Source Coding in Quantum Information Theory

Nilanjana Datta
Statistical Laboratory
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road, Cambridge CB30WB
email: n.datta@statslab.cam.ac.uk

Tony C. Dorlas
Dublin Institute for Advanced Studies
School of Theoretical Physics
10 Burlington Road, Dublin 4, Ireland.
email: dorlas@stp.dias.ie

Keywords: information source; data compression; Shannon entropy; von Neumann entropy; memoryless sources; ergodic sources; Shannon-McMillan-Breimann theorem; typical sequences; typical subspaces; fidelity; Schumacher compression; universal data compression; quantum lattice systems; variable-length coding; Lempel-Ziv algorithm.

1 Introduction

Two key issues of classical and quantum information theory are storage and transmission of information. An information source produces some outputs (or signals) more frequently than others. Due to this redundancy one can reduce the amount of space needed for its storage without compromising on its content. This data compression is done by a suitable encoding of the output of the source. In contrast, in the transmission of information through a channel, it is often advantageous to add redundancy to a message, in order to combat the effects of noise. This is done in the form of error-correcting codes. The amount of redundancy which needs to be added to the original message depends on how much noise is present in the channel (see e.g. [18, 36, 30]). Hence redundancy plays complementary roles in data compression and transmission of data through a noisy channel. In this review we focus only on data compression in Quantum Information Theory.

In Classical Information Theory, Shannon showed that there is a natural limit to the amount of compression that can be achieved. It is given by the Shannon entropy. The analogous concept in Quantum Information Theory is the von Neumann entropy. Here we review some of the main results of quantum data compression and the significance of the von Neumann entropy in this context.

The review is structured on follows. We first give a brief introduction to the Shannon entropy and classical data compression. This is followed by a discussion of quantum entropy and the idea behind quantum source coding. We elaborate on data compression schemes for three different classes of quantum sources, namely memoryless (or i.i.d.) sources, ergodic sources and sources modelled by Gibbs states of quantum spin systems. In the bulk of the review, we concentrate on source-dependent, fixed-length coding schemes. We conclude with a brief discussion of universal and variable-length coding.

In this short review, we have only concentrated on certain aspects of data compression in Quantum Information Theory. For other important developments see e.g. [1, 16, 29, 36, 37, 8, 21, 2].

2 Classical Data Compression

2.1 Entropy and source coding

A simple model of a classical information source consists of a sequence of discrete random variables $X_1, X_2, \ldots, X_n$, whose values represent the output of the source. Each random variable X_i, $1 \leq i \leq n$ takes values x_i from a finite set, the *source alphabet* $\mathcal{X}$. Hence $\underline{X}^{(n)} := (X_1, \ldots, X_n)$ takes values $\underline{x}^{(n)} := (x_1, \ldots, x_n) \in \mathcal{X}^n$. The sequences $\underline{x}^{(n)}$ constitute the output or signal of the source. We recall the definition of entropy (or information content) of a source:

If the discrete random variables $X_1, \ldots, X_n$ which take values from a finite alphabet $\mathcal{X}$ have joint probabilities

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = p_n(x_1, \ldots, x_n)$$

then the **Shannon entropy** of this source is defined by

$$H(X_1, \ldots, X_n) = -\sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_n \in \mathcal{X}_n} p_n(x_1, \ldots, x_n) \log p_n(x_1, \ldots, x_n). \tag{2.1}$$

Here and in the following, the logarithm is taken to the base 2. This is because the fundamental unit of classical information is a **bit**, which takes two values, 0 and 1. Notice that $H(X_1, \ldots, X_n)$ in fact only depends on the (joint) probability mass function (p.m.f.) p_n and can also be denoted as $H(p_n)$.

There are several other concepts of entropy, e.g. relative entropy, conditional entropy and mutual information. See for example [10, 30].

Two important properties of $H(X)$ are as follows:

$$(a)\, 0 \leq H(X) \leq \log |\mathcal{X}| \quad \text{and} \quad (b)\, H(X) \text{ is concave in } X. \tag{2.2}$$

In the next section, analogous quantities are introduced for quantum information and the corresponding properties are stated.

Suppose that the random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.). Then the entropy of each random variable modelling the source is the same and can be denoted by $H(X)$. From the point of view of Classical Information Theory, the Shannon entropy has an important operational definition. It quantifies the minimal physical resources needed to store data from a classical information source and provides a limit to which data can be compressed *reliably* (i.e., in a manner in which the original data can be recovered later with a low probability of error). Shannon showed that the original data can be reliably obtained from the compressed version only if the rate of compression is greater than the Shannon entropy. This result is formulated in Shannon's Noiseless Channel Coding Theorem [35, 10, 30] given in Section 2.3.

2.2 The Asymptotic Equipartition Property

The main idea behind Shannon's Noiseless Channel Coding Theorem is to divide the possible values $x_1, x_2, \ldots \ldots x_n$ of random variables $X_1, \ldots, X_n$ into two classes – one consisting of sequences which have a high probability of occurrence, known as *typical sequences* and the other consisting of sequences which occur rarely, known as *atypical sequences*. The idea is that there are far fewer typical sequences than the total number of possible sequences, but they occur with high probability. The existence of typical sequences follows from the so-called **Asymptotic Equipartition Property**:

Theorem 2.1 (AEP) *If $X_1, X_2, X_3, \ldots$ are i.i.d. random variables with p.m.f $p(x)$, then*

$$-\frac{1}{n} \log p_n(X_1, \ldots, X_n) \xrightarrow{\mathbb{P}} H(X), \tag{2.3}$$

where $H(X)$ is the Shannon entropy for a single variable, and $p_n(X_1, \ldots, X_n)$ denotes the random variable taking values $p_n(x_1, \ldots, x_n) = \prod_{i=1}^n p(x_i)$ with probabilities $p_n(x_1, \ldots, x_n)$.

This theorem has been generalised to the case of sequences of dependent variables $(X_n)_{n\in\mathbb{Z}}$ which are ergodic for the shift transformation defined below. It is easiest to formulate this for an information stream which extends from $-\infty$ to $+\infty$:

Definition A sequence $(X_n)_{n\in\mathbb{Z}}$ is called **stationary** if for any $n_1 < n_2$ and any $x_{n_1}, \ldots, x_{n_2} \in \mathcal{X}$,

$$\mathbb{P}(X_{n_1} = x_{n_1}, \ldots, X_{n_2} = x_{n_2}) = \mathbb{P}(X_{n_1+1} = x_{n_1}, \ldots, X_{n_2+1} = x_{n_2}).$$

We define the shift transformation τ by

$$\tau\left((x_n)_{n\in\mathbb{Z}}\right) = (x'_n)_{n\in\mathbb{Z}}, \text{ where } x'_n = x_{n-1}. \tag{2.4}$$

Then $(X_n)_{n\in\mathbb{Z}}$ is called **ergodic** if it is stationary and if every subset $A \subset \mathcal{X}^{\mathbb{Z}}$ such that $\tau(A) = A$, has probability 0 or 1: $\mathbb{P}((X_n)_{n\in\mathbb{Z}} \in A) = 0$ or 1.

It is known that $(X_n)_{n\in\mathbb{Z}}$ is ergodic if and only if its probability distribution is extremal in the set of invariant probability measures.The generalisation of Theorem 2.1 [28, 9] now reads:

Theorem 2.2 (Shannon-McMillan-Breiman Theorem) *Suppose that the sequence $(X_n)_{n\in\mathbb{Z}}$ is ergodic. Then*

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\log p_n(X_1,\ldots,X_n)\right\} = h_{KS} \text{ with prob. } 1, \tag{2.5}$$

*where h_{KS} is the **Kolmogorov-Sinai entropy** defined by*

$$h_{KS} = \lim_{n\to\infty}\frac{1}{n}H(X_1,\ldots,X_n) = \inf_n \frac{1}{n}H(X_1,\ldots,X_n). \tag{2.6}$$

Remark. One can prove, using convexity, that the sequence $\frac{1}{n}H(p_n)$ is decreasing, and bounded below.

We now define the *set of typical sequences* as follows:

Definition Let $X_1,\ldots,X_n$ be i.i.d. random variables with p.m.f. $p(x)$. Given $\epsilon > 0$, the **typical set** $T_\epsilon^{(n)}$ is the set of sequences $(x_1\ldots x_n)$ for which

$$2^{-n(H(X)+\epsilon)} \le p(x_1\ldots x_n) \le 2^{-n(H(X)-\epsilon)}. \tag{2.7}$$

In the case of an ergodic sequence, $H(X)$ is replaced by h_{KS} in (2.7).

Let $|T_\epsilon^{(n)}|$ denote the total number of typical sequences and $\mathbb{P}\{T_\epsilon^{(n)}\}$ denotes the probability of the *typical set*. Then the following is an easy consequence of Theorem 2.1.

Theorem 2.3 (Theorem of Typical Sequences) *For any $\delta > 0$ $\exists$ $n_0(\delta) > 0$ such that $\forall n \ge n_0(\delta)$ the following hold:*
(a) $\mathbb{P}\{T_\epsilon^{(n)}\} > 1-\delta$, and (b) $(1-\delta)\,2^{n(H(X)-\epsilon)} \le |T_\epsilon^{(n)}| \le 2^{n(H(X)+\epsilon)}$.

2.3 Shannon's Noiseless Channel Coding Theorem

Shannon's Noiseless Channel Coding Theorem is a simple application of the Theorem of Typical Sequences and says that the optimal rate at which one can *reliably* compress data from an i.i.d. classical information source is given by the Shannon entropy $H(X)$ of the source.

A **compression scheme** C^n of rate R maps possible sequences $\underline{x} = (x_1,\ldots,x_n)$ to a binary string of length $\lceil nR\rceil$: $C^n : \underline{x} \mapsto \underline{y} = (y_1,\ldots,y_{\lceil nR\rceil})$, where $x_i \in \mathcal{X}$; $|\mathcal{X}| = d$ and $x_i \in \{0,1\}$ $\forall$ $1 \le i \le \lceil nR\rceil$. The corresponding *decompression scheme* takes the $\lceil nR\rceil$ compressed bits and maps them back to a string of n letters from the alphabet $\mathcal{X}$: $D^n : \underline{y} \in \{0,1\}^{\lceil nR\rceil} \mapsto \underline{x'} = (x'_1,\ldots,x'_n)$. A compression–decompression scheme is said to be **reliable** if the probability that $\underline{x'} \neq \underline{x}$ tends to 0 as $n \to \infty$. Shannon's Noiseless Channel Coding Theorem [35, 10] now states

Theorem 2.4 (Shannon) *Suppose that $\{X_i\}$ is an i.i.d. information source, with $X_i \sim p(x)$ and Shannon entropy $H(X)$. If $R > H(X)$ then there exists a reliable compression scheme of rate R for the source. Conversely, any compression scheme with rate $R < H(X)$ is not be reliable.*

Proof (sketch) Suppose $R > H(X)$. Choose $\epsilon > 0$ such that $H(X) + \epsilon < R$. Consider the set $T_\epsilon^{(n)}$ of typical sequences. The method of compression is then to examine the output of the source, to see if it belongs to $T_\epsilon^{(n)}$. If the output is a typical sequence, then we compress the data by simply storing an index for the particular sequence using $\lceil nR \rceil$ bits in the obvious way. If the input string is *not* typical, then we compress the string to some fixed $\lceil nR \rceil$ bit string e.g. $(00\ldots000)$. In this case, data compression effectively fails, but, in spite of this, the compression–decompression scheme succeeds with probability *one* as $n \to \infty$ by Theorem 2.3.

If $R < H(X)$ then any compression scheme of rate R is not reliable. This also follows from Theorem 2.3 by the following argument. Let $\mathcal{S}(n)$ be a collection of sequences $\underline{x}^{(n)}$ of size $|\mathcal{S}(n)| \leq 2^{\lceil nR \rceil}$. Then the subset of atypical sequences is highly improbable, whereas the subset of typical sequences has probability bounded by $2^{nR}2^{-nH(X)} \to 0$. ∎

3 Quantum Data Compression

3.1 Quantum Sources and Entropy

A quantum information source is defined by a sequence of density matrices ρ_n, acting on finite-dimensional Hilbert spaces $\mathcal{H}_n$ representing the possible states of the source, n labelling the size of the emitted signal (i.e., the message).

A **density matrix** ρ is a positive definite operator on a Hilbert space $\mathcal{H}$ with $\mathrm{Tr}(\rho) = 1$, and the expected value of an operator A on $\mathcal{H}$ is given by

$$\phi(A) = \mathrm{Tr}\,(\rho\, A). \tag{3.1}$$

The functional ϕ on $\mathcal{M} = \mathcal{B}(\mathcal{H})$, the algebra of linear operators on $\mathcal{H}$, is positive (i.e. $\phi(A) \geq 0$, if $A \geq 0$) and maps the identity $\mathbf{1} \in \mathcal{M}$ to 1. Such a functional is, confusingly, also called a **state**. Conversely, given such a state on a finite-dimensional algebra $\mathcal{M}$, there exists a unique density operator ρ_ϕ such that (3.1) holds, so the concepts can be used interchangeably. (This is not true in the infinite-dimensional case.)

Diagonalising ρ_n,

$$\rho_n = \sum_{k=1}^{N_n} \lambda_k^{(n)} |\psi_k^{(n)}\rangle\langle\psi_k^{(n)}|, \tag{3.2}$$

one can interpret $\lambda_k^{(n)}$ as the probability with which the eigenstate $|\psi_k^{(n)}\rangle$ occurs. In (3.2) $N_n = \dim\, \mathcal{H}_n$. One can usually consider $\mathcal{H}_n$ to be an n-fold tensor product of the same Hilbert space $\mathcal{H}$. The case that ρ_n is also a tensor product corresponds to an independent source.

In Classical Information Theory, the optimal rate of data compression is given by the Shannon entropy of the source. The analogous quantity in Quantum Information Theory is the **von Neumann entropy**

$S(\phi)$ also written as $S(\rho_\phi)$. It is defined by

$$S(\phi) = -\mathrm{Tr}\big(\rho_\phi \log \rho_\phi\big) = -\sum_{k=1}^{N} \lambda_k \log \lambda_k, \tag{3.3}$$

(The logarithm is taken to base 2 as before.) It has properties analogous to $H(X)$, in particular [31, 30].

Proposition 3.1 *If $\mathcal{H}$ is a finite-dimensional Hilbert space, then for states ϕ on $\mathcal{B}(\mathcal{H})$,*

(a) $0 \leq S(\phi) \leq \log(\dim(\mathcal{H}))$ *and* *(b)* $S(\phi)$ *is concave in* ϕ.

3.2 Compression schemes and fidelity

Consider a quantum information source $\{\rho_n, \mathcal{H}_n\}$, where ρ_n has spectral decomposition given by (3.2). To compress data from such a source one encodes each eigenstate $|\psi_k^{(n)}\rangle$, by a state $\tilde{\rho}_k^{(n)} \in \mathcal{B}(\tilde{\mathcal{H}}_n)$ where $\dim \tilde{\mathcal{H}}_n = d_c(n) < N_n$. Thus, a compression scheme is now a map $\mathcal{C}^n : |\psi_k^{(n)}\rangle \mapsto \tilde{\rho}_k^{(n)} \in \mathcal{B}(\tilde{\mathcal{H}}_n)$. Obviously, the goal is to keep the dimension d_c as small as possible. As in the classical case, i.e. Theorem 2.4, we are interested in finding the optimal *limiting rate of data compression*, which in this case is given by

$$R_\infty := \lim_{n\to\infty} \frac{\log d_c}{n}. \tag{3.5}$$

The reliability or *fidelity* of a compression scheme can be measured in various ways. We use the following definition of fidelity:

$$F_n = \sum_k \lambda_k^{(n)} \langle\psi_k^{(n)}| \, [\tilde{\rho}_k^{(n)}]^{1/2} |\psi_k^{(n)}\rangle. \tag{3.6}$$

This fidelity satisfies $0 \leq F_n \leq 1$ and $F_n = 1$ if $\tilde{\rho}_k^{(n)} = |\psi_k^{(n)}\rangle\langle\psi_k^{(n)}|$ for all k. A compression–decompression scheme is said to be *reliable* if $F_n \to 1$ as $n \to \infty$.

Analogous to the classical case, reliable coding is achieved by looking for a typical subspace $\mathcal{T}_\epsilon^{(n)}$ of the Hilbert space $\mathcal{H}_n$ for a given $\epsilon > 0$. If $P_\epsilon^{(n)}$ is the projection onto $\mathcal{T}_\epsilon^{(n)}$, we put

$$|\tilde{\psi}_k^{(n)}\rangle := \frac{P_\epsilon^{(n)}|\psi_k^{(n)}\rangle}{||P_\epsilon^{(n)}|\psi_k^{(n)}\rangle||}, \; ; \; \alpha_k := ||P_\epsilon^{(n)}|\psi_k^{(n)}\rangle|| \; ; \; \beta_k = ||(\mathbf{1} - P_\epsilon^{(n)})|\psi_k^{(n)}\rangle||, \tag{3.7}$$

and set

$$\tilde{\rho}_k^{(n)} := \alpha_k^2 |\tilde{\psi}_k^{(n)}\rangle\langle\tilde{\psi}_k^{(n)}| + \beta_k^2 |\psi_0\rangle\langle\psi_0|, \tag{3.8}$$

where $|\psi_0\rangle$ is any fixed state in $\left(\mathcal{T}_\epsilon^{(n)}\right)^\perp$. The fidelity of this compression scheme satisfies

$$\begin{aligned} F_n &\geq \sum_k \lambda_k^{(n)} \langle\psi_k^{(n)}|\, \tilde{\rho}_k^{(n)} |\psi_k^{(n)}\rangle \\ &= \sum_k \lambda_k^{(n)} \left[\alpha_k^2 |\langle\psi_k^{(n)}|\tilde{\psi}_k^{(n)}\rangle|^2 + \beta_k^2 |\langle\psi_k^{(n)}|\psi_0\rangle|^2\right] \\ &\geq \sum_k \lambda_k^{(n)} \alpha_k^2 |\langle\psi_k^{(n)}|\tilde{\psi}_k^{(n)}\rangle|^2 = \sum_k \lambda_k^{(n)} \alpha_k^4 \\ &\geq \sum_k \lambda_k^{(n)} (2\alpha_k^2 - 1) := 2A_n - 1. \end{aligned} \tag{3.9}$$

In the following subsections we discuss different classes of quantum sources for which one can find typical subspaces $\mathcal{T}_\epsilon^{(n)}$ such that the quantity A_n (and hence the fidelity F_n) tends to 1 as $n \to \infty$.

3.3 Schumacher's theorem for memoryless quantum sources

The notion of a typical subspace was first introduced in the context of quantum information theory by Schumacher [33]. He considered the simplest class of quantum sources namely *memoryless* or i.i.d sources for which the density matrix ρ_n is a tensor product $\rho_n = \pi^{\otimes n}$. It is obvious that the von Neumann entropy of such a source is given by

$$S(\rho_n) \equiv S(\pi^{\otimes n}) = nS(\pi) \tag{3.10}$$

In this case the eigenvalues $\lambda_k^{(n)}$ in (3.2), can clearly be written as

$$\lambda_{k_1,\ldots,k_n}^{(n)} = \prod_{i=1}^n \lambda_{k_i}, \tag{3.11}$$

where λ_k are the eigenvalues of π. Since $\sum_k \lambda_k = 1$ we can consider the probability distribution defined by these eigenvalues and define the corresponding classical typical subset $T_\epsilon^{(n)}$ of indices $(k_1, \ldots, k_n)$ as in Theorem 2.3. Defining $\mathcal{T}_\epsilon^{(n)}$ as the space spanned by the eigenvectors $|\psi_{k_1,\ldots,k_n}^{(n)}\rangle$ with $(k_1, \ldots, k_n) \in T_\epsilon^{(n)}$ we obtain immediately the quantum analogue of the Theorem of Typical Sequences:

Theorem 3.2 (Typical Subspace Theorem) *Fix $\epsilon > 0$. Then for any $\delta > 0$ $\exists$ $n_0(\delta) > 0$ such that $\forall n \geq n_0(\delta)$ and $\rho_n = \pi^{\otimes n}$, the following are true:*
(a) $\mathrm{Tr}\left(P_\epsilon^{(n)} \rho_n\right) > 1 - \delta$, *and* *(b)* $(1-\delta)\, 2^{n(S(\pi)-\epsilon)} \leq \dim(\mathcal{T}_\epsilon^{(n)}) \leq 2^{n(S(\pi)+\epsilon)}$,
where $P_\epsilon^{(n)}$ is the orthogonal projection onto the subspace $\mathcal{T}_\epsilon^{(n)}$.

(Notice that $S(\pi) = H(X)$, where X is a random variable with probability distribution given by $\{\lambda_k\}$.) Moreover, $\mathrm{Tr} P_\epsilon^{(n)} \rho_n = \mathbb{P}\left[(k_1, \ldots, k_n) \in \mathcal{T}_\epsilon^{(n)}\right]$.)

Using the above theorem, Schumacher [33, 24] proved the following analogue of Shannon's Noiseless Channel Coding Theorem:

Theorem 3.3 (Schumacher's Quantum Coding Theorem) *Let $\{\rho_n, \mathcal{H}_n\}$ be an i.i.d. quantum source: $\rho_n = \pi^{\otimes n}$ and $\mathcal{H}_n = \mathcal{H}^{\otimes n}$. If $R > S(\pi)$ then there exists a reliable compression scheme of rate R. If $R < S(\pi)$ then any compression scheme of rate R is not reliable.*

Proof (i) $R > S(\pi)$: Choose $\epsilon > 0$ such that $R > S(\pi) + \epsilon$.

For a given $\delta > 0$, choose the typical subspace as above and choose n large enough so that (a) and (b) in the Typical Subspace Theorem hold. Then note that

$$\alpha_k^2 = \langle \psi^{(n)}_{k_1,\ldots,k_n} |\, P_\epsilon^{(n)} | \psi^{(n)}_{k_1,\ldots,k_n} \rangle \tag{3.12}$$

and hence $A_n = \mathrm{Tr}\,(P_\epsilon^{(n)} \rho_n) > 1 - \delta$.

(ii) Analogous to the proof of Shannon's theorem, if $\rho_n \mapsto \tilde{\rho}_n$ is a compression scheme with rate $R < S(\pi)$ then, there are subspaces $\mathcal{K}_n = \mathrm{supp}(\tilde{\rho}_n)$ of dimension $\leq 2^{\lceil nR \rceil}$ (assuming for simplicity that $\mathcal{H}$ has dimension 2), on which $\tilde{\rho}_n$ is concentrated. Considering the projections onto $\mathcal{T}_\epsilon^{(n)}$ and its complement, we then find that the fidelity which tends to 0 as $n \to \infty$. ∎

3.4 Ergodic Quantum Sources

A quantum generalisation of classical ergodic sources is defined as follows. First consider the analogue of an infinite sequence of random variables which is a state on the infinite tensor product of a finite-dimensional $*$-algebra $\mathcal{M}$. The latter is given by the norm-closure of the increasing sequence of finite tensor products

$$\mathcal{M}_\infty = \overline{\bigcup_n \otimes_{k=-n}^{n} \mathcal{M}} \tag{3.13}$$

A translation-invariant state ϕ_∞ on $\mathcal{M}_\infty$ is said to be **ergodic** if it cannot be decomposed as a (non-trivial) convex combination of other translation-invariant states. The analogue of the Kolmogorov-Sinai entropy (2.6) for an ergodic state ϕ_∞ is called the **mean entropy** and is given by

$$S_M(\phi_\infty) = \lim_{n\to\infty} \frac{1}{n} S(\phi_n), \tag{3.14}$$

where ϕ_n is the restriction of ϕ_∞ to $\mathcal{M}_n := \mathcal{M}^{\otimes n}$.

Following Hiai and Petz [17], we define the following quantity for any state ϕ on an arbitrary finite-dimensional $*$-algebra $\mathcal{M}$ and a given $\delta > 0$:

$$\beta_\delta(\phi) = \inf \left\{ \log \mathrm{Tr}(q) :\ q \in \mathcal{M}, q^* = q, q^2 = q, \phi(q) \geq 1 - \delta \right\}. \tag{3.15}$$

We also define a state ϕ_∞ on $\mathcal{M}_\infty$ to be **completely ergodic** if it is ergodic under transformations on $\mathcal{M}_\infty$, induced by l-fold shifts on $\mathbb{Z}$, for arbitrary $l \in \mathbb{N}$. The following theorem is due to Hiai and Petz [17], who proved it in a slightly more general setting:

Theorem 3.4 (Hiai & Petz) *Suppose that ϕ_∞ is a completely ergodic state on $\mathcal{M}_\infty$ and $d := \dim \mathcal{M} < \infty$, and set $\phi_n = \phi_\infty \restriction_{\mathcal{M}_n}$. Then, for any $\delta > 0$, the following hold:*

$$(\mathbf{1}) \qquad \limsup_{n\to\infty} \frac{1}{n} \beta_\delta(\phi_n) \leq S_M(\phi_\infty), \tag{3.16}$$

and $$(\mathbf{2}) \qquad \liminf_{n\to\infty} \frac{1}{n} \beta_\delta(\phi_n) \geq S_M(\phi_\infty) - \delta \log d. \tag{3.17}$$

Proof of (1): Choose $r > S_M(\phi_\infty)$ and let $\epsilon < r - S_M(\phi_\infty)$ and $h = r - \epsilon$. By the definition of $S_M(\phi_\infty)$, there exists $l \in \mathbb{N}$ such that $S(\phi_l) < l\,h$. Let $\{|e_i\rangle\}_{i=1}^{l\,d}$ be an orthonormal set of eigenvectors of ρ_{ϕ_l}, with corresponding eigenvalues λ_i, i.e. let

$$\rho_{\phi_l} = \sum_{i=1}^{l\,d} \lambda_i p_i, \tag{3.18}$$

where $p_i = |e_i\rangle\langle e_i|$ is the projection onto $|e_i\rangle$, be the spectral decomposition for ρ_{ϕ_l}. Denote the spectrum $\mathcal{X} = \{\lambda_i\}_{i=1}^{l\,d}$. For $n \in \mathbb{N}$, introduce the probability measures ν_n on $\mathcal{X}^n$ by

$$\nu_n(A) = \phi_{nl}(q_A), \tag{3.19}$$

where, for any $A \subset \mathcal{X}^n$, the projection q_A is defined by

$$q_A = \sum_{(\lambda_{i_1},\ldots,\lambda_{i_n})\in A} p_{i_1} \otimes \ldots \otimes p_{i_n}. \tag{3.20}$$

Similarly, we define ν_∞ on $\mathcal{X}^{\mathbb{Z}}$. The sequence of random variables $(X_n)_{n\in\mathbb{Z}}$ with distribution ν_∞ is then ergodic since ϕ_∞ is completely ergodic (and hence l-ergodic).

By the Shannon-McMillan-Breiman theorem 2.2,

$$-\frac{1}{n}\log \nu_n(\{(x_1,\ldots,x_n)\} \to h_{KS} \tag{3.21}$$

almost surely w.r.t. ν_∞, where h_{KS} is the Kolmogorov-Sinai entropy. The latter is given by $h_{KS} = \lim_{n\to\infty} \frac{1}{n} H_n = \inf_{n\in\mathbb{N}} \frac{1}{n} H_n$, where

$$H_n = -\sum_{(x_1,\ldots,x_n)\in\mathcal{X}^n} \nu_n(\{(x_1,\ldots,x_n)\}) \log \nu_n(\{(x_1,\ldots,x_n)\}). \tag{3.22}$$

Notice in particular that

$$h_{KS} \le H_1 = S(\phi_l) < l\,h. \tag{3.23}$$

If we let $T_\epsilon^{(n)}$ be the (typical) subset of $\mathcal{X}^n$ such that

$$-\frac{1}{n}\log \nu_n(\{(x_1,\ldots,x_n)\}) \in (h_{KS}-\epsilon, h_{KS}+\epsilon), \tag{3.24}$$

for $(x_1,\ldots,x_n) \in T_\epsilon^{(n)}$ then we have $\nu_\infty(T_\epsilon^{(n)}) \ge 1-\delta$ for n large enough. Moreover, since $\nu_n(\{(x_1,\ldots,x_n)\}) \ge e^{-n(h_{KS}+\epsilon)}$ for all $(x_1,\ldots,x_n) \in T_\epsilon^{(n)}$, and the total measure is 1,

$$|T_\epsilon^{(n)}| \le e^{n(h_{KS}+\epsilon)} \le e^{n(l\,h+\epsilon)}. \tag{3.25}$$

It follows that $\mathrm{Tr}(q_{T_\epsilon^{(n)}}) \le e^{n(l\,h+\epsilon)}$ whereas $\phi_{nl}(q_{T_\epsilon^{(n)}}) = \nu_n(T_\epsilon^{(n)}) \ge 1-\delta$ and we conclude that

$$\frac{1}{nl}\beta_\delta(\phi_{nl}) \le \frac{n(l\,h+\epsilon)}{nl} < r \tag{3.26}$$

from which (3.16) follows upon taking $n\to\infty$, since $r > S_M(\phi_\infty)$ was arbitrary. (Notice that $\beta_\delta(\phi_n)$ is decreasing in n since $\mathcal{M}_n \subset \mathcal{M}_{n+1}$.)

Proof of (2): Given $\epsilon, \delta > 0$ and $n \in \mathbb{N}$, choose a projection q_n with $\phi_n(q_n) \ge 1-\delta$ and $\log \mathrm{Tr}(q_n) < \beta_\delta(\phi_n) + \epsilon$. Since $S_M(\phi_\infty) = \inf \frac{1}{n} S(\phi_n)$ we have $S_M(\phi_\infty) \le \frac{1}{n} S(\phi_n)$. We now use the following simple lemma:

Lemma 3.5 *If ϕ is a state on a finite-dimensional $*$-algebra $\mathcal{M}$, and $q \in \mathcal{M}$ is a projection, then*

$$S(\phi) \leq H(p) + \phi(q) \log \mathrm{Tr}(q) + (1 - \phi(q)) \log \mathrm{Tr}(1 - q), \tag{3.27}$$

where $H(p) = -p \log p - (1-p) log(1-p)$ (the binary entropy) with $p = \phi(q)$.

Proof . First notice that if $[\rho_\phi, q] = 0$ then the result (3.27) follows from the simple inequality

$$-\sum_{i=1}^{m} \tilde{\lambda}_i \log \tilde{\lambda}_i \leq \log m \quad \text{if} \quad \sum_{i=1}^{m} \tilde{\lambda}_i = 1. \tag{3.28}$$

Indeed, diagonalising ρ_ϕ, the eigenvalues λ_i divide into two subsets with corresponding eigenvectors belonging to the range of q respectively its complement. Considering the first set, we have, if $m = \dim(\mathrm{Ran}(q))$, and taking $\tilde{\lambda}_i = \lambda_i / \left(\sum_{i=1}^{m} \lambda_i\right)$ in (3.28),

$$\begin{aligned} -\sum_{i=1}^{m} \lambda_i \log \lambda_i &\leq -\left(\sum_{i=1}^{m} \lambda_i\right) \log \left(\frac{1}{m} \sum_{i=1}^{m} \lambda_i\right) \\ &= -\mathrm{Tr}(q\rho_\phi) \left[\log \mathrm{Tr}(q\rho_\phi) - \log \mathrm{Tr}(q) \right]. \end{aligned}$$

Adding the analogous inequality for the part of the spectrum corresponding to $1-q$, we obtain (3.27).

In the general case, i.e. if $[\rho_\phi, q] \neq 0$, define the unitary $u = 2q - 1$ and the state

$$\phi'(x) = \frac{1}{2}[\phi(x) + \phi(uxu)]. \tag{3.29}$$

Then $[\rho_{\phi'}, q] = 0$ and by concavity of $S(\phi)$ and the result for the previous case

$$H(X) + \phi(q)\mathrm{Tr}(q) + (1 - \phi(q))\mathrm{Tr}(1 - q)) \geq S(\phi') \geq S(\phi) \tag{3.30}$$

since $\phi'(q) = \phi(q)$. ∎

Continuing with the proof of **(2)**, we conclude that

$$\begin{aligned} S(\phi_n) &\leq H(p) + \phi_n(q_n) \log \mathrm{Tr}(q_n) + (1 - \phi(q_n)) \log \mathrm{Tr}(1 - q_n) \\ &\leq 1 + \beta_\delta(\phi_n) + \epsilon + \delta n \log d. \end{aligned}$$

Dividing by n and taking the limit we obtain (3.17). ∎

It follows from this theorem that we can define a typical subspace in the same way as in Schumacher's theorem. Indeed, given $\delta > 0$ and $\epsilon > 0$, we have that for n large enough, there exists a subspace $\mathcal{T}_\epsilon^{(n)}$ equal to the range of a projection q_n such that $\phi_n(q_n) > 1 - \delta$ and $e^{n(S_M(\phi_\infty) - \delta \log d - \epsilon)} < \dim(\mathcal{T}_\epsilon^{(n)}) = \mathrm{Tr}(q_n) < e^{n(S_M(\phi_\infty) + \epsilon)}$. The proof of the quantum analogue of the Shannon-McMillan Theorem is then the same as that of Schumacher's theorem [32, 5]:

Theorem 3.6 *Let ϕ_∞ be a completely ergodic stationary state on the infinite tensor product algebra $\mathcal{M}_\infty$. If $R > S_M(\phi_\infty)$ then there exists a reliable quantum code of rate R. Conversely, if $R < S_M(\phi_\infty)$ then any quantum compression scheme of rate R is not reliable.*

Remarks. Theorem 3.4 also holds for higher-dimensional information streams, with essentially the same proof. (The existence of the mean entropy is more complicated in that case.) The condition of complete ergodicity in this theorem is unnecessary. Indeed, Bjelakovic et al. [5] showed that the result remains valid (also in more than one dimensions) if the state ϕ_∞ of the source is simply ergodic. They achieved this by decomposing a general ergodic state into a finite number of l-ergodic states, and then applying the above strategy to each. Moreover, they also proved [6] an analogue of Breiman's almost-sure version of the Shannon-McMillan theorem [9]. It should also be mentioned that a weaker version of Theorem 3.4 was proved in 1998 by King and Lesniewski [26]. They considered the entropy of an associated classical source, but did not show that this classical entropy can be optimised to approximate the von Neumann entropy. This had in fact already been proved by Hiai & Petz [17] in 1991. The relevance of this work for quantum information theory was finally pointed out by Mosonyi and Petz [32].

3.5 Source coding for quantum spin systems

In this section we consider a class of quantum sources modelled by Gibbs states of a finite strongly interacting quantum spin system in $\Lambda \subset \mathbb{Z}^d$ with $d \geq 2$. Due to the interaction between spins, the density matrix of the source is not given by a tensor product of the density matrices of the individual spins and hence the quantum information source is non-i.i.d. We consider the density matrix to be written in the standard Gibbsian form:

$$\rho^{\omega,\Lambda} = \frac{e^{-\beta H_\Lambda^\omega}}{\Xi^{\omega,\Lambda}}, \tag{3.31}$$

where $\beta > 0$ is the inverse temperature. Here ω denotes the boundary condition, i.e., the configuration of the spins in $\Lambda = \mathbb{Z}^d \setminus \Lambda$, and H_Λ^ω is the Hamiltonian acting on the spin system in Λ under this boundary condition. (See [13] for precise definitions of these quantities). The denominator on the right-hand side of (3.31) is the partition function.

Note that any faithful density matrix can be written in the form (3.31) for some self–adjoint operator H_Λ^ω with discrete spectrum, such that $e^{-\beta H_\Lambda^\omega}$ is trace class. However, we consider H_Λ^ω to be a small quantum perturbation of a classical Hamiltonian and require it to satisfy certain hypotheses (see [13]). In particular, we assume that $H_\Lambda = H_{0\Lambda} + \lambda V_\Lambda$, where (i) $H_{0\Lambda}$ is a classical, finite-range, translation-invariant Hamiltonian with a finite number of periodic ground states, and the excitations of these ground states have an energy proportional to the size of their boundaries (Peierls condition; see e.g. [7, 12]); (ii) λV_Λ is a translation-invariant, exponentially decaying, quantum perturbation, λ being the perturbation parameter. These hypotheses ensure that the Quantum Pirogov Sinai theory of phase transitions in lattice systems (see [7], [12]) applies.

The power of Quantum Pirogov-Sinai theory is such that, in proving reliable data compression for such sources, we do not need to invoke the concept of ergodicity.

Using the concavity of the von Neumann entropy $S(\rho^{\omega,\Lambda})$ one can prove that the von Neumann entropy rate (or mean entropy) of the source

$$h := \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{S(\rho^{\omega,\Lambda})}{|\Lambda|}$$

exists. For a general van Hove sequence, this follows from the strong subadditivity [27] of the von Neumann entropy.

Let $\rho^{\omega,\Lambda}$ have a spectral decomposition

$$\rho^{\omega,\Lambda} = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|,$$

where the eigenvalues λ_j, $1 \leq j \leq 2^{|\Lambda|}$, and the corresponding eigenstates $|\psi_j\rangle$ depend on ω and Λ. Let $\mathcal{P}^{\omega,\Lambda}$ denote the probability distribution $\{\lambda_j\}$ and consider a random variable $K^{\omega,\Lambda}$ which takes a value λ_j with probability λ_j:

$$K^{\omega,\Lambda}(\psi_j) = \lambda_j \quad ; \quad \mathcal{P}^{\omega,\Lambda}(K^{\omega,\Lambda} = \lambda_j) = \lambda_j.$$

The data compression limit is related to asymptotical properties of the random variables $K^{\omega,\Lambda}$ as $\Lambda \nearrow \mathbb{Z}^d$. As in the case of i.i.d. sources, we prove the reliability of data compression by first proving the existence of a *typical subspace*. The latter follows from Theorem 3.7 below. The proof of this crucial theorem relies on results of Quantum Pirogov Sinai theory [12, 7].

Theorem 3.7 *Under the above assumptions, for β large and λ small enough, for all $\epsilon > 0$*

$$\lim_{\Lambda \nearrow \mathbb{Z}^d} \mathcal{P}^{\omega,\Lambda}\left(|\frac{-1}{|\Lambda|}\log K^{\omega,\Lambda} - h| \leq \epsilon\right) = \lim_{\Lambda \nearrow \mathbb{Z}^d} \sum_j \lambda_j \, \chi_{\{|-|\Lambda|^{-1}\log\lambda_j - h| \leq \epsilon\}} = 1, \tag{3.32}$$

where $\chi_{\{\dots\}}$ denotes an indicator function.

Theorem 3.7 is essentially a Law of Large Numbers for random variables $(-\log K^{\omega,\Lambda})$. The statement of the theorem can be alternatively expressed as follows. For any $\epsilon > 0$,

$$\lim_{\Lambda \nearrow \mathbb{Z}^d} \mathcal{P}^{\omega,\Lambda}\left(2^{-|\Lambda|(h+\epsilon)} \leq K^{\omega,\Lambda} \leq 2^{-|\Lambda|(h-\epsilon)}\right) = 1. \tag{3.33}$$

Thus we can define a typical subspace $\mathcal{T}_\epsilon^{\omega,\Lambda}$ by

$$\mathcal{T}_\epsilon^{\omega,\Lambda} := \text{span}\,\{|\psi_j\rangle \,:\, 2^{-|\Lambda|(h+\epsilon)} \leq \lambda_j \leq 2^{-|\Lambda|(h-\epsilon)}\}. \tag{3.34}$$

It clearly satisfies the analogues of *(a)* and *(b)* of the Typical Subspace Theorem, which implies as before that a compression scheme of rate R is reliable if and only if $R > h$.

3.6 Universal and variable length data compression

Thus far we discussed source-dependent data compression for various classes of quantum sources. In each case data compression relied on the identification of the typical subspace of the source, which in turn required a knowledge of its density matrix. In Classical Information Theory, there exists a generalisation of the Theorem of Typical Sequences due to Cziszar and Körner [11] where the typical set is **universal**, in that it is typical for every possible probability distribution with a given entropy. This result was used by Jozsa et al. [22] to construct a universal compression scheme for quantum i.i.d sources with a given von Neumann entropy S using a counting argument for symmetric subspaces. This was generalised to ergodic sources by Kaltchenko and Yang [25] along the lines of Theorem 3.4. Hayashi and Matsumoto [14, 15] supplemented the work of [22] with an estimation of the eigenvalues of the source (using the measurement smearing technique) to show that a reliable compression scheme exists for *any* quantum i.i.d source, independent of the value of its von Neumann entropy S, the

limiting rate of compression being given by S. If one admits variable length coding, the Lempel-Ziv algorithm gives a completely universal compression scheme, independent of the value of the entropy, in the classical case [38, 10]. This algorithm was generalised to the quantum case for i.i.d sources by Jozsa & Presnell [23], and to sources modelled by Gibbs states of free bosons or fermions on a lattice by Suhov and Johnson [19, 20]. A more general analysis of variable length coding, and in particular prefix-free codes was initiated by Schumacher and Westmoreland [34].

Another important question is the *efficiency* of the the various coding schemes. The above-mentioned schemes for quantum i.i.d. sources are not efficient, in the sense that they have no polynomial time implementation. Recently, it was shown by Bennett et al. [3] that an efficient, universal compression scheme for i.i.d sources can be constructed by employing quantum state tomography.

Acknowledgements The authors would like to thank Y.M.Suhov for helpful discussions.

References

[1] C. Ahn, A. Doherty, P. Hayden & A. Winter. On the distributed compression of quantum information. quant-ph/0403042, 2004.

[2] H. Barnum et al. On quantum coding for ensembles of mixed states. *J. Phys. A* **34**, 6767–6786, 2001.

[3] C. H. Bennett, A. W. Harrow & S. Lloyd. Universal quantum data compression via gentle tomography. quant-ph/0403078, 2004.

[4] P. Billingsley. *Ergodic Theory and Information.* John Wiley & Sons, New York etc. 1965.

[5] I. Bjelaković, T. Krüger, R. Siegmund-Schultze & A. Szkoła. The Shannon-McMillan theorem for ergodic quantum lattice systems. *Invent. math.* **155**, 203–222, 2004.

[6] I. Bjelaković, T. Krüger, R. Siegmund-Schultze & A. Szkoła. Chained typical subspaces–a quantum version of Breiman's theorem. quant-ph/0301177.

[7] C. Borgs, R. Kotecký, D. Ueltschi. Low temperature phase diagrams for quantum perturbations of classical spin systems. *Commun. Math. Phys.*, **181**, 409–46, 1996.

[8] S. Braunstein, C. Fuchs, D. Gottesman & H.-K. Lo. A quantum analog of Huffman coding. quant-ph/9805080, 1998.

[9] L. Breiman. The individual ergodic theorem of information theory. *Ann. Math. Stat.* **28**, 809–811 (1957). Correction note, *ibid.* **31**, 809-810 (1960).

[10] T. M. Cover & J. A. Thomas. *Elements of information theory.* John Wiley and Sons, New York, 1991.

[11] I. Csiszár & J. Körner. *Information Theory. Coding Theorems for Discrete Memoryless Systems.* Akadémiai Kiadó, Budapest, 1981.

[12] N. Datta, R. Fernández, & J. Fröhlich. Low-temperature phase diagrams of quantum lattice systems. I. Stability for quantum perturbations of classical systems with finitely-many ground states. *J. Stat. Phys.*, **84**, 455–534, 1996.

[13] N. Datta & Y. Suhov. Data Compression Limit for an Information Source of Interacting Qubits. *Quantum Information Processing* **1**(4), 257-281, 2002.

[14] M. Hayashi & K. Matsumoto. Quantum universal variable length source coding. quant-ph/0202001, 2002.

[15] M. Hayashi & K. Matsumoto. Simple construction of quantum universal variable length source coding. quant-ph/0209124, 2002.

[16] P. Hayden, R. Jozsa & A. Winter. Trading quantum for classical resources in quantum data compression. *J. Math. Phys.* **43**, 4404–4444, 2002; quant-ph/0209124, 2002.

[17] F. Hiai & D. Petz. The proper formula for the relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.* **143**, 257–281, 1991.

[18] A. S. Holevo. Quantum coding theorems. *Russ. Math. Surveys* **53**, 1295–1331, 1998.

[19] O.Johnson & Y.M.Suhov. The von Neumann entropy and information rate for integrable quantum Gibbs ensembles. *Quantum computers and computing* **3/1**, 3-24, 2002.

[20] O.Johnson & Y.M.Suhov. The von Neumann entropy and information rate for integrable quantum Gibbs ensembles 2. *Quantum computers and computing* **4/1**, 128-143, 2003.

[21] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Optics* **41**, 2314–2323, 1994.

[22] R. Jozsa , M. Horodecki, P. Horodecki & R. Horodecki, . Universal quantum information compression. *Phys. Rev. Lett.* **81**, 1714–1717, 1998.

[23] R. Jozsa & S. Presnell. Universal quantum information compression and degrees of prior knowledge. quant-ph/0210196, 2002.

[24] R. Jozsa & B. Schumacher. A new proof of the quantum noiseless coding theorem. *J. Mod. Optics* **41**, 2343–2349, 1994.

[25] A. Kaltchenko & E.-H. Yang. Universal compression of ergodic quantum sources. *Quant. Inf. Comput.* **3**, 359–375, 2003.

[26] C. King & A. Lesniewski. Quantum sources and a quantum coding theorem. *J. Math. Phys.* **39**, 88–101, 1998.

[27] E. H. Lieb & M. B. Ruskai. Proof of the strong subadditivity of quantum mechanical entropy. *J. math. Phys.* **14**, 1938–1941, 1973.

[28] B. McMillan. The basic theorems of information theory. *Ann. Math. Stat.* **24**, 196–219, 1953.

[29] Y. Mitsumori, et al. Experimental demonstration of quantum source coding. quant-ph/0304036, 2003.

[30] M. A. Nielson & I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, 2000.

[31] M. Ohya & D. Petz. *Quantum Entropy and Its Use.* Springer Verlag, Heidelberg, 1993.

[32] D. Petz & M. Mosonyi. Stationary quantum source coding. *J. Math. Phys.* **42**, 4857–4864, 2001.

[33] B. Schumacher. Quantum Coding. *Phys. Rev. A* **51**, 2738-2747, 1995.

[34] B. Schumacher & M. D. Westmoreland. Indeterminate length quantum coding. quant-ph/0011014, 2002.

[35] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.* **27**, 379-423 and 623-656, 1918.

[36] A. Winter. *Coding theorems for quantum information theory.* Ph. D. thesis, Univ. Bielefeld. quant-ph/9907077, 1999.

[37] A. Winter. Compression of sources of probability distributions and density operators. quant-ph/0208131, 2002.

[38] J. Ziv & A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory* **IT-23**, 3337–, 1977.