Quantum sources and a quantum coding theorem

Christopher King[a)]
Department of Mathematics, Northeastern University, Boston, Massachusetts 02115

Andrzej Leśniewski[b)]
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 28 March 1996; accepted for publication 16 September 1997)

We define a large class of quantum sources and prove a quantum analog of the asymptotic equipartition property. Our proof relies on using local measurements on the quantum source to obtain an associated classical source. The classical source provides an upper bound for the dimension of the relevant subspace of the quantum source, via the Shannon–McMillan noiseless coding theorem. Along the way we derive a bound for the von Neumann entropy of the quantum source in terms of the Shannon entropy of the classical source, and we provide a definition of ergodicity of the quantum source. Several explicit models of quantum sources are also presented. © *1998 American Institute of Physics.* [S0022-2488(98)00401-0]

I. INTRODUCTION

A. Context

The possibility of building a quantum computer has stimulated new interest in the quantum analog of classical information theory (see Ref. 1 for an introduction and review of current ideas). Shannon, McMillan, Khinchin, and others provided a firm foundation for the classical theory, and used the mathematics of stochastic processes to prove important theorems. In particular they obtained limits on the amount of information that can be transmitted through a channel. Although a quantum computer does not yet exist, it is reasonable to suppose that similar issues of channel capacity are relevant to its operation. So it is interesting to investigate this question, using our current understanding of quantum mechanics.

There has been much work done on this and related questions. In particular, Schumacher stated and proved a capacity result for a quantum channel in Refs. 2 and 3. Our particular interest is in the *extended* quantum signal source described by Schumacher. This is a quantum system whose state space is a (tensor) product of many copies of one fundamental state space M. The source produces a signal that is encoded by a state in M; the ensemble of possible signals is represented by a density operator ρ on M. The extended source corresponds to a sequence of such states, and has a natural interpretation as a message. The probabilistic character of the message is contained in the density operator on the tensor product of copies of M. One choice of density operator is $\rho\otimes\cdots\otimes\rho$. This corresponds to independent signals at all times, and there are no correlations between signals in the message. For this reason we call this a quantum Bernoulli source (a precise definition is provided in Sec. II). Schumacher proves his quantum noiseless coding theorem for such a source. He introduces the notion of fidelity of a quantum channel, and states his results in terms of this. On the most basic level his results show that the state space for the extended source has a relevant subspace whose dimension is determined by the von Neumann entropy of ρ. As far as the information content of the source is concerned, the rest of the state space can be ignored.

[a)]Electronic mail: king@neu.edu
[b)]Electronic mail: lesniewski@huhepl.harvard.edu

0022-2488/98/39(1)/88/14/$15.00

B. Results

Our main interest in this paper is to extend Schumacher's result to a large class of quantum sources. We define precisely what we mean by this in Sec. II. Roughly speaking, we consider sources that allow correlations on all time scales between signals in a message. In classical information theory the corresponding result is the Shannon–McMillan theorem, which shows that under very general assumptions the ensemble of all possible messages can be split into a relevant and an irrelevant part. The criterion for splitting is provided by the entropy of the source. The corresponding result for the quantum theory should be a splitting of the state space into a relevant subspace and an irrelevant subspace, with the von Neumann entropy as the criterion. This is precisely what Schumacher proved for the quantum Bernoulli source. We obtain such a splitting in the general case, and derive an estimate for the dimension of the relevant subspace, by computing the entropy of a classical source that is obtained by making measurements on the quantum system.

To summarize, we show that a general quantum source can be encoded by a quantum system whose dimension is smaller than the original. We give estimates for the dimension of the reduced space, based on the results of local measurements of the source. For the case of a quantum source emitting orthogonal states our estimate is tight, and for a Bernoulli source reproduces Schumacher's result. We also derive an inequality relating this experimental entropy to the true von Neumann entropy.

C. Layout

The paper is organized as follows. In sec. II we define quantum sources and recall some standard results about the construction of infinite quantum systems. We also provide some explicit examples that serve to illustrate our ideas. In Sec. III we recall the classical notion of an ergodic source, and propose a definition of ergodicity for a quantum source. As a check we prove that a quantum Bernoulli source is ergodic. In Sec. IV we recall the notion of quantum entropy. In Sec. V we pursue the quantum analog of the Shannon–McMillan theorem. To do this we use the notion of a positive operator-valued measure, and construct an associated classical source. The entropy of this associated source satisfies a lower bound involving the von Neumann entropy of the quantum source. It also determines the dimension of a relevant subspace of the quantum system, which in turn yields our quantum noiseless coding theorem.

II. QUANTUM SOURCES

A. Setup

According to the mathematical theory of information, see, e.g., Refs. 4–6, a (classical) *source* is a stochastic process. In this paper, we are concerned with discrete sources that can be described as follows. We are given a finite set $\mathscr{X}$, called the alphabet, and consider the space $\mathscr{X}^{\infty}$ of all infinite sequences $\underline{x}=\{x_n\}_{n\in Z}$, called messages. The time evolution is given by the shift T: $\mathscr{X}^{\infty}\to\mathscr{X}^{\infty}$ defined by $(T\underline{x})_n:=x_{n+1}$. The stochastic character of the process is governed by a probability measure μ on $\mathscr{X}^{\infty}$. The triple $(\mathscr{X}^{\infty},T,\mu)$ is called a source. We say that the source is *stationary*, if T preserves the measure μ.

We begin by describing informally the properties of a quantum source. In the next section we will provide a rigorous mathematical description. By analogy with the classical case, a quantum source will be a triple, consisting of *quantum messages*, the *time shift*, and a *probability distribution* for the messages. For technical reasons, it is useful to describe the space of quantum messages (which is a linear space) by the algebra of observables on it.

A *quantum source* sends a series of signals, each of which is a vector in a finite-dimensional Hilbert space $\mathscr{H}$. We assume that the source is discrete, i.e., each signal is an element of a finite set $\mathscr{S}=\{|\psi_1\rangle,\ldots,|\psi_s\rangle\}$ of normalized vectors in $\mathscr{H}$. To avoid unnecessary redundancy, we assume that the space $\mathscr{H}$ is spanned by $\mathscr{S}$, i.e. $\mathscr{H}\simeq\mathbb{C}^d$, where $d\leqslant s$. We do not assume that the elements of $\mathscr{S}$ are orthogonal to each other or even linearly independent. Indeed, this is an important difference from the classical situation, which does not allow for forming linear superpositions of states. Denote by p_j the *a priori* probability of the state $|\psi_j\rangle$ being sent. The density matrix corresponding to the ensemble of signals $\mathscr{S}$ is then given by

$$\rho = \sum_{1 \leq j \leq s} p_j |\psi_j\rangle\langle\psi_j|. \tag{II.1}$$

As a consequence of our assumptions, $\mathrm{tr}(\rho)=1$. Clearly, an ensemble of signals $\mathscr{S}$ and the associated *a priori* probabilities determines uniquely the density matrix ρ. On the other hand, any given density matrix corresponds to infinitely many different sets of signals. For a discussion of this point, see Ref. 7.

The observables associated with quantum signals are $d \times d$ Hermitian matrices; more formally (we will need this viewpoint shortly) they are elements of the $\mathbb{C}^*$-algebra $\mathscr{A}=\mathscr{L}(\mathscr{H})$ of linear operators on $\mathscr{H}$. Abusing slightly the language, we will refer to $\mathscr{A}$ as the algebra of observables. Using the language adopted in the operator algebra approach to quantum physics, see, e.g., Refs. 8 and 9, we define the *state* on the algebra of observables $\mathscr{A}$ associated with the density matrix ρ to be

$$\tau_1(A) := \mathrm{tr}(A\rho) = \sum_{1 \leq j \leq s} p_j \langle\psi_j|A|\psi_j\rangle. \tag{II.2}$$

B. Definition

We now propose a formal definition of a quantum source. To this end, we construct the state on the algebra of observables associated with entire (infinite) quantum messages rather than individual signals. This is technically somewhat delicate, as it involves infinite tensor products of Hilbert spaces. Let $I \subset \mathbb{Z}$ be a finite set of the form $\{M, M+1, ..., N-1, N\}$, where $M<N$, i.e., I is a finite collection of consecutive integers. By $\mathscr{T}$ we denote the partially ordered set of all such I's. We set $H_I := \otimes_{j \in I} \mathscr{H}_j$, where $\mathscr{H}_j = \mathscr{H}$ for all $j \in I$, and define the corresponding observable algebra $\mathsf{A}_I := \mathscr{L}(H_I) \simeq \otimes^{|I|} \mathscr{L}(\mathscr{H})$, where $|I|$ denotes the number of elements in I. For $I \subset J$, there is a natural embedding $\mathscr{L}(H_I) \hookrightarrow \mathscr{L}(H_J)$, and so we can form the union $\mathsf{A}_{\mathrm{loc}} := \cup_{I \in \mathscr{T}} \mathsf{A}_I$. The latter is a normed algebra, and we refer to its elements as *local observables*. Roughly, $\mathsf{A}_{\mathrm{loc}}$ is a collection of operators acting on the infinite tensor product $\otimes_{j \in \mathbb{Z}} \mathscr{H}_j$; every element of $\mathsf{A}_{\mathrm{loc}}$ acts as the identity on all but a finite number of factors in this product. For a local observable $A \in \mathsf{A}_{\mathrm{loc}}$, we let $\mathrm{supp}(A)$ denote its support, i.e., the smallest $I \in \mathscr{T}$ such that $A \in \mathsf{A}_I$. The norm closure A of $\mathsf{A}_{\mathrm{loc}}$ is a $\mathbb{C}^*$-algebra called the algebra of *quasilocal observables*. These concepts are borrowed from algebraic field theory and statistical mechanics, and we refer the reader to Refs. 8 and 9 for a thorough presentation.

We will use the net $\{\mathsf{A}_I\}_{I \in \mathscr{T}}$ of matrix algebras to construct the Hilbert space of states of a quantum source. Assume that we have a family $\{\Pi_I\}_{I \in \mathscr{T}}$, $\Pi_I \in \mathsf{A}_I$, satisfying the following assumptions:

1°. Each Π_I is a positive operator.
2°. If $|I|=1$, then $\Pi_I = \rho$.
3°. Let $I \subset J$ be such that $J \setminus I \in \mathscr{T}$. Then

$$\mathrm{tr}_{H_I} \Pi_J = \Pi_I, \tag{II.3}$$

where tr_{H_I} denotes the partial trace over the factor H_I in the tensor product $H_J = H_I \otimes H_{J \setminus I}$ or $H_J = H_{J \setminus I} \otimes H_I$.

Note that the last condition implies, in particular, that $\mathrm{tr}\, \Pi_I = 1$. In other words, $\{\Pi_I\}_{I \in \mathscr{T}}$ is a *consistent* family of density matrices, and it can be thought of as a quantum mechanical counterpart of a consistent family of cylinder measures.

For each $I \in \mathscr{I}$ we define a state τ_I on A_I by

$$\tau_I(A) := \mathrm{tr}(A\Pi_I), \tag{II.4}$$

and observe that $|\tau_I(A)| \leq \|A\|$, uniformly in I. The consistency condition (II.3) implies that τ_I is well defined, and that the generalized limit $\tau(A) := \lim_{I \nearrow \mathbb{Z}} \tau_I(A)$ exists for all $A \in \mathsf{A}_{\mathrm{loc}}$, and satisfies $|\tau(A)| \leq \|A\|$. As a consequence, τ can be uniquely extended to a state on the $\mathbb{C}^*$-algebra A of quasilocal observables. We use the same symbol τ to denote this extension.

Let H, π be the GNS representation, see, e.g., Ref. 9, associated with the state τ. The Hilbert space H is the state space of the quantum source. If no confusion arises, we will write A instead of $\pi(A)$.

The additive group $\mathbb{Z}$ underlying the above construction plays the role of (discrete) time translations. Its action on the algebra of local observables is defined as follows:

$$A_I \ni A \simeq A \otimes I \rightarrow \alpha(A) := I \otimes A \simeq A \in A_{I+1}. \tag{II.5}$$

In other words, α pushes the observable to the right by one unit of time. Clearly, $\|\alpha(A)\| = \|A\|$, and so α has a unique extension to all of A, which we will denote by the same symbol. The family of automorphisms $\{\alpha^n\}_{n \in \mathbb{Z}}$ defines then a representation of $\mathbb{Z}$. The triple $(\mathsf{A}, \alpha, \tau)$ is called a *quantum source*.

We say that the quantum source is *stationary*, if the state τ is invariant under α, i.e.

$$\tau(\alpha(A)) = \tau(A), \tag{II.6}$$

for all $A \in \mathsf{A}$. From now on we will be assuming that our source is stationary. A standard result in operator algebras implies that the automorphism α is unitarily implementable on the GNS Hilbert space associated with the invariant state τ, i.e., $\alpha(A) = FAF^{-1}$ on H. We will call the unitary operator F a *quantum shift*.

Because of stationarity, we can always assume that $I \in \mathscr{I}$ is of the form $\{1,...,n\}$. To simplify the notation, we will write then Π_n instead of Π_I, and τ_n instead of τ_I.

C. Examples

The simplest example of a quantum source is a *Bernoulli source*, which we now describe. As in the classical case, a Bernoulli source produces messages that are sequences of independent signals. Accordingly, the family of density matrices $\{\Pi_I\}$ is given as follows:

$$\Pi_I = \otimes_{i \in I} \rho_i, \tag{II.7}$$

where each ρ_i equals ρ. It immediately follows that the consistency condition (II.3) is satisfied, and that the source is stationary. This is the class of sources considered by Schumacher.[2] One special feature of a Bernoulli source is that whenever $\mathrm{supp}(A) \cap \mathrm{supp}(B) = \emptyset$, we have $\tau(AB) = \tau(A)\tau(B)$.

There are many examples of non-Bernoulli sources. We present here a special class of stationary sources. These are all described by a signal density matrix ρ and another matrix $R \in \mathscr{L}(\mathscr{H} \otimes \mathscr{H})$ satisfying several consistency and positivity conditions. The most basic conditions are

$$R = R^*, \quad \mathrm{tr}_1((\rho \otimes I)R) = \rho, \quad \mathrm{tr}_2(R) = I. \tag{II.8}$$

Here I is the identity matrix on $\mathscr{H}$, and we introduce the notation tr_i for the partial trace over the ith factor in the n-fold tensor product $\otimes^n \mathscr{H}$. The density matrices $\{\Pi_n\}$ are then constructed recursively as follows:

$$\Pi_1=\rho,$$

$$\Pi_{n+1}=\tfrac{1}{2}(\Pi_n\otimes I_1)(I_{n-1}\otimes R)+\tfrac{1}{2}(I_{n-1}\otimes R)(\Pi_n\otimes I_1).$$

We have denoted by I_n the identity matrix on the product $\otimes^n\mathscr{H}$. The consistency of this definition [(II.3)] is immediate. The positivity of the density matrices Π_n is a further constraint on R. We have several explicit examples for which the positivity can be proven.

In the simplest situation the matrix R satisfies the following additional conditions:

$$R\geqslant 0,$$

$$[\rho\otimes I_1,R]=0,$$

$$[I_1\otimes R,R\otimes I_1]=0. \tag{II.9}$$

It follows readily that all the matrices Π_n are positive, for all $n\geqslant 1$. For example, suppose $\rho=\Sigma_{j=1}^{d}\lambda_j P_j$, where $\{\lambda_j\}$ are the eigenvalues of ρ, and $\{P_j\}$ are the corresponding orthogonal projections. Then we can take $R=\Sigma_{j=1}^{d}P_j\otimes P_j$, and all the above properties are easily seen to hold.

For our second example $\mathscr{H}=\mathbb{C}^2$, and we assume that ρ is strictly positive. Let $\{\sigma_j\}$ be the Pauli matrices:

$$\sigma_1=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad \sigma_2=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

By choosing a suitable basis we can write

$$\rho=\frac{1}{2}I+\frac{a}{2}\sigma_3, \tag{II.10}$$

where $|a|<1$. Then we take the matrix R to be

$$R=I\otimes\rho+\left(\frac{a}{2}I-\frac{1}{2}\sigma_3\right)\otimes(b\sigma_1+c\sigma_2). \tag{II.11}$$

The consistency conditions are easily verified. Positivity of all matrices $\{\Pi_n\}$ holds for $|b|$, $|c|$ sufficiently small; the proof is given in the Appendix.

D. Reduction to classical source

The density matrix Π_n can also be written in terms of the states $\{\psi_j\}$ that span $\mathscr{S}$ [compare (II.1)] as follows:

$$\Pi_n=\sum_{1\leqslant j_1,\ldots,j_n\leqslant s}p_{j_1,\ldots,j_n}|\psi_{j_1}\rangle\langle\psi_{j_1}|\otimes\cdots\otimes|\psi_{j_n}\rangle\langle\psi_{j_n}|.$$

The numbers $p_{j_1,\ldots,j_n}$ satisfy obvious consistency conditions. If the states $\{\psi_j\}$ are linearly independent, then they are also non-negative. Further, if the states are orthogonal, the density matrices $\{\Pi_n\}$ commute. In this case, the quantum source is equivalent to a classical source.

III. ERGODIC QUANTUM SOURCES

A. Definition

According to the individual ergodic theorem,[10] for any function $f \in L^1(\mathscr{X}^{\cdot\infty})$, the sequence of time averages

$$\langle f \rangle_N(\underline{x}) := \frac{1}{N} \sum_{0 \leqslant n \leqslant N-1} f(T^n \underline{x}), \tag{III.1}$$

converges almost everywhere to a limit $\overline{f}$. The function $\overline{f}$ is invariant under T. The source $(\mathscr{X}^{\cdot\infty}, T, \mu)$ is called *ergodic*, if the only functions invariant under T are constants. Consequently, for an ergodic source $\overline{f} = \int_{\mathscr{X}^{\cdot\infty}} f(\underline{x}) d\mu(\underline{x})$. This condition is one of the several equivalent statements that could be used to define ergodicity; see, e.g., Ref. 10. There is no natural concept of convergence almost everywhere in quantum mechanics. Consequently, we adopt a notion of convergence that is natural for the algebra A. Our definition of ergodicity of a quantum source is specifically designed to suit the purposes of this paper, although it may have broader applications.

We say that a quantum source is *ergodic* if the following condition is satisfied. For any $A \in \mathsf{A}$, the time averages,

$$\langle A \rangle_N := \frac{1}{N} \sum_{0 \leqslant n \leqslant N-1} F^n A F^{-n}, \tag{III.2}$$

converge in a suitable sense to the limit $\tau(A)I$, as $N \to \infty$. Specifically, we require weak convergence, i.e., for all ϕ, $\psi \in H$,

$$\lim_{N \to \infty} (\phi, (\langle A \rangle_N - \tau(A)I)\psi) = 0. \tag{III.3}$$

For $A \in \mathsf{A}$, we let $[A]$ denote the corresponding element of H. Substituting $\phi = [I]$ and $\psi = [B]$, $B \in \mathsf{A}_{\mathrm{loc}}$ in (III.3), we find that for an ergodic source,

$$\lim_{N \to \infty} \tau(\langle A \rangle_N B) = \tau(A)\tau(B), \tag{III.4}$$

for all $A, B \in \mathsf{A}_{\mathrm{loc}}$.

B. Bernoulli source

The simplest example of an ergodic source is a Bernoulli source.

Proposition III.1: A quantum Bernoulli source is ergodic.

Proof: Let $A, B, C \in \mathsf{A}_{\mathrm{loc}}$. We assert that

$$\tau(B^\dagger \langle A \rangle_N C) \to \tau(B^\dagger C)\tau(A), \tag{III.5}$$

which is equivalent to (III.3) with $\phi = [B]$, $\psi = [C]$. To prove this, we observe that there is n_0, such that

$$\tau(B^\dagger F^n A F^{-n} C) = \tau(F^n A F^{-n})\tau(B^\dagger C) = \tau(A)\tau(B^\dagger C),$$

for $n > n_0$ (this follows from the fact that the supports of A, B, and C are all finite). Consequently,

$$\tau(B^\dagger \langle A \rangle_N C) = \frac{1}{N} \sum_{0 \leqslant n \leqslant n_0} \tau(B^\dagger F^n A F^{-n} C) + \frac{N - n_0 - 1}{N} \tau(A)\tau(B^\dagger C).$$

But

$$\frac{1}{N}\left|\sum_{0\le n\le n_0}\tau(B^\dagger F^n A F^{-n} C)\right|\le\frac{n_0+1}{N}\|A\|\|B\|\|C\|\to 0,$$

and so $\lim_{N\to\infty}\tau(B^\dagger\langle A\rangle_N C)=\tau(A)\tau(B^\dagger C)$.

The remainder of the proof is a series of straightforward approximation arguments.

1°. We claim that (III.5) holds for all $A,B,C\in\mathsf{A}$. To prove this, observe that for all N,

$$\|\langle A\rangle_N\|\le\|A\|. \tag{III.6}$$

For $A\in\mathsf{A}$, let $A_j\in\mathsf{A}_{\mathrm{loc}}$ be such that $\|A-A_j\|\to 0$, as $j\to\infty$. For any $B,C\in\mathsf{A}_{\mathrm{loc}}$, write

$$\tau(B^\dagger\langle A\rangle_N C-\tau(A)B^\dagger C)=\tau(B^\dagger\langle A_j\rangle_N C-\tau(A_j)B^\dagger C)$$
$$+\tau(B^\dagger\langle A-A_j\rangle_N C)+\tau(A_j-A)\tau(B^\dagger C),$$

and choose j such that $\|A-A_j\|\le\epsilon/(3\|B\|\|C\|)$. Now choose $N_0=N_0(j)$ such that $|\tau(B^\dagger\langle A_j\rangle_N C-\tau(A_j)B^\dagger C)|<\epsilon/3$, for all $N>N_0$. Then, using (III.6),

$$|\tau(B^\dagger\langle A\rangle_N C-\tau(A)B^\dagger C)|<\epsilon/3+\|\langle A-A_j\rangle_N\|\|B\|\|C\|+\|A_j-A\|\|B\|\|C\|$$
$$<2\|A-A_j\|\|B\|+\epsilon/3\le\epsilon,$$

for all $N>N_0$. We have thus shown that (III.5) holds for all $A\in\mathsf{A}$ and $B,C\in\mathsf{A}_{\mathrm{loc}}$. Repeating twice almost *verbatim* the above 3ϵ argument we establish (III.5) for all $A,B,C\in\mathsf{A}$.

2°. Having established (III.3) for $\phi=[B]$, $\psi=[C]$, and $A\in\mathsf{A}$, we now show that it holds for arbitrary ϕ, $\psi\in\mathsf{H}$. Let $\phi\in\mathsf{H}$, and let B_j be such that $\|\phi-[B_j]\|<\epsilon/(2\|[C]\|\|A\|)$. Write

$$(\phi,(\langle A\rangle_N-\tau(A)I)[C])=([B_j],(\langle A\rangle_N-\tau(A)I)[C])+(\phi-[B_j],(\langle A\rangle_N-\tau(A)I)[C]),$$

and choose $N_0=N_0(j)$ so that $|([B_j],(\langle A\rangle_N-\tau(A)I)[C])|<\epsilon/2$, for all $N>N_0$. Then

$$|(\phi,(\langle A\rangle_N-\tau(A)I)[C])|<\epsilon/2+\|\phi-[B_j]\|\|[C]\|\|A\|<\epsilon.$$

Repeating this argument we establish (III.3) for all $\psi\in\mathsf{H}$. The proof of the proposition is complete.

IV. ENTROPY OF A QUANTUM SOURCE

A. Definition

In this section we construct the entropy of a quantum source. Our construction is largely standard; see, e.g., Refs. 11 and 12, but we include most of the details to make the presentation self-contained. The key mathematical input is the following lemma, known as Klein's inequality,[13] whose proof can be found in Ref. 12, as well as on p. 1122 of Ref. 11.

Lemma IV.1: Let A and B be positive trace class operators on a Hilbert space. Then

$$\mathrm{tr}(A\log A-A\log B)\ge\mathrm{tr}(A-B). \tag{IV.1}$$

We define the entropy associated with a sequence of n signals to be

$$H_n(\Pi):=-\mathrm{tr}_{\mathscr{H}^{\otimes n}}(\Pi_n\log\Pi_n). \tag{IV.2}$$

Substituting $A=\Pi_{m+n}$ and $B=\Pi_m\otimes\Pi_n$ in (IV.1), we obtain the following subadditivity property of $H_n(\Pi)$ (this is a special case of the well-known subadditivity of quantum mechanical entropy; see Ref. 11):

$$H_{m+n}(\Pi)\leqslant H_m(\Pi)+H_n(\Pi). \tag{IV.3}$$

A standard argument, see, e.g., Ref. 6, pp. 48–49, shows that the limit

$$h(\Pi)=\lim_{n\to\infty}\frac{1}{n}H_n(\Pi) \tag{IV.4}$$

exists. We call $h(\Pi)$ the entropy of the quantum source. Notice also that subadditivity implies the inequality

$$h(\Pi)\leqslant-\mathrm{tr}_{\mathscr{H}}(\rho\ \log\ \rho)\leqslant\log\ d. \tag{IV.5}$$

where ρ is the density matrix for the signals, and d is the dimension of the signal Hilbert space $\mathscr{H}$.

B. Examples

We can easily compute the entropy of two basic types of quantum sources introduced in Sec. II.

For a Bernoulli source, $\Pi_n=\rho\otimes\cdots\otimes\rho$, and so

$$\begin{aligned}\frac{1}{n}H_n(\Pi)&=-\frac{1}{n}\,\mathrm{tr}_{\mathscr{H}^{\otimes n}}((\otimes^n\rho)(\log\ \otimes^n\rho))\\&=-\frac{1}{n}\sum_{1\leqslant j\leqslant n}\mathrm{tr}_{\mathscr{H}^{\otimes n}}((\otimes^n\rho)(I\otimes\cdots\otimes\log\ \rho\otimes\cdots\otimes I))\\&=-\mathrm{tr}_{\mathscr{H}}(\rho\ \log\ \rho).\end{aligned}$$

As a result, $h(\Pi)=-\mathrm{tr}(\rho\ \log\rho)$, i.e., the entropy of a Bernoulli source is equal to the von Neumann entropy of the signal density matrix.

For the non-Bernoulli source defined in (II.9) a similar calculation yields $h(\Pi)=-\mathrm{tr}((\rho\otimes I)R\ \log R)$. We have not found a closed form expression for the entropy of the other non-Bernoulli source described in Sec. II.

V. ASYMPTOTIC EQUIPARTITION PROPERTY

A. Classical result

The classical asymptotic equipartition property, also known as the Shannon–McMillan theorem—see, e.g., Refs. 4 and 6—states that if $(\mathscr{X}^{\infty},T,\mu)$ is an ergodic source with entropy $h(\mu)$, then the sequence,

$$f_n(\underline{x}):=-\frac{1}{n}\log\ \mu(\{\underline{y}\in\mathscr{X}^{\infty}:y_1=x_1,\ldots,y_n=x_n\}), \tag{V.1}$$

converges in measure to $h(\mu)$. In other words, given $\delta,\epsilon>0$, there is n_0 such that for all $n\geqslant n_0$,

$$\mu(\{\underline{x}\in\mathscr{X}:|f_n(\underline{x})-h(\mu)|>\delta\})<\epsilon. \tag{V.2}$$

This is interpreted as saying that for any length n, there are two categories of messages sent by a source: (i) a small fraction of ‘‘likely’’ messages, each of which carries equal probability,

$$\mu(\{\underline{y}\in\mathscr{X}^{\infty}:y_1=x_1,\ldots,y_n=x_n\})\sim e^{-nh(\mu)}, \tag{V.3}$$

and (ii) the bulk of ''unlikely'' messages, whose total probability goes to zero as n goes to infinity. There are approximately $e^{nh(\mu)}$ likely messages, which is much less than the total number of messages $e^{n\log r}$ [unless $h(\mu)$ happens to equal $\log r$].

The goal of this section is to establish an analogous result for quantum sources. Our theorem generalizes Schumacher's result[2,3] to general, not necessary Bernoulli, sources.

B. POM

Let $\mathbf{A}=\{A_1,\ldots,A_r\}$, $r<\infty$ be a family of observables on $\mathscr{H}$ such that $A_j\geqslant 0$, for all j, and

$$A_1+\cdots+A_r=I. \tag{V.4}$$

Such a family is called a *positive operator-valued measure* (POM); see Ref. 7 and references therein. A POM is called *pure* if each A_J is a rank one operator. For example, any family of $d=\dim\mathscr{H}$ pairwise orthogonal projections on $\mathscr{H}$ satisfies the above conditions, and so is a pure POM. We will call the set $\mathscr{X}_{\mathbf{A}}=\{1,...,r\}$ the classical alphabet associated with the POM $\mathbf{A}$, and denote by $\mathscr{X}_{\mathbf{A}}^{\infty}$ the space of all infinite messages over the alphabet $\mathscr{X}_{\mathbf{A}}$. We can define a probability measure on $\mathscr{X}_{\mathbf{A}}^{\infty}$ associated with the quantum source ($\mathbf{A}$, τ, α). For $\{k_1,\ldots,k_n\}\in\mathscr{X}^{n}$, we define

$$\mu_n^{\mathbf{A}}(\{\underline{x}:x_1=k_1,\ldots,x_n=k_n\}):=\tau_n(A_{k_1}\otimes\cdots\otimes A_{k_n}). \tag{V.5}$$

This defines a consistent family of cylinder measures, as

$$\sum_{k\in\mathscr{X}_{\mathbf{A}}}\mu_{n+1}^{\mathbf{A}}(\{\underline{x}:x_1=k_1,\ldots,x_n=k_n,x_{n+1}=k\})=\tau_{n+1}\left(A_{k_1}\otimes\cdots\otimes A_{k_n}\otimes\sum_{k\in\mathscr{X}_{\mathbf{A}}}A_k\right)=\tau_{n+1}(A_{k_1}\otimes\cdots\otimes A_{k_n}\otimes I)=\tau_n(A_{k_1}\otimes\cdots\otimes A_{k_n})=\mu_n^{\mathbf{A}}(\{\underline{x}:x_1=k_1,\ldots,x_n=k_n\}).$$

Let $\mu^{\mathbf{A}}$ denote the probability measure obtained from $\{\mu_n^{\mathbf{A}}\}$ by means of Kolmogorov's extension theorem. By T we denote the shift operator on $\mathscr{X}_{\mathbf{A}}^{\infty}$. Then the triple $(\mathscr{X}_{\mathbf{A}}^{\infty},T,\mu^{\mathbf{A}})$ forms a classical information source. We emphasize that it depends on the choice of $\mathbf{A}$. Obviously, this source is stationary.

C. Classical ergodicity from quantum

In fact, this source is ergodic if the underlying quantum source is ergodic. We state it as the following lemma.

Lemma V.1: If (A, α, τ) *is ergodic, then for any choice of* $\mathbf{A}$, *the classical source constructed above is ergodic.*

Proof: We show that $(\mathscr{X}_{\mathbf{A}}^{\infty},T,\mu^{\mathbf{A}})$ has the following property: for every $f\in L^1(\mathscr{X}^{\infty},d\mu^{\mathbf{A}})$, the sequence $\langle f\rangle_N$ converges in L^1 to $\int f\,d\mu^{\mathbf{A}}$. Using Fatou's lemma, and recalling the individual ergodic theorem, this implies that the only functions invariant under T are constants, and so the classical source is ergodic. We first observe that it is sufficient to show that if $\mathscr{C}$ is a cylinder set, and $\chi_{\mathscr{C}}$ denotes the corresponding characteristic function, then

$$\int_{\mathscr{X}_{\mathbf{A}}^{\infty}}|\langle\chi_{\mathscr{C}}\rangle_N(\underline{x})-\mu^{\mathbf{A}}(\mathscr{C})|\,d\mu^{\mathbf{A}}(\underline{x})\to 0, \tag{V.6}$$

as $N\to\infty$. A standard ''$\epsilon/3$'' argument then implies that for all $f\in L^1(\mathscr{X}^{\infty},d\mu^{\mathbf{A}})$, $\langle f\rangle_N$ converges to $\int f\,d\mu^{\mathbf{A}}$ in the L^1 norm. Furthermore, since $d\mu^{\mathbf{A}}$ is a probability measure, (V.6) will follow from convergence of $\langle \chi_{\mathscr{C}} \rangle_N$ to $\mu^{\mathbf{A}}(\mathscr{C})$ in the L^2 norm. This in turn is implied by the following stronger result. Suppose $\mathscr{C}$ and $\mathscr{D}$ are cylinder sets, and $\chi_{\mathscr{C}}$ and $\chi_{\mathscr{D}}$ are the corresponding characteristic functions; then

$$\int_{\mathscr{X}_{\mathbf{A}}^{\infty}} \langle \chi_{\mathscr{C}} \rangle_N(\underline{x}) \chi_{\mathscr{D}}(\underline{x}) d\mu^{\mathbf{A}}(\underline{x}) \to \mu^{\mathbf{A}}(\mathscr{C}) \mu^{\mathbf{A}}(\mathscr{D}), \tag{V.7}$$

as $N \to \infty$.

In order to prove (V.7), let $\mathscr{C} = \{\underline{x}: x_1 = k_1, \ldots, x_n = k_n\}$, $\mathscr{D} = \{\underline{x}: x_{j+1} = l_1, \ldots, x_{j+m} = l_m\}$. The corresponding observables are given by

$$G(\mathscr{C}) = A_{k_1} \otimes \cdots \otimes A_{k_n}, \quad G(\mathscr{D}) = A_{l_1} \otimes \cdots \otimes A_{l_m}.$$

There is N_0 such that for all $n > N_0$, $T^n(\mathscr{C}) \cap \mathscr{D} = \emptyset$. Therefore,

$$\frac{1}{N} \int_{\mathscr{X}_{\mathbf{A}}^{\infty}} \sum_{i=N_0+1}^{N-1} \mathscr{X}_{\mathscr{C}}(T^i \underline{x}) \mathscr{X}_{\mathscr{D}}(\underline{x}) d\mu^{\mathbf{A}}(\underline{x}) = \frac{1}{N} \tau\left(\sum_{i=N_0+1}^{N-1} \alpha^i(G(\mathscr{C})) G(\mathscr{D}) \right)$$

$$= \tau(\langle G(\mathscr{C}) \rangle_N) G(\mathscr{D})) - \frac{1}{N} \tau\left(\sum_{i=0}^{N_0} \alpha^i(G(\mathscr{C})) G(\mathscr{D}) \right).$$

By our assumption of quantum ergodicity, the first term above converges to $\mu^{\mathbf{A}}(\mathscr{C}) \mu^{\mathbf{A}}(\mathscr{D})$, as $N \to \infty$. The second term is bounded as follows:

$$\left| \frac{1}{N} \tau\left(\sum_{i=0}^{N_0} \alpha^i(G(\mathscr{C})) G(\mathscr{D}) \right) \right| \leqslant \frac{N_0+1}{N} \prod_{p=1}^{n} \|A_{k_p}\| \prod_{q=1}^{l} \|A_{l_q}\|,$$

which converges to 0 as $N \to \infty$. Finally,

$$\frac{1}{N} \int_{\mathscr{X}_A^{\infty}} \sum_{i=0}^{N_0} \chi_{\mathscr{C}}(T^i \underline{x}) \chi_{\mathscr{D}}(\underline{x}) d\mu^A(\underline{x}) \geqslant \frac{N_0+1}{N} \to 0,$$

and the proof is complete.

D. Bounds for the entropy

Let $h_{\mathbf{A}}$ denote the Shannon entropy of the classical source $(\mathscr{X}_{\mathbf{A}}^{\infty}, T, \mu^{\mathbf{A}})$ (Refs. 5 and 6). In other words,

$$h_{\mathbf{A}} = -\lim_{n \to \infty} \frac{1}{n} \sum_{k_1, \ldots, k_n \in \mathscr{X}_{\mathbf{A}}} \mu_{\mathbf{A}}(\{\underline{x}: x_1 = k_1, \ldots, x_n = k_n\}) \times \log \mu_{\mathbf{A}}(\{\underline{x}: x_1 = k_1, \ldots, x_n = k_n\})$$

$$= -\lim_{n \to \infty} \frac{1}{n} \sum_{k_1, \ldots, k_n \in \mathscr{X}_{\mathbf{A}}} \tau(A_{k_1} \otimes \cdots \otimes A_{k_n}) \log \tau(A_{k_1} \otimes \cdots \otimes A_{k_n}).$$

A remarkable fact about $h_{\mathbf{A}}$ is that it can be bounded from below in terms of $h(\Pi)$, the quantum entropy of the source, and a quantity depending exclusively on the statistical properties of the signal ensemble.

Theorem V.2: *For any POM* **A** *the following inequality holds*:

$$h_{\mathbf{A}} \geqslant h(\Pi) - \sum_{1 \leqslant k \leqslant r} \mathrm{tr}(A_k \rho) \log \mathrm{tr}(A_k). \tag{V.8}$$

Remark: In particular, if **A** is a pure POM consisting of d mutually orthogonal projections, then $h_{\mathbf{A}} \geqslant h(\Pi)$.

Proof: The proof of this theorem is based on the following lemma.

Lemma V.3: Let $\mathscr{F}$ be a Hilbert space of finite dimension N, and let B be a density matrix on $\mathscr{F}$. If $A_1, \ldots, A_r$ are positive operators such that $A_1 + \cdots + A_r = I$, then

$$\sum_{1 \leqslant j \leqslant r} \mathrm{tr}(A_j B) \log \mathrm{tr}(A_j B) \leqslant \mathrm{tr}(B \log B) + \sum_{1 \leqslant j \leqslant r} \mathrm{tr}(A_j B) \log \mathrm{tr}(A_j). \tag{V.9}$$

Proof: Let A_j^{kl} denote the matrix entries of A_j in an orthonormal basis consisting of eigenvectors of B. Then $\mathrm{tr}(A_j B) = \Sigma_k \lambda_k A_j^{kk}$, where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of B. The function $[0,1] \ni x \to f(x) := x \log x$ is convex, and so by Jensen's inequality,

$$\begin{aligned} \sum_j f\left(\sum_k \lambda_k A_j^{kk} \right) &= \sum_j \mathrm{tr}(A_j) f\left(\sum_k \frac{A_j^{kk}}{\mathrm{tr}(A_j)} \lambda_k \right) + \sum_j \sum_k \frac{A_j^{kk}}{\mathrm{tr}(A_j)} \lambda_k f(\mathrm{tr}(A_j)) \\ &\leqslant \sum_j \sum_k A_j^{kk} f(\lambda_k) + \sum_j \mathrm{tr}(A_j B) \log \mathrm{tr}(A_j) \\ &= \sum_k f(\lambda_k) + \sum_j \mathrm{tr}(A_j B) \log \mathrm{tr}(A_j), \end{aligned}$$

which implies (V.9).

As a consequence of this lemma, and condition 3° in the definition of a consistent family of density matrices,

$$\begin{aligned} &\sum_{k_1, \ldots, k_n \in \mathscr{X}_{\mathbf{A}}} \mathrm{tr}((A_{k_1} \otimes \cdots \otimes A_{k_n}) \Pi_n) \log \mathrm{tr}((A_{k_1} \otimes \cdots \otimes A_{k_n} \Pi_n)) \\ &\leqslant \mathrm{tr}(\Pi_n \log \Pi_n) + \sum_{k_1, \ldots, k_n \in \mathscr{X}_{\mathbf{A}}} \mathrm{tr}((A_{k_1} \otimes \cdots \otimes A_{k_n}) \Pi_n) \log \mathrm{tr}(A_{k_1} \otimes \cdots \otimes A_{k_n}) \\ &= \mathrm{tr}(\Pi_n \log \Pi_n) + \sum_{1 \leqslant j \leqslant n} \sum_{k_1, \ldots, k_n \in \mathscr{X}_{\mathbf{A}}} \mathrm{tr}((A_{k_1} \otimes \cdots \otimes A_{k_n}) \Pi_n) \log \mathrm{tr}(A_{k_j}) \\ &= \mathrm{tr}(\Pi_n \log \Pi_n) + n \sum_{1 \leqslant k \leqslant r} \mathrm{tr}(A_k \rho) \log \mathrm{tr}(A_k), \end{aligned}$$

and the claim follows.

E. Main theorem

The theorem below is the main result of this section. It can be regarded as a quantum version of the Shannon–McMillan theorem.

Theorem V.4: *Let $(\mathsf{A}, \tau, \alpha)$ be an ergodic source, and let* **A** *be a POM, for which the operators $\{A_j\}$, $1 \leqslant j \leqslant r$, are orthogonal projections. Let $h_{\mathbf{A}}$ be the Shannon entropy of the associated classical source, and define*

$$M = \max_{1 \leqslant j \leqslant r} \mathrm{rank}(A_j), \quad m = \min_{1 \leqslant j \leqslant r} \mathrm{rank}(A_j).$$

Then, given $\delta,\epsilon>0$, *there is* n_0, *such that for all* $n\geqslant n_0$,

$$\mathscr{H}^{\otimes n}=\mathscr{S}_n\oplus\mathscr{S}_n^{\perp},$$

where $\mathscr{S}_n$ *is a subspace whose dimension satisfies*

$$\frac{\log m-\delta}{\log d}\leqslant\frac{\log \dim \mathscr{S}_n}{\log \dim \mathscr{H}^{\otimes n}}-\frac{h_{\mathbf{A}}}{\log d}\leqslant\frac{\log M+\delta}{\log d}. \tag{V.10}$$

Further, let $P_{\mathscr{S}_n}$ *be the orthogonal projection onto* $\mathscr{S}_n$. *Then for any observable* $C\in\mathscr{L}(\mathscr{H}^{\otimes n})$,

$$|\tau(CP_{\mathscr{S}_n})-\tau(C)|<\epsilon\|C\|. \tag{V.11}$$

Remark 1: We can regard $\mathscr{S}_n$ as a significant subspace of $\mathscr{H}^{\otimes n}$ in the sense that the expectation of any observable is almost completely determined by its restriction to $\mathscr{S}_n$. If the states $\{|\psi_j\rangle\}$ are orthogonal, we can take the d orthogonal projections $\{|\psi_j\rangle\langle\psi_j|\}$ for the POM. In this case the entropy $h_{\mathbf{A}}$ is equal to the von Neumann entropy, since the density operators all commute. Then there is a direct correspondence with the classical Shannon–McMillan theorem, and the quantum theory is just a restatement of the classical result.

Remark 2: In the Bernoulli case we can take the POM $\mathbf{A}$ to be d orthogonal projections onto the eigenvectors of ρ, in which case the inequality of Theorem V.2 is saturated. If ρ has a simple spectrum, this means that the Shannon entropy equals the von Neumann entropy of the quantum source. Our result then agrees with Schumacher's conclusion that the information contained in the quantum source resides in a subspace whose dimension is asymptotically $e^{nh_{\mathbf{A}}}$. In the general case we obtain only an upper bound for the dimension of the relevant subspace, and this upper bound depends on the choice of POM. For example, if each operator A_j in the POM $\mathbf{A}$ is equal to $(1/d)I$, where I is the identity, then the Shannon entropy is $h_{\mathbf{A}}=\log d$. Since this is the maximum possible entropy for a POM with d operators, we can conclude that all information about the quantum source has been lost in this measurement process. As these results show, it is advantageous to use a POM composed of orthogonal projections.

Proof: we use the POM to construct the classical source $\mathscr{X}_{\mathbf{A}}^{\infty}$ with entropy $h_{\mathbf{A}}$. Let $f_n(\underline{x})$ be the empirical entropy of a message $\underline{x}$ of length n defined in (V.1). Given $\epsilon,\delta>0$, let us define the sets

$$U_{n,\delta}=\{\underline{x}\in\mathscr{X}:|f_n(\underline{x})-h_{\mathbf{A}}|>\delta\},$$

$$L_{n,\delta}=\{\underline{x}\in\mathscr{X}:|f_n(\underline{x})-h_{\mathbf{A}}|\leqslant\delta\}.$$

By the Shannon–McMillan theorem, given $\epsilon,\delta>0$, there is n_0, such that for all $n\geqslant n_0$,

$$\sum_{\mathbf{x}\in U_{n,\delta}}\mu_{\mathbf{A}}(\{\underline{x}\})<\epsilon.$$

Since the operators $\{A_j\}$ are orthogonal projections, each tensor product $A_{k_1}\otimes\cdots\otimes A_{k_n}$ is an orthogonal projection, and hence so is the sum of these operators over the set $L_{n,\delta}$. Let $\mathscr{S}_n$ denote the range of this projection, and let $P_{\mathscr{S}_n}$ denote the orthogonal projection onto this subspace. Then for any observable C, we have

$$|\tau(CP_{\mathscr{S}_n})-\tau(C)|\leqslant\epsilon\|C\|.$$

It remains to estimate the dimension of $\mathscr{S}_n$. Since the projections $\{A_j\}$ are orthogonal, its dimension is given by

$$\dim(\mathscr{S}_n)=\sum_{\mathbf{x}\in L_{n,\delta}}\prod_{j=1}^{n}\operatorname{rank}(A_{x_j}).$$

Therefore,

$$m^n|L_{n,\delta}|\leqslant\dim(\mathscr{S}_n)\leqslant M^n|L_{n,\delta}|,$$

where $|L_{n,\delta}|$ is the size of the set $L_{n,\delta}$. The Shannon–McMillan theorem implies that

$$(1-\epsilon)e^{n(h_{\mathbf{A}}+\delta)}\geqslant|L_{n,\delta}|\geqslant e^{n(h_{\mathbf{A}}-\delta)}.$$

This leads to

$$\frac{\log m}{\log d}-\frac{\delta}{\log d}\leqslant\frac{\dim(\mathscr{S}_n)}{n\log d}-\frac{h_{\mathbf{A}}}{\log d}\leqslant\frac{\log M}{\log d}+\frac{\delta}{\log d}+\frac{\log(1-\epsilon)}{n},$$

and the result follows.

ACKNOWLEDGMENTS

Supported in part by the National Science Foundation under Grant No. DMS-9424344. We would like to thank Evelyn Wright for helpful discussions, and the referee for useful suggestions.

APPENDIX: PROOF OF POSITIVITY

We present here the proof that the R matrix (II.11) defines a family of positive density matrices, when $|b|$ and $|c|$ are sufficiently small. It is convenient to introduce the matrices

$$\omega=\frac{a}{2}I-\frac{1}{2}\sigma_3,$$

$$Q=\rho^{-1}(b\sigma_1+c\sigma_2).$$

Note that we assume $|a|<1$, so ρ^{-1} exists. Then for all $n\geqslant 2$ we define

$$S_n=I_{n-2}\otimes\omega\otimes Q.$$

It follows by direct calculation that for all $n\geqslant 1$,

$$\Pi_{n+1}=\Pi_n\otimes\rho+\tfrac{1}{2}(\Pi_n\otimes\rho)S_{n+1}+\tfrac{1}{2}S_{n+1}^{\dagger}(\Pi_n\otimes\rho).$$

In order to proceed we make the inductive assumption that $\Pi_n>0$; this implies, in particular that $(\Pi_n+u)^{-1}$ is bounded for every $u\geqslant 0$. We will prove that $(\Pi_{n+1}+u)^{-1}$ is bounded for every $u\geqslant 0$; together with the positivity of $\Pi_1=\rho$, this will establish the desired result.

Note first that $\|S_n\|\leqslant\|\omega\|\|Q\|$, and this bound is uniform in n. For convenience we define

$$w=\|\omega\|,\quad q=\|Q\|.$$

Furthermore, for any $u\geqslant 0$,

$$(\Pi_{n+1}+u)^{-1}=(\Pi_n\otimes\rho+u)^{-1}(I+\tfrac{1}{2}(\Pi_n\otimes\rho)S_{n+1}(\Pi_n\otimes\rho+u)^{-1}$$
$$+\tfrac{1}{2}S_{n+1}^{\dagger}(\Pi_n\otimes\rho)(\Pi_n\otimes\rho+u)^{-1})^{-1}.$$

Our inductive assumption implies that $(\Pi_n\otimes\rho+u)^{-1}$ is bounded for $u\geqslant 0$. Choosing $|b|$, $|c|$ sufficiently small guarantees that $\|S_n^{\dagger}\|<\epsilon$ for any $\epsilon>0$. Furthermore,

$$(\Pi_n\otimes\rho)S_{n+1}(\Pi_n\otimes\rho+u)^{-1}=([\Pi_n(I_{n-1}\otimes\omega)\Pi_n^{-1}]\otimes Q^{\dagger})(\Pi_n\otimes\rho)(\Pi_n\otimes\rho+u)^{-1}.$$

Since $\|Q^{\dagger}\|=q$ can be made arbitrarily small by choosing $|b|$, $|c|$ sufficiently small, the boundedness of $(\Pi_{n+1}+u)^{-1}$ will follow from a bound for the operator $\Pi_n(I_{n-1}\otimes\omega)\Pi_n^{-1}$, which is uniform in n. Accordingly let us define for $n\geqslant 1$,

$$A_n=\Pi_n(I_{n-1}\otimes\omega)\Pi_n^{-1}.$$

By imitating the derivation above, we obtain the recursion relation

$$A_n=(I+\tfrac{1}{2}A_{n-1}\otimes Q^{\dagger}+\tfrac{1}{2}S_n^{\dagger})(I_{n-1}\otimes\omega)(I+\tfrac{1}{2}A_{n-1}\otimes Q^{\dagger}+\tfrac{1}{2}S_n^{\dagger})^{-1}. \tag{A.1}$$

It is immediate that $\|A_1\|=\|\omega\|\leqslant 1$. We make the inductive assumption that $\|A_{n-1}\|\leqslant 1$; then (A1) implies the estimate

$$\|A_n\|\leqslant\left(1+\frac{q}{2}+\frac{wq}{2}\right)w\left(1-\frac{q}{2}-\frac{wq}{2}\right)^{-1}. \tag{A.2}$$

If we choose

$$q<2\frac{1-w}{(1+w)^2}, \tag{A.3}$$

then (A2) implies that $\|A_n\|\leqslant 1$. Hence by choosing (A3) we obtain that A_n is uniformly bounded for all n, and hence that $(\Pi_{n+1}+u)^{-1}$ is bounded for all $u\geqslant 0$. Therefore the positivity of the density matrices is proved.

[1] C. H. Bennett, ''Quantum information and computation,'' Phys. Today **48**, 24–30 (1995).
[2] B. Schumacher, ''Quantum coding,'' Phys. Rev. A **51**, 2738–2747 (1995).
[3] R. Jozsa and B. Schumacher, ''A new proof of the quantum noiseless coding theorem,'' J. Mod. Opt. **41**, 2343–2349 (1994).
[4] R. B. Ash, *Information Theory* (Dover, New York, 1990).
[5] P. Billingsley, *Ergodic Theory and Information* (Krieger, Huntington, NY, 1978).
[6] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).
[7] L. P. Hughston, R. Jozsa, and W. Wootters, ''A complete classification of quantum ensembles having a given density matrix,'' Phys. Lett. A **183**, 14–18 (1993).
[8] R. Haag, *Local Quantum Physics* (Springer-Verlag, Berlin, 1992).
[9] O. Bratteli and D. W. Robinson *Operator Algebras and Quantum Statistical Mechanics* (Springer-Verlag, Berlin, 1981), Vols. 1 and 2.
[10] I. P. Cornfeld, S. V. Fomin, and Ya. Sinai, *Ergodic Theory* (Springer-Verlag, Berlin, 1982).
[11] O. E. Lanford and D. W. Robinson, ''Mean entropy of states in quantum-statistical systems,'' J. Math. Phys. **9**, 1120–1125 (1968).
[12] A. Wehrl, ''General properties of entropy,'' Rev. Mod. Phys. **50**, 221–260 (1978).
[13] O. Klein, Z. Phys. **72**, 767–775 (1931).