Universal Quantum Information Compression

Richard Jozsa*
School of Mathematics and Statistics, University of Plymouth, Plymouth, Devon PL4 8AA, England

Michał Horodecki†
Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland

Paweł Horodecki‡
Faculty of Applied Physics and Mathematics, Technical University of Gdańsk, 80-952 Gdańsk, Poland

Ryszard Horodecki§
Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland
(Received 7 May 1998)

Suppose that a finite dimensional quantum source is known to have von Neumann entropy less than or equal to S but is otherwise completely unspecified. We describe a method of universal quantum data compression which will faithfully compress the quantum information of any such source to S qubits per signal (in the limit of large block lengths). [S0031-9007(98)06944-0]

PACS numbers: 03.67.–a

The question of compressibility of information is one of the central issues in information theory. For classical information Shannon's noiseless coding theorem [1,2] provides a tight bound (equal to the Shannon entropy of the source) on the extent to which information may be compressed. For quantum information an analogous tight bound (equal to the von Neumann entropy of the source) was established by Schumacher [3] and further developed in [4,5]. The methods of information compression which are generally used to establish these results are *source specific;* i.e., they apply only to each given source separately. As elaborated below, the classical compression protocol requires knowledge of the probability distribution of the source, and the quantum compression protocol requires knowledge of the density matrix of the source. In this Letter we will consider the question of *universal* quantum information compression. Is there a protocol which will faithfully compress quantum information even if we do not know the density matrix of the source? More precisely, suppose that all we know about the source is that its von Neumann entropy does not exceed some given value S. Is it then still possible to faithfully compress the quantum information to S qubits per signal? Remarkably, in the case of classical information such universal compression schemes are known to exist. An explicit example is a scheme based on the theory of types developed by Csiszar and Körner [6] (which is also described in Sec. 12.3 of [2]). In this Letter we will establish the existence of universal compression schemes for *quantum* information generated by sources of finite dimension. (Henceforth all sources will be assumed to have this property.)

We begin with an outline of some source-specific compression schemes which may be used to realize the Shannon and Schumacher bounds. Later our main results will be related to an extension of constructions occurring in these schemes. Consider a source of classical information which generates signal i with probability p_i. Note that the signals may be faithfully represented using $\log N$ bits/signal by just using their names (here N is the number of signals; in this Letter logarithms are always to base 2). Let $S = -\sum_i p_i \log p_i$ be the Shannon entropy of the source (which is always $\leq \log N$). Shannon's theorem asserts that the signals may be represented asymptotically faithfully using only S bits/signal and no fewer number of bits can suffice for this task. Thus a sender (Alice) can communicate the sequence of generated signals to a receiver (Bob) by sending S bits/signal, and this transmission rate is optimal. The compression may be achieved by the following method of *block coding*, i.e., processing long sequences of signals rather than individual signals themselves separately. Note that we do not require that Bob is able to recover the signals perfectly but only that the probability of any error tends to zero in the limit of increasing block length. (This is the meaning of the term "asymptotically faithfully.") Consider all possible signal sequences $i_1 i_2 \dots i_n$ of length n (with associated probability $p_{i_1} p_{i_2} \dots p_{i_n}$). Let $\mathrm{SEQ}(n)$ be the set of all such sequences of length n. Our basic ingredient is the theorem of typical sequences [2] which asserts the following.

Theorem of typical sequences.—For any given $\epsilon > 0$ and $\delta > 0$ and for all sufficiently large n there is a subset $\mathrm{TYP}(n) \subseteq \mathrm{SEQ}(n)$ which has size $2^{n(S+\delta)}$ [i.e., an exponentially small fraction of $\mathrm{SEQ}(n)$] but whose total probability exceeds $1 - \epsilon$ (i.e., is as high as desired). The sequences in $\mathrm{TYP}(n)$ are called *typical* sequences, and those not in $\mathrm{TYP}(n)$ are called *atypical* sequences.

Intuitively this theorem asserts that (for all sufficiently large n) any sequence of signals generated by the source may be assumed with arbitrarily high probability, to be

 0031-9007/98/81(8)/1714(4)$15.00

a typical sequence. Thus to achieve compression to S bits/signal Alice and Bob set up a list of names of all the typical sequences [requiring $n(S + \delta)$ bits per typical sequence]. Then for sequences of length n generated by the source Alice sends the name of the sequence if it is a typical sequence and the name of some fixed chosen typical sequence if it is atypical. In the latter case Bob will be unable to regenerate the correct message and an error will have occurred. However, according to the theorem of typical sequences, this can be arranged to occur with arbitrarily small probability by choosing n large enough.

The compression of quantum information was first considered by Schumacher [3], who developed a quantum analog of Shannon's theorem. The quantum compression protocol was subsequently simplified by Jozsa and Schumacher [4] (hereafter referred to as the JS protocol) and later Barnum *et al.* [5] showed that the limit of compression provided by the JS protocol is optimal, i.e., that no other conceivable compression protocol can provide further asymptotically faithful compression.

Consider a source of quantum states which produces pure states $|\psi_i\rangle \in \mathcal{H}$ with probabilities p_i. Let $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ be the density matrix of the source and let $S(\rho) = -\mathrm{tr}\, \rho \log \rho$ be its von Neumann entropy. Then the JS protocol [4] provides asymptotically faithful compression to $S(\rho)$ qubits per signal state. The method rests again on the theorem of typical sequences above. Note that the density matrix of all signal sequences of length n is just $\rho^{\otimes n} = \rho \otimes \ldots \otimes \rho$. Let λ_i denote the eigenvalues of ρ so that the eigenvalues of $\rho^{\otimes n}$ are given by all products of the form $\lambda_{i_1 \ldots i_n} = \lambda_{i_1} \ldots \lambda_{i_n}$. Let $\Lambda(n)$ be the subspace of $\mathcal{H}^{\otimes n}$ given by the span of all eigenstates $|\lambda_{i_1} \ldots \lambda_{i_n}\rangle$ corresponding to all *typical* sequences $i_1 \ldots i_n$ of eigenvalues. $\Lambda(n)$ is called the *typical subspace* (for block length n). Since the Shannon entropy of the distribution λ_i is equal to the von Neumann entropy $S(\rho)$ we see that $\dim \Lambda(n) = 2^{n(S(\rho)+\delta)}$; i.e., the typical subspace occupies about $nS(\rho)$ qubits. Let Π denote the projection onto the typical subspace. Then by considering $\rho^{\otimes n}$ in its eigenbasis and recalling the theorem of typical sequences we easily see that

$$\mathrm{tr}\, \rho^{\otimes n} \Pi > 1 - \epsilon \,. \tag{1}$$

This gives the JS compression protocol: for sufficiently large n Alice accumulates a sequence of n signal states $|\psi_{\mathrm{in}}\rangle = |\psi_{j_1}\rangle \ldots |\psi_{j_n}\rangle$ and performs a measurement which determines whether the joint state lies in $\Lambda(n)$ or its orthogonal complement; i.e., the joint state is projected into one or other of these complementary subspaces. If the state projects to $\Lambda(n)$ Alice sends the resulting $n(S(\rho) + \delta)$ qubits to Bob. If it projects to the orthogonal complement (which occurs with probability $< \epsilon$) she sends to Bob any chosen state of $\Lambda(n)$. Now, as proved in Ref. [4], Eq. (1) implies that

$$\overline{\langle\psi_{\mathrm{in}}|\rho_{\mathrm{out}}|\psi_{\mathrm{in}}\rangle} > 1 - 2\epsilon \,,$$

where ρ_{out} is the state obtained by Bob if $|\psi_{\mathrm{in}}\rangle$ was generated by the source (and the average denoted by the overbar is taken over all input blocks of signals $|\psi_{\mathrm{in}}\rangle$). Thus, Bob receives the state $|\psi_{\mathrm{in}}\rangle$ with arbitrarily high fidelity [3,7] and in the limit of $\delta \to 0$ only S qubits/signal were transmitted.

We now come to the issue of *universal* compression. The above compression schemes based on typical sequences and the typical subspace are source specific. For classical compression we need to know the probability distribution of the source in order to identify the typical sequences. For quantum compression we need to know the density matrix of the source to identify the typical subspace. In the case of classical universal compression [6] explicit knowledge of the probability distribution is not required—knowledge of a bound S on the Shannon entropy of the source suffices to compress the information to S bits/signal. Our main result below will show that, similarly, knowledge of the density matrix of a quantum source is not required to achieve faithful compression of quantum information—there exists a universal quantum compression protocol which will faithfully asymptotically compress any quantum source with von Neumann entropy $\leq S$ to S qubits per signal [8].

But first, to illustrate the utility of the result, consider the recently investigated problem of compression of quantum information with incomplete data [9]. Namely, suppose that the information about the source is obtained via measurements performed over a subensemble of the generated signal sequence. Suppose further that the set of measured observables was too small to ensure a complete reconstruction of the density matrix of the source. The question was as follows: what is the maximal possible compression rate R allowing faithful transmission in this case? It has been pointed out in [9] that the Jaynes maximal entropy principle [10] places a lower bound on R,

$$R \geq S_J \,. \tag{2}$$

Here S_J is maximal entropy admissible by the measured mean values (Jaynes entropy). It has also been shown that for any *qubit* source the inequality passes into equality. Now, applying the universal quantum compression protocol we obtain that the equality holds in the general case, so that the Jaynes entropy gives the optimal compression with incomplete experimental data characterizing the source.

We will now briefly outline a method of classical universal data compression. Suppose we have a classical source and we know only that its Shannon entropy is less than some given number S (and we do not know its probability distribution). Then a result of Csiszar and Körner [6] shows that there exists a set of sequences $\mathrm{CK}(n) \subseteq \mathrm{SEQ}(n)$ of length n (whose description depends only on the value of S) which satisfies all of the properties enjoyed by $\mathrm{TYP}(n)$ in the theorem of typical sequences *not only for some one probability distribution with Shannon entropy S but simultaneously for all distributions with entropy $\leq S$;* i.e., the total probability of CK(n) with respect to any such distribution exceeds $1 - \epsilon$ and the size of CK(n) is $2^{n(S+\delta)}$. The explicit construction of CK(n) is also described in Sec. 12.3 of [2]. Hence if we replace TYP(n) by CK(n) in the classical compression scheme described previously we will have a universal compression scheme which faithfully asymptotically compresses any source with Shannon entropy $\leq S$ to S bits/signal.

Consider next the prospect of replacing TYP(n) by CK(n) in the JS protocol. It is not difficult to see that this modified protocol will faithfully compress to S qubits/signal all those quantum sources whose density matrices *commute with* ρ and have von Neumann entropy $\leq S$. Thus this does not provide a fully universal quantum compression scheme: if we consider *all* possible sources with von Neumann entropy $\leq S$ then their density matrices need not commute. Below we describe an alternative quantum compression scheme which is fully universal.

For any given ρ let Ξ be the subspace of $\mathcal{H}^{\otimes n}$ in the modified JS protocol, which is spanned by all eigenstates of $\rho^{\otimes n}$ labeled by CK sequences; i.e., Ξ is the analog of the typical subspace $\Lambda(n)$. Thus projection onto Ξ will achieve faithful compression for all sources with von Neumann entropy $\leq S$ whose density matrices commute with ρ. A set of mutually commuting density matrices is characterized by the corresponding common eigenbasis, and this may be any chosen orthonormal basis of $\mathcal{H}$. Thus as ρ varies over *all* possible density matrices with von Neumann entropy $\leq S$ there will be a subspace Ξ associated with each choice of orthonormal basis of $\mathcal{H}$. We make this dependence explicit by writing $\Xi(B)$ (where B denotes an orthonormal basis of $\mathcal{H}$), and we suppress explicit mention of the values of n and S on which Ξ also depends.

Now let Υ be the smallest subspace of $\mathcal{H}^{\otimes n}$ which contains $\Xi(B)$ for *all* choices of basis B. Then projection into Υ will achieve quantum compression for all sources with von Neumann entropy $\leq S$. Below we will prove that

$$\dim \Upsilon \leq (n + 1)^{d^2} 2^{n(S+\delta)}, \qquad (3)$$

where $d = \dim \mathcal{H}$, n is the block length, and $\delta > 0$ may be as small as desired. Thus we will achieve universal compression to R qubits/signal where R is given by

$$R = \lim_{n\to\infty} \frac{\log \dim \Upsilon}{n} \leq \lim_{n\to\infty} d^2 \frac{\log(n + 1)}{n} + S + \delta,$$

which tends to $S + \delta$ qubits/signal. Since δ can be as small as desired, asymptotically we have S qubits/signal. This is our universal quantum information compression scheme.

To prove (3) let $B^0 = \{e_1^0, \ldots, e_d^0\}$ be any fixed chosen orthonormal basis of $\mathcal{H}$. Then any other basis $B = \{e_1, \ldots, e_d\}$ is obtained from B^0 by applying some $d \times d$ unitary transformation U. Now $\Xi(B)$ is the span of $2^{n(S+\delta)}$ states of the form $e_{i_1} \otimes \ldots \otimes e_{i_n}$ (where we choose all CK sequences of the labels). Denote this basis by CK(B). Hence $\Xi(B)$ is precisely the subspace obtained by applying $U^{\otimes n}$ to $\Xi(B^0)$ (where $U^{\otimes n}$ is the unitary transformation on $\mathcal{H}^{\otimes n}$ given by $U \otimes \ldots \otimes U$). Then Υ is the span of all $\Xi(B)$ as B ranges over all bases, which in turn equals the span of all $U^{\otimes n}\phi$ where U ranges over all $d \times d$ unitary matrices and ϕ ranges over CK(B^0). Let M_d denote the linear space of all $d \times d$ complex matrices. Since M_d contains all unitary matrices we get

$$\Upsilon \subseteq \mathrm{span}\{A^{\otimes n}\phi : A \in M_d,\ \phi \in \mathrm{CK}(B^0)\}. \qquad (4)$$

For any fixed ϕ let

$$\mathcal{H}_\phi = \mathrm{span}\{A^{\otimes n}\phi : A \in M_d\}. \qquad (5)$$

We will show that

$$\dim \mathcal{H}_\phi \leq (n + 1)^{d^2}. \qquad (6)$$

Then using (4) and the fact that $\dim \Xi(B^0) = 2^{n(S+\delta)}$ we will immediately obtain our desired result (3).

To prove (6) we use the notion of the symmetric subspace.

Definition.—The symmetric subspace of a space $\mathcal{H}^{\otimes n}$ is the space SYM($\mathcal{H}$) of the vectors which are invariant under any permutation of the positions in the tensor product.

The symmetric subspace has found various applications in quantum information theory [11,12]. In [12] it is proved that the space SYM($\mathcal{H}$) has the following properties: (i) It is spanned by the vectors of the form $\psi^{\otimes n}$; (ii) its dimension is equal to $\binom{n+d-1}{d-1}$ where $d = \dim \mathcal{H}$. In fact, by considering the symmetrization of a product basis of $\mathcal{H}^{\otimes n}$ it is easy to obtain the simpler overestimate $\dim \mathrm{SYM}(\mathcal{H}) \leq (n + 1)^d$ which will suffice for our purposes.

An important point to note is that for fixed d and varying n the size of SYM($\mathcal{H}$) grows only polynomially with n, whereas the full space $\mathcal{H}^{\otimes n}$ (of dimension d^n) grows exponentially. Thus SYM($\mathcal{H}$) becomes exponentially small inside $\mathcal{H}^{\otimes n}$ as n grows.

Since M_d is a linear space we can consider $M_d^{\otimes n}$ and the symmetric subspace $\mathrm{SYM}(M_d) \subseteq M_d^{\otimes n}$. According to (i)

$$\mathrm{SYM}(M_d) = \mathrm{span}\{A^{\otimes n} : A \in M_d\},$$

and hence (5) gives

$$\mathcal{H}_\phi = \mathrm{span}\{B\phi : B \in \mathrm{SYM}(M_d)\}.$$

Now, we can define a linear mapping Γ from the space SYM(M_d) to $\mathcal{H}_\phi$ by

$$\mathrm{SYM}(M_d) \ni B \rightarrow \Gamma(B) = B\phi \in \mathcal{H}_\phi . \quad (7)$$

This mapping is onto the space $\mathcal{H}_\phi$, and since it is linear it cannot increase dimension. Hence

$$\dim \mathcal{H}_\phi \leq \dim \mathrm{SYM}(M_d) .$$

Recalling that $\dim M_d = d^2$, (ii) gives that

$$\dim \mathrm{SYM}(M_d) = \binom{n + d^2 - 1}{d^2 - 1} \leq (n + 1)^{d^2},$$

which proves (6) and completes the proof of (3).

Thus we have shown that for any given S and sufficiently large n, projection into $\Upsilon(S, n)$ will provide universal quantum data compression to S qubits/signal for all sources of pure quantum states with von Neumann entropy $\leq S$. The same method will also work faithfully for all sources of *mixed* states ρ_i where the von Neumann entropy of $\rho = \sum_i p_i \rho_i$ does not exceed S. Indeed, we may always represent each of these mixed states as a probabilistic mixture of pure states whose identities we have forgotten. Also according to the results of Barnum *et al.* [5] our compression scheme is optimal—compression beyond S qubits/signal cannot be faithful for sources of entropy equal to S and hence cannot be faithful for all sources of entropy $\leq S$.

Finally we remark that our bound (3) on $\dim \Upsilon$, although sufficient for our purposes, is not generally tight. Indeed, all we needed to show was that $\dim \Upsilon$ is some *polynomial* (in n) multiple of $\dim \Xi(B^0)$. It is interesting to note that $\dim \Upsilon$ can be calculated exactly for the case of $S = 0$. Here we are considering all possible trivial sources $\Sigma(\psi)$ which generate repeatedly one and the same vector ψ (i.e., have von Neumann entropy zero). For $\Sigma(\psi)$ the subspace Ξ and the typical subspace are both just the one dimensional $\mathrm{span}\{\psi \otimes \ldots \otimes \psi\} \in \mathcal{H}^{\otimes n}$. Hence Υ is the span of all states of the form $\psi \otimes \ldots \otimes \psi$, and by (i) we see that Υ equals $\mathrm{SYM}(\mathcal{H})$ in this case. As noted previously, $\mathrm{SYM}(\mathcal{H})$ becomes vanishingly small inside $\mathcal{H}^{\otimes n}$ as n increases so the number of qubits per signal used for faithful transmission tends to zero with increasing n.

M. H. and P. H. gratefully acknowledge the support from the Foundation for Polish Science. R. J. is supported in part by the European TMR network ERB-FMRX-CT96-0087.

*Email address: rjozsa@plymouth.ac.uk
†Email address: michalh@iftia.univ.gda.pl
‡Email address: pawel@mifgate.mif.pg.gda.pl
§Email address: fizrh@univ.gda.pl

[1] E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, New York, 1991).
[3] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[4] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[5] H. Barnum, Ch. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. A **54**, 4707 (1996).
[6] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Academic Press, New York, 1981).
[7] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).
[8] C. Krattenthaler and P. B. Slater, in http://xxx.lanl.gov/quant-ph/9612043, have considered a different approach to the issue of universal quantum coding and obtained results for a parametrized family of distributions of two-level quantum systems.
[9] M. Horodecki, R. Horodecki, and P. Horodecki, Acta Phys. Slovaca **48**, 133 (1998). Also available at quant-ph/9803080.
[10] E. Jaynes, Phys. Rev. **108**, 171 (1957); **108**, 620 (1957); Am. J. Phys. **31**, 66 (1963).
[11] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995); R. Derka, V. Bužek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998); R. Werner, Report No. quant-ph/9804001.
[12] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello, SIAM J. Comput. **26**, 1541 (1997). Also available at quant-ph/9604028.