} \log n)$. This matches the error rate and inefficiency of the proof of [3], though not their overflow exponent. The procedure of [4], on the other hand, can only achieve a compression rate of $S(\rho) + O(n^{-s})$ by incurring an error rate of $O(n^{-1/2+s(1+d^2)})$ (possibly up to logarithmic factors) and an overflow exponent of zero. For example, compressing qubits with constant error is only possible at a rate of $S(\rho) + O(n^{-1/10})$.

There are a number of directions left for future research. Besides finding asymptotically optimal compression and gentle tomography trade-off curves for the i.i.d. case, it would be interesting to find a method for ergodic sources analogous to Lempel-Ziv-Walsh coding that adaptively created a quantum dictionary and compressed quantum information on the fly. Alternatively, it may be that single-pass quantum compression with sublinear quantum memory cannot achieve an asymptotically vanishing error.

The method we have described in our paper could be easily made more efficient in a number of ways. However, we hope it will serve as a simple demonstration of how even unknown quantum states are amenable to classical techniques in the limit of many identical copies.

ACKNOWLEDGMENTS

This work was partially supported by the Hewlett Packard-MIT foundation (HP-MIT), by the ARO under a MURI program, by ARDA via NRO, and by the NSA and ARDA under Contract No. DAAD19-01-1-06. We are grateful to I. Chuang, K. Matsumoto, R. Schack, and B. Schumacher for helpful discussions.

[1] B. Schumacher, Phys. Rev. A **51**, 2738 (1995); R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[2] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **81**, 1714 (1998).
[3] M. Hayashi and K. Matsumoto, Phys. Rev. A **66**, 022311 (2002).
[4] R. Jozsa and S. Presnell, Proc. R. Soc. London, Ser. A **459**, 3061 (2003).
[5] C. M. Caves, K. S. Thorne, R. W. P. Drever, V. D. Sandberg, and M. Zimmerman, Rev. Mod. Phys. **52**, 341 (1980).
[6] S. Lloyd and Jean-Jacques E. Slotine, Phys. Rev. A **62**, 012307 (2000).
[7] A. Harrow and S. Lloyd (unpublished).
[8] P. R. Chernoff, J. Funct. Anal. **2**, 238 (1968).
[9] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).
[10] L. Vandenberghe and S. Boyd, SIAM Rev. **38**, 49 (1996).
[11] I. L. Chuang and D. S. Modha, IEEE Trans. Inf. Theory **46**, 1104 (2000).
[12] M. Fannes, Commun. Math. Phys. **31**, 291 (1973).