SCIENTIFIC REPORTS

OPEN

Lossless quantum data compression with exponential penalization: an operational interpretation of the quantum Rényi entropy

Guido Bellomo[1], Gustavo M. Bosyk[2], Federico Holik[2] & Steeve Zozor[3]

Based on the problem of quantum data compression in a lossless way, we present here an operational interpretation for the family of quantum Rényi entropies. In order to do this, we appeal to a very general quantum encoding scheme that satisfies a quantum version of the Kraft-McMillan inequality. Then, in the standard situation, where one is intended to minimize the usual average length of the quantum codewords, we recover the known results, namely that the von Neumann entropy of the source bounds the average length of the optimal codes. Otherwise, we show that by invoking an exponential average length, related to an exponential penalization over large codewords, the quantum Rényi entropies arise as the natural quantities relating the optimal encoding schemes with the source description, playing an analogous role to that of von Neumann entropy.

One of the main concerns in classical and quantum information theory is the problem of encoding information by using fewest resources as possible. This task is known as *data compression* and it can be carried out either in a *lossy* or a *lossless* way, depending on whether the original data can be recovered with or without errors, respectively.

Here, we are interested in lossless *quantum* data compression. In order to state our proposal, let us first recall how this task works in the classical domain. The mathematical foundations of classical data compression can be found in the seminal paper of Shannon[1] (see e.g.[2] for an introduction to the topic), although we can summarize it as follows. Let $S = \{p_i, s_i\}$ be a classical source where each symbol s_i has associated a probability of occurrence p_i. The idea is to assign to each symbol a codeword $c(s_i)$ of some alphabet $A = \{0, \ldots, k-1\}$ in an adequate way. In particular, a k-ary classical code c of S is said *uniquely decodable* if this assignment of codewords is injective for any possible concatenation. A celebrated result states that any uniquely decodable code necessarily satisfies the *Kraft-McMillan inequality*[3,4]: $\sum_i k^{-\ell_i} \leq 1$ where ℓ_i is the length of the codeword $c(s_i)$ (measured in bits if $k = 2$). Conversely, given a set of codewords lengths $\{\ell_i\}$, there exists a uniquely decodable code with these lengths. Thus, lossless data compression consists in finding a uniquely decodable code taking into account the statistical description of the source. Formally, this is carried out by minimizing the average codeword length $L = \sum_i p_i \ell_i$ subject to the Kraft-McMillan inequality. In the end, one obtains a *variable-length* code where shorter codewords are assigned to symbols with a high probability of occurrence, whereas larger codewords are assigned to symbols with low probability (see[2], chap. 5). Moreover, one has that (in the limit of the large number of independent and identically-distributed sources) the average length of the optimal code is arbitrarily close to the *Shannon entropy*[1] of the source, $H(p) = -\sum_i p_i \log_k p_i$.

As noticed by Campbell, the previous solution has the disadvantage that it can happen that the codeword length turns out to be very large for symbols with a sufficiently low probability of occurrence[5]. Indeed, the use of

[1]CONICET-Universidad de Buenos Aires, Instituto de Investigación en Ciencias de la Computación (ICC), Buenos Aires, Argentina. [2]Instituto de Física La Plata, UNLP, CONICET, Facultad de Ciencias Exactas, Casilla de Correo 67, 1900 La Plata, La Plata, Argentina. [3]Univ. Grenoble Alpes, CNRS, Grenoble INP Institute of Engineering, GIPSA-Lab, 38000, Grenoble, France. Correspondence and requests for materials should be addressed to G.M.B. (email: gbosyk@fisica.unlp.edu.ar)

average codewords length as a criterion of performance has the implicit assumption that the cost varies linearly with the codeword length, which is not always desirable. For instance, it could be the case that adding a letter to a large codeword may have a larger impact than adding a letter to a shorter codeword, for instance in terms of memory needed to store a codeword. This problem has given place to the proposal of several other measures of codeword lengths (see e.g.[6–9]), for which the average length is a limiting case. In particular, a generalized average t-length, also called exponential average, is defined as[5] $L_t = \frac{1}{t} \sum_i \log_k \sum_i p_i k^{t \ell_i}$, where $t \geq 0$ is a parameter related to the cost. Notice that in the limiting case $t \to 0$ one recovers $L_t \to L$ and, as t increases, a greater *penalization* over the large codewords holds. Indeed, Campbell has obtained a source coding theorem taking into account such a penalization. His theorem is similar to the standard one, but the encoding is made in such a way that the generalized codeword length turns out to be arbitrarily close to the *Rényi entropy*[10] of the source, $H_\alpha(p) = \frac{1}{1-\alpha} \log_k \sum_i p_i^\alpha$ with $\alpha = \frac{1}{t+1}$. This remarkable result provides an operational interpretation of the Rényi entropy as the natural information measure for the problem of optimal data compression with penalization over large codewords (see also[11] for a discussion of an axiomatic derivation of entropy related to the coding problem).

As we have seen, variable-length codes arise naturally in the problem of lossless classical data compression. In the quantum information theory realm, the formulation of this problem presents intrinsic difficulties. These difficulties are mainly related to the fact that a quantum source can possibly send mutually non-orthogonal states. Thus, one has to deal with superpositions of quantum codewords. Even worse, these superpositions may correspond to codewords of different lengths. Schumacher and Westmoreland were the first in establishing a general approach to the problem of quantum variable-length coding[12]. Furthermore, they have provided the first quantum version of the Kraft-McMillan inequality and have found that the von Neumann entropy of the source ρ, $S(\rho) = -\mathrm{Tr}(\rho \log_2 \rho)$ (binary logarithm for coding in qubits), plays an analogous role to that of the Shannon entropy in the classical source coding theorem. Several other authors have contributed to this subject proposing alternative or extended schemes[12–20]. In general, these approaches face the same disadvantage as in the classical case: namely they do not consider the fact that large codewords, even appearing with low probabilities, may have large impact in terms of resources needed for the encoding. This drawback is even more relevant nowadays, due to the fact that the practical implementation of quantum information protocols pose the challenge of manipulating coherent superpositions of qubits. While the use of chains of qubits of arbitrary length may arise naturally in some theoretical considerations, it can be very expensive and difficult to implement large chains in the lab, specially at the early stages of the development of quantum information technology devices. Thus, our goal is to provide a quantum version of Campbell's strategy for the problem of coding with penalization of large codewords. As a consequence, we show that in this framework the quantum Rényi entropies emerge as the natural quantities relating the optimal encoding schemes with the source description. Accordingly, we provide an operational interpretation for those entropies.

Results

Uniquely decodable quantum code and quantum Kraft-McMillan inequality.
In this section, we summarize some definitions and results of the literature related to our proposal. We begin by pointing out the problem of *lossless quantum compression*.

Lossless quantum compression consists in compressing a quantum source given by an ensemble of quantum states, by using a variable-length quantum code so that the original states can be exactly recovered, *i.e.*, without error. More precisely, the situation to deal with is the following. Let us assume that a quantum source produces an ensemble of quantum states $\mathcal{S} = \{p_n, |s_n\rangle\}_{n=1}^N$, where $p_n \geq 0$, $\sum_{n=1}^N p_n = 1$ and $|s_n\rangle \in \mathcal{H}_\mathcal{S} \equiv \mathbb{C}^d$. The first task is to encode in an unambiguous or uniquely decodable way not only every single quantum state $|s_n\rangle$ of the source, but also any string of quantum states of the source. In this sense, let us first introduce a very general definition of a *uniquely decodable quantum source code*.

Definition 1. A uniquely decodable quantum source code of $\mathcal{S}$ over a quantum k-ary alphabet $\mathcal{A} = \{|0\rangle, \ldots, |k-1\rangle\} \subset \mathcal{H}_\mathcal{A} \equiv \mathbb{C}^k$, with $k \in \mathbb{N}^* \backslash \{1\}$, is a linear isometry map $U : \mathcal{F}_\mathcal{S} \to \mathcal{F}_\mathcal{A}$ where $\mathcal{F}_\mathcal{X} \equiv \oplus_{\ell=0}^\infty \mathcal{H}_\mathcal{X}^{\otimes \ell}$ is a Fock space, where $\mathcal{X} = \mathcal{S}$ or $\mathcal{A}$.

In this way, the fact that U is an isometry guarantees an injective mapping which assigns for each string of the form $\otimes_{m=1}^M |s_{i_m}\rangle$, with $i_m \in \{1, \ldots, N\}$ and $M \in \mathbb{N}^*$, a quantum codeword $U \otimes_{m=1}^M |s_{i_m}\rangle \in \mathcal{F}_\mathcal{A}$. Let us see how our definition works for single code words. A single quantum codeword over $\mathcal{A}$ is a quantum pure state that belongs to the Fock space $\mathcal{F}_\mathcal{A}$ (we are taking here strings with a single component). Thus, we can write $U|s_n\rangle = \sum_j a_{j,n} |a_{j,n}\rangle$, where $a_{j,n} \in \mathbb{C}$, $\sum_j |a_{j,n}|^2 = 1$ and $|a_{j,n}\rangle \in \mathcal{H}_\mathcal{A}^{\otimes l_j} \subset \mathcal{F}_\mathcal{A}$. Notice that the number of non-vanishing coefficients in the set $\{a_{j,n}\}_{j=1}^\infty$ could be infinite in principle. In the following we will restrict to the finite case (*i.e.*, $a_{j,n} = 0$ for almost all j).

Up to now, we have given a very formal definition of uniquely decodable quantum source code. In order to show an encoding scheme that satisfies definition 1, we mainly follow the proposal given in[19,20]. First, let us precise the definition of a k-ary classical uniquely decodable code for the symbols source $S = \{1, \ldots, d\}$ over an alphabet $A = \{0, \ldots, k-1\}$. Let F_A be the set $F_A = \cup_{\ell=0}^\infty A^\ell$. Then, $c : S \to F_A$ is a classical uniquely decodable code if and only if for any $M \geq 1$, any concatenation $c^M(i_1, \ldots, i_M) = c(i_1) \cdots c(i_M)$ of M codewords is an injective function (see e.g.[2]). We denote by ℓ_i the length of the i-th codeword, *i.e.*, the number of "letters" of A appearing in the codeword $c(i)$. Hereafter, we consider the isometries $U : \mathcal{H}_\mathcal{S} \to \mathcal{F}_\mathcal{A}$ of the form[19,20],

$$U = \sum_{i=1}^d |c(i)\rangle \langle e_i|,$$

(1)

where $\{|e_i\rangle\}_{i=1}^d$ is a basis of $\mathcal{H}_S$ and c a classical uniquely decodable code of S. Clearly, by construction, one has $|c(i)\rangle c(i) \in \mathcal{H}_\mathcal{A}^{\otimes \ell_i} \subset \mathcal{F}_\mathcal{A}$ and $\{|c(i)\rangle\}$ forms an orthonormal set, so that $U^\dagger U = I$ (but notice that, in general, $UU^\dagger$ can be different from the identity operator). We refer any isometry of the form (1) as *lossless quantum encoding scheme*. Note that contrary to a classical code, $|c(i)\rangle$ here does not encode any quantum state of the source $\{|s_n\rangle\}$ but the base state $|e_i\rangle$, except when $|s_n\rangle = |e_i\rangle$ for some n, i. As introduced, the codeword associated to a superposition of source states is the superposition of the codewords. Moreover, $U|s_n\rangle$ does not necessarily belong to a space of the form $\mathcal{H}_\mathcal{A}^{\otimes \ell}$ for some ℓ. Notice now that a quantum coding scheme U can be extended to a map $U^M : \mathcal{H}_S^{\otimes M} \to \mathcal{F}_\mathcal{A}$ on sentences a follows:

$$U^M = \sum_{i_1=1}^d \cdots \sum_{i_M=1}^d |c(i_1)\cdots c(i_M)\rangle \langle e_{(i_1)}\cdots e_{(i_M)}|. \tag{2}$$

The above map is well defined for all $M \in \mathbb{N}^*$. A map such as U^M can be naturally considered as an operator acting in the Fock space $\mathcal{F}_S$ by viewing $|e_{i_1}\cdots e_{i_M}\rangle \in \mathcal{H}_S^{\otimes M} \subset \mathcal{F}_S$ as follows. Consider a state $|\phi\rangle \in \mathcal{H}_S^{\otimes M'}$. Then, we write $U^M|\phi\rangle = \delta_{M,M'}\sum_{i_1=1}^d \cdots \sum_{i_M=1}^d |e_{i_1}\cdots e_{i_M}|\phi\rangle|c(i_1)\cdots c(i_M)\rangle$. Now, with this observation we can define an operator $U^\infty : \mathcal{F}_S \to \mathcal{F}_\mathcal{A}$ as

$$U^\infty = \sum_{M=1}^\infty U^M. \tag{3}$$

The physical interpretation of U^∞ is that for each sentence $\otimes_{m=1}^M |s_{i_m}\rangle$ of the source, we will obtain the right coded sentence for each $M \in \mathbb{N}^*$. It is important to remark that all these coding schemes are lossless in the sense of definition 1.

As it is well known in classical data compression, the Kraft-McMillan inequality gives a necessary and sufficient condition for the existence of a uniquely decodable code (see *e.g.*[2],). This result has been originally extended to the quantum domain in[12], introducing a particular formalism. We proceed here to obtain a *quantum version of the Kraft-McMillan inequality*, compatible with the previous construction.

Let us first introduce the *length observable*, which allows to get a further notion of codeword length.

Definition 2. The length observable Λ acting on $\mathcal{F}_\mathcal{A}$ is defined as

$$\Lambda \equiv \sum_{\ell=0}^\infty \ell \Pi_\ell, \tag{4}$$

where Π_ℓ denotes the orthogonal projector onto the subspace $\mathcal{H}_\mathcal{A}^{\otimes \ell} \subset \mathcal{F}_\mathcal{A}$.

Now, the quantum Kraft-McMillan inequality reads as follows.

Theorem 1. For any losless quantum encoding scheme U given by Eq. (1), the following inequality must be satisfied:

$$\mathrm{Tr}(U^\dagger k^{-\Lambda} U) \leq 1. \tag{5}$$

The proof of this theorem, which mainly relies in its classical counterpart, is given in the section Methods, along with the proofs of the subsequent theorems.

Source coding and von Neumann entropy bounds.   As in the classical case, we are interested in quantum codes that minimize the amount of resources involved. However, in the quantum case arises an extra difficulty to quantify the number of resources since there is no a unique way of defining the notion of length of a quantum codeword. For a given encoding scheme U, the standard definition of *quantum codeword length* is the following.

Definition 3. The quantum codeword length of $|\omega\rangle \equiv U|s\rangle$ for some $|s\rangle \in \mathcal{H}_S$ is given by the expectation value

$$\ell(|\omega\rangle) \equiv \langle \omega|\Lambda|\omega\rangle = \sum_{i=1}^d |\langle e_i|s\rangle|^2 \ell_i. \tag{6}$$

Thus, from this definition, the codewords may not have definite length in the sense that they are not eigenstates of the length operator in the general case. For that reason a quantum code given by the encoding scheme (1) is sometimes called *quantum indeterminate-length code*[12].

As we have noticed, one can introduce another important measure of the length of a quantum codeword. One used in the literature is the *base length*[13]:

Definition 4. The base length of a quantum codeword $|\omega\rangle \equiv U|s\rangle$ is given by

$$l(|\omega\rangle) \equiv \max\{\ell \in \mathbb{N} \langle \omega|\Pi_\ell|\omega\rangle \neq 0\} \quad = \max_{\{i \in \{1,\dots,d\}|e_i|s \neq 0\}} \{\ell_i\}. \tag{7}$$

Notice that the base length plays a key role as it determines the minimum size of the quantum register necessary to store a quantum codeword.

The base length of a quantum codeword is an integer whereas the quantum codeword length is not, in general. However, there is a relation between both lengths given by $\ell(|\omega\rangle) = \sum_{\ell=0}^{l(\omega)} \ell\langle\omega|\Pi_\ell|\omega\rangle \leq l(|\omega\rangle)\sum_\ell\langle\omega|\Pi_\ell|\omega\rangle$. Immediately, one has $\ell(|\omega\rangle) \leq l(|\omega\rangle)$, with equality if and only if $|\omega\rangle = U|s\rangle$ is an eigenstate of Λ, *i.e.*, if $|s\rangle$ is an eigenstate of U.

Henceforth, we consider that the state of the quantum source $\mathcal{S}$ is given by the density operator ρ, *i.e.*, a positive semi-definite operator of trace one acting on $\mathbb{C}^d$. We will write the density operator using the decomposition on ensemble's states, *i.e.*, $\rho = \sum_{n=1}^N p_n|s_n\rangle\langle s_n|$, or equivalently, considering the spectral decomposition, *i.e.*, $\rho = \sum_{i=1}^d \rho_i|\rho_i\rangle\langle\rho_i|$, where ρ_i is the eigenvalue corresponding to the eigenstate $|\rho_i\rangle$. In addition, we will denote as

$$C(\rho) \equiv U\rho U^\dagger = \sum_{i,i'=1}^d \langle e_i|\rho|e_{i'}\rangle|c(i)\rangle\langle c(i')| \tag{8}$$

the output of the quantum encoder (1). Then, according to definition 3, the *average codeword length* of $\mathcal{S}$ is given by

$$\ell(C(\rho)) \equiv \mathrm{Tr}(C(\rho)\Lambda) = \sum_{n=1}^N p_n \sum_{i=1}^d |\langle e_i|s_n\rangle|^2 \ell_i. \tag{9}$$

On the other hand, according to definition 4, the *base length* of $\mathcal{S}$ is

$$l(C(\rho)) \equiv \max\{l(U|s_n\rangle)\}_{n=1}^N = \max\left\{\max_{\{i\in\{1,\dots,d\}\,|\,\langle e_i|s_n\rangle\neq 0\}}\{\ell_i\}\right\}_{n=1}^N. \tag{10}$$

We have now all the ingredients to introduce *optimal quantum lossless codes*.

Definition 5. A quantum encoding scheme U is optimal for the source $\mathcal{S}$ if it minimizes the average codeword length, that is,

$$U^{\mathrm{opt}} \equiv \operatorname*{argmin}_{\mathrm{Tr}(U^\dagger k^{-\Lambda}U)\leq 1} \mathrm{Tr}(U\rho U^\dagger\Lambda) \tag{11}$$

and thus the minimal average codeword length for the source $\mathcal{S}$ is given by

$$\ell(C^{\mathrm{opt}}(\rho)) = \mathrm{Tr}(C^{\mathrm{opt}}(\rho)\Lambda), \tag{12}$$

where $C^{\mathrm{opt}}(\rho) \equiv U^{\mathrm{opt}}\rho U^{\mathrm{opt}\dagger}$.

In the classical setting to search for the optimal code, one has to find for the set of integers $\{\ell_i\}$ that minimizes the averaged length subjected to the Kraft-McMillan inequality. It is well known that *Huffman code* provides the optimal solution[21]. Let us see that the quantum optimal code or the quantum version of Huffman code is obtained for an encoding scheme U with basis given by the eigenstates of ρ and the classical code c given by the Huffman code for the symbols $\{1, \dots, d\}$ with probabilities given by the eigenvalues of ρ.

Theorem 2. *The optimal quantum code of the quantum source $\mathcal{S}$ writes*

$$U^{\mathrm{opt}} = \sum_{i=1}^d |c^{\mathrm{opt}}(i)\rangle\langle\rho_i|, \tag{13}$$

where $\{c^{\mathrm{opt}}(i)\}$ is the classical optimal code given by the Huffman code[21] of the symbols $\{1, \dots, d\}$ with corresponding probabilities $\{\rho_1, \dots, \rho_d\}$.

Let us recall that there is no an analytic formula for the individual lengths ℓ_i of the classical Huffman code in the general case. On the other hand, if one drops the integer restriction of $\{\ell_i\}$ in the minimization problem, one obtains the optimum "lengths" $-\log_k\rho_i$. To take integer values, one can consider the excess integer part of these values, $l_i = \lceil-\log_k\rho_i\rceil$, and construct a corresponding code using the Kraft tree (see[2] for more details). This is a well known method called the *Shannon coding* for which the average length is close to the optimal one (which is given by the Huffman code). Accordingly, we can say that the quantum version of the Shannon code is given by an encoding scheme of the form (13), where the classical code c is now given by the Shannon code. Nevertheless, without explicitly expressing the optimal code, it is possible to upper and lower bound the optimal average codeword length in terms of the von Neumann entropy of the source, as previously proved in a different formalism in[12].

Theorem 3. *The average length of the optimal code is lower and upper bounded as follows*

$$S(\rho) \leq \ell(C^{\mathrm{opt}}(\rho)) < S(\rho) + 1, \tag{14}$$

where $S(\rho) = -\mathrm{Tr}(\rho\log_k\rho)$ is the von Neumann entropy of a density operator ρ, and $\log_k$ is the logarithm of base k.

According to theorem 3, the entropy of the source bounds the compression capacity. Moreover, one can attain the lower bound for the case of K independent and identical preparations of the source for large K. Let $\rho^{\otimes K}$ be the corresponding density operator, and denote by $\frac{1}{K}\ell(C^{\mathrm{opt}}(\rho^{\otimes K}))$ the optimal average length code per

source, where $C(\rho^{\otimes K})$ is defined via the concatenation (2). Then, from $S(\rho^{\otimes K}) = KS(\rho)$ and theorem 3 one has $S(\rho) \leq \frac{1}{K}\ell(C^{\text{opt}}(\rho^{\otimes K})) < S(\rho) + \frac{1}{K}$, so that

$$\lim_{K\to\infty} \frac{1}{K}\ell(C^{\text{opt}}(\rho^{\otimes K})) = S(\rho). \tag{15}$$

We end this section discussing what happens to the average codeword length when the encoding scheme is designed for a "wrong" density operator τ instead of the correct one ρ. This could be useful for the case where τ is the best estimation of the state of the source for instance. In such a situation, the average code length of the quantum Shannon code corresponding to τ is again bounded, as follows (see, e.g., refs[19,20]).

Theorem 4. *Let τ be a density operator whose diagonal form is $\tau = \sum_{i=1}^{d}\tau_i|\tau_i\rangle\langle\tau_i|$. Let us consider the quantum Shannon code $U^{\text{Sh}} = \sum_{i=1}^{d}|c(i)\rangle\langle\tau_i|$ designed for τ, where $c(i)$ are classical codewords of the Shannon code, with lengths $\ell_i = \lceil -\log_k\tau_i\rceil$. The average length of such a quantum encoding is bounded as follows*

$$S(\rho) + S(\rho\|\tau) \leq \ell(C^{\text{Sh}}(\rho)) < S(\rho) + S(\rho\|\tau) + 1, \tag{16}$$

where $C^{\text{Sh}}(\rho) \equiv U^{\text{Sh}}\rho U^{\text{Sh}\dagger}$.

Notice that this gives an operational interpretation to the quantum relative entropy as follows: $S(\rho\|\tau)$ measures the deviation from the average codeword length of the quantum Shannon code, when the code is designed using a density operator which differs from density operator associated to the source (see also[22,23], for a further understanding of the role of quantum relative entropy in the context of data compression).

Source coding and quantum Rényi entropy bounds.

Let us first note that the definition 5 of optimal code and the results given above are closely linked to the standard definition 3 of the length of a quantum codeword. However, there could be problems for which the relevant measure of length is not the usual one. In this sense, Müller *et al.* have used the average of the base lengths of the source in order to define a different optimal code[18] and have obtained a complementary result to the one given by theorem 3. In this section we follow an alternative strategy, which is based on an extension of Campbell's proposal to the quantum case[5]. Let us first introduce a notion of *exponential* quantum codeword length. The standard quantum codeword and base lengths turn out to be particular cases of our definition.

Definition 6. The t-exponential length of a quantum codeword $|\omega\rangle \equiv U|s\rangle$ for some $|s\rangle \in \mathcal{H}_S$ is given by the expectation value

$$\ell_t(|\omega\rangle) \equiv \frac{1}{t}\log_k\langle\omega|k^{t\Lambda}|\omega\rangle = \frac{1}{t}\log_k\left(\sum_{i=1}^{d}\langle e_i|s\rangle^2 k^{t\ell_i}\right), \tag{17}$$

where $t \geq 0$ is a parameter related to the cost assigned to large codewords. In the limiting cases, one has

$$\ell_0(|\omega\rangle) \equiv \lim_{t\to 0}\ell_t(|\omega\rangle) = \ell(|\omega\rangle) \qquad \text{and} \qquad \ell_\infty(|\omega\rangle) \equiv \lim_{t\to\infty}\ell_t(|\omega\rangle) = l(|\omega\rangle). \tag{18}$$

Notice that $t \mapsto \ell_t(\omega)$ is a continuous nondecreasing function, i.e., $\ell_t(|\omega\rangle) \leq \ell_{t'}(|\omega\rangle)$ for $t \leq t'$. Thus, by changing the parameter t, one can move continuously and increasingly from the standard quantum codeword length to the base length. In other words, the *t*-exponential codeword length will allow to make a compromise between minimizing the average length and the base length. Finally, note that if $|\omega\rangle \in \mathbb{C}^\ell$, i.e., the quantum codeword is an eigenstate of the length observable, then $\ell_t(|\omega\rangle) = \ell$, which is a reasonable property for a quantum codeword length measure.

According to definition 6, the *t*-exponential average codeword length of the quantum source $\mathcal{S}$ is given by

$$\ell_t(C(\rho)) \equiv \frac{1}{t}\log_k\text{Tr}(C(\rho)k^{t\Lambda}) = \frac{1}{t}\log_k\left(\sum_{n=1}^{N}p_n\sum_{i=1}^{d}|\langle e_i|s_n\rangle|^2 k^{t\ell_i}\right). \tag{19}$$

We introduce now the notion of optimal quantum code corresponding to our previously defined *t*-exponential codeword length. A natural choice is as follows:

Definition 7. A quantum encoding scheme U is *t*-exponential optimal for the source $\mathcal{S}$ if it minimizes the *t*-exponential average codeword length, that is,

$$U_t^{\text{opt}} \equiv \underset{\text{Tr}(U^\dagger k^{-\Lambda}U)\leq 1}{\text{argmin}} \frac{1}{t}\log_k\text{Tr}(U\rho U^\dagger k^{t\Lambda}) \tag{20}$$

and thus the minimal *t*-exponential average codeword length for the source $\mathcal{S}$ is given by

$$\ell_t(C_t^{\text{opt}}(\rho)) = \frac{1}{t}\log_k\text{Tr}(C_t^{\text{opt}}(\rho)k^{t\Lambda}), \tag{21}$$

where $C_t^{\text{opt}}(\rho) \equiv U_t^{\text{opt}}\rho U_t^{\text{opt}\dagger}$.

In the classical setting to search for the t-exponential optimal code, as for the standard context, one has to look for the set of integers $\{\ell_i\}$ that minimizes the t-exponential averaged length subjected to the Kraft-McMillan inequality. This problem has been already solved in[7,24,25]. In the quantum context, we prove here that the optimal code is again obtained for an encoding scheme U with basis given by the eigenstates of ρ and the classical t-exponential optimal code c_t for the symbols $\{1, \ldots, d\}$ with probabilities given by the eigenvalues of ρ.

Theorem 5. *The quantum code that minimizes the t-exponential average codeword length of the quantum source $\mathcal{S}$ writes*

$$U_t^{\text{opt}} = \sum_{i=1}^{d} |c_t^{\text{opt}}(i)\rangle\langle\rho_i|, \tag{22}$$

where $\{c_t^{\text{opt}}(i)\}$ is the classical code minimizing the t-exponential average code length of the symbols $\{1, \ldots, d\}$ with corresponding probabilities $\{\rho_1, \ldots, \rho_d\}$.

As for the standard case, there is no an analytic formula for the individual optimal integer lengths ℓ_i leading to the minimum t-exponential average length of the classical code. But, again, if one drops the integer restriction of $\{\ell_i\}$ in the minimization problem, one obtains now the optimum "lengths" $-\log_k \rho_{t_i}$ where the ρ_{t_i} are the "escort probabilities", eigenvalues of the "escort" density operator

$$\rho_t \equiv \frac{\rho^{\frac{1}{1+t}}}{\text{Tr}\rho^{\frac{1}{1+t}}}, \tag{23}$$

acting on $\mathcal{H}_\mathcal{S}$. To take integer values, one can again consider the excess integer part of these values, $\ell_i = \lceil -\log_k \rho_{t_i} \rceil$, and construct a corresponding code using the Kraft tree, that is the Shannon code corresponding to the escort probabilities $\{\rho_{t_i}\}$. However, independently of the explicit expression of the generalized optimal code (20), it is possible to upper and lower bound the optimal t-exponential average quantum codeword length (21) in terms of the quantum Rényi entropy of the source.

Theorem 6. *The t-exponential average length of the t-exponential optimal code is lower and upper bounded as follows*

$$S_{\frac{1}{1+t}}(\rho) \leq \ell_t(C_t^{\text{opt}}(\rho)) < S_{\frac{1}{1+t}}(\rho) + 1, \tag{24}$$

where $S_\alpha(\rho) = \frac{1}{1-\alpha}\log_k \text{Tr}\rho^\alpha, \ \alpha \geq 0$, is the quantum Rényi entropy of the density operator of the source ρ.

We recall that our aim is to provide a scheme to address the problem of how to codify codewords of a quantum source allowing chains of variable length, but considering a penalization for large codewords. This aim can be achieved by appealing to definitions 6 and 7 and theorems 5 and 6. In particular, we can interpret theorem 6 as the quantum version of Campbell's source coding theorem[5]. Hence, the quantum Rényi entropy plays a role similar to that of von Neumann's in the standard quantum source coding theorem, when an exponential penalization is considered. Indeed, theorem 3 results as a particular case of our theorem 6 (with $t = 0$), recovering the results of Schumacher and Westmoreland[12]. This situation is completely analogous to that of the classical setting, with regard to the roles played by Rényi and Shannon measures for the cases with and without penalization, respectively. Consequently, this allows us to provide a natural operational interpretation for the quantum Rényi entropy in relation with the problem of lossless quantum data compression. Finally, notice that this is an alternative approach to that of Müeller *et al.*[18], where they have studied an analogous problem, but minimizing the average of the individual base lengths of the source instead of considering a penalization over large codewords.

According to theorem 6, the quantum Rényi entropy of the source bounds the compression capacity when an exponential penalization is considered. As in the case with no penalization, one can attain the lower bound for the case of K independent and identically prepared sources for large K. Thus, consider a density operator $\rho^{\otimes K}$ and denote by $\frac{1}{K}\ell_t(C_t^{\text{opt}}(\rho^{\otimes K}))$ to the t-exponential optimal average length code per source. Then, using that $S_\alpha(\rho^{\otimes K}) = KS_\alpha(\rho)$ and theorem 6, one has $S_{\frac{1}{1+t}}(\rho) \leq \frac{1}{K}\ell_t(C_t^{\text{opt}}(\rho^{\otimes K})) < S_{\frac{1}{1+t}}(\rho) + \frac{1}{K}$. In this way

$$\lim_{K\to\infty} \frac{1}{K}\ell_t(C_t^{\text{opt}}(\rho^{\otimes K})) = S_{\frac{1}{1+t}}(\rho). \tag{25}$$

Let us point out that the quantum Rényi entropy appears also naturally in the determination of the exponent of the average error of the quantum fixed-length source coding[26,27], which is closely related to the Chernoff exponent appearing in classical discrimination problems. This exponent provides thus another interpretation of the quantum Rényi entropy. Our approach differs in that we study the role played by the quantum Rényi entropy in the problem of lossless quantum data compression with penalization.

As in the end of the previous section, we now discuss what happens with the t-exponential average codeword length when the encoding scheme is designed for a density operator τ, *i.e.*, using the escort density operator $\tau_t \equiv \frac{\tau^{\frac{1}{1+t}}}{\text{Tr}\tau^{\frac{1}{1+t}}}$. In that case the t-exponential average codeword length of the quantum Shannon code corresponding to τ_t is again bounded as follows.

Theorem 7. *Let τ be a density operator whose diagonal form is $\tau = \sum_{i=1}^{d} \tau_i |\tau_i\rangle\langle\tau_i|$. Let us consider the quantum Shannon code $U_t^{\mathrm{Sh}} = \sum_{i=1}^{d} |c(i)\rangle\langle\tau_i|$ designed for the escort density operator τ_t, where $c(i)$ are classical codewords of the Shannon code, with lengths $\ell_i = \lceil -\log_k \tau_{t_i} \rceil$. The t-exponential average length of such a quantum encoding is bounded as follows*

$$S_{\frac{1}{1+t}}(\rho) + S_{1+t}(\rho_t \| \tau_t) \leq \ell_t(C_t^{\mathrm{Sh}}(\rho)) < S_{\frac{1}{1+t}}(\rho) + S_{1+t}(\rho_t \| \tau_t) + 1, \tag{26}$$

where $C_t^{\mathrm{Sh}}(\rho) \equiv U_t^{\mathrm{Sh}} \rho U_t^{\mathrm{Sh}\dagger}$.

It is important to remark that this theorem provides an operational interpretation for the quantum Rényi divergence as follows: $S_{1+t}(\rho_t \| \tau_t)$ quantifies the deviation from the t-exponential average codeword length of the quantum Shannon code, when the code is designed using an escort density operator which differs from density operator associated to the source.

It would be desirable to have some expression that indicates how the standard average and the base length of the t-exponential optimal code behave when an exponential penalization is considered. However, this is not possible as there is no an analytic formula for the individual codeword length in this case, in general. An interesting alternative is analyzing how the the standard average of the quantum Shannon code is affected by an exponential penalization.

Theorem 8. Let $U_t^{\mathrm{Sh}} = \sum_{i=1}^{d} |c(i)\rangle\langle\rho_i|$ be the quantum Shannon code designed for the escort density operator ρ_t, for which the classical codewords lengths are given by $\ell_i = \lceil -\log_k \rho_{t_i} \rceil$. The average length of this code is bounded as follows

$$\frac{1}{1+t}S(\rho) + \frac{t}{1+t}S_{\frac{1}{1+t}}(\rho) \leq \ell(C_t^{\mathrm{Sh}}(\rho)) < \frac{1}{1+t}S(\rho) + \frac{t}{1+t}S_{\frac{1}{1+t}}(\rho) + 1. \tag{27}$$

Notice that the bounds are basically a convex combination of the von Neumann entropy (related to the minimum average length) and the Rényi entropy (related to the minimal t-exponential average length) of the source. Since $S_{\frac{1}{1+t}}(\rho) \geq S(\rho)$, the average length $\ell(C_t^{\mathrm{Sh}}(\rho))$ increases with respect to t; in particular, $\ell(C_t^{\mathrm{Sh}}(\rho)) \geq \ell(C^{\mathrm{Sh}}(\rho))$. On the contrary, for the base length, one can see that when ρ_i is small enough, there exists a parameter t sufficiently large so that $l(C_t^{\mathrm{Sh}}(\rho)) < l(C^{\mathrm{Sh}}(\rho))$. In particular, the base length can be lessen up to $\lceil S_0(\rho) \rceil = \lceil \log_k \mathrm{rank}\rho \rceil$. So, there is a tradeoff between $\ell(C_t^{\mathrm{Sh}}(\rho))$ and $l(C_t^{\mathrm{Sh}}(\rho))$ with respect to t. The optimal choosing of the cost parameter depends on the particularities of the problem in question (e.g., the size of the quantum register, etc). Finally, notice that for the exceptional case that all $-\log_k \rho_{ti}$ are integers, the quantum Shannon code hence designed coincides with the t-exponential optimal code $\mathbb{C}_t^{\mathrm{opt}}$ of theorem 5 and the lower bound of (24) is achieved.

Discussion

We have addressed the problem of lossless quantum data compression. In particular, we have considered the case in which codification of large codewords is penalized. Our work can be regarded as a quantum version of Campbell's work[5].

First, we have provided an expression for the optimal code for the case with exponential penalization (theorem 5) in terms of its classical counterpart[7,24,25]. We have shown that this penalization affects the optimal code in such a way that the Rényi entropy of the source bounds the t-exponential average codeword length (theorem 6). As a corollary, in the limit of a large number of independent and identically prepared sources, we have found that the capacity of compression equals the Rényi entropy of the source. Thus, the quantum Rényi entropy acquires a natural operational interpretation. In addition, we have found that a wrong description of the source produces an excess term in the bound of the average codeword length, which is related to the quantum Rényi divergence (theorem 7). Given that we recover the results by Schumacher and Westmoreland[12] when penalization is negligible, our work can be seen as an generalization of theirs.

Finally, we have discussed how the average and base lengths of the quantum Shannon code behave in terms of the cost parameter, which is related to the penalization (theorem 8). Indeed, there is a tradeoff between these two quantities, in the sense that it is possible to reduce the base length, but with the side effect of increasing the average length and viceversa.

It is worth noticing that our approach provides an alternative to that of Müeller *et al.*[18], where they have studied an analogous problem, but minimizing the average of the individual base lengths of the source. Our results are complementary to theirs.

Methods

In this section, we give the proofs of all theorems.

Proof of theorem 1.

Proof. Notice first that $U^{\dagger}k^{-\Lambda}U = \sum_{i=1}^{d} k^{-\ell_i}|e_i\rangle\langle e_i|$ due to

$$\langle c(i)|\Pi_{\ell}|c(i')\rangle = \delta_{\ell,\ell_{i'}}\delta_{i,i'}, \tag{28}$$

where the ℓ_i are the lengths of the classical codewords $c(i)$. Then, one directly obtains $\text{Tr}(U^\dagger k^{-\Lambda} U) = \sum_{i=1}^d k^{-\ell_i}$. Given that the code c is uniquely decodable, the proof ends by appealing to the classical Kraft-McMillan inequality.

Proof of theorem 2.

Proof. Let us first notice that the quantum Kraft-McMillan constraint is independent of the basis $\{|e_i\rangle\}$ of a given lossless quantum encoding scheme $U = \sum_i |c(i)\rangle\langle e_i|$. Then, to prove the theorem one can do it in two steps. On the one hand, let us first fix a classical code c and minimize $\ell(C(\rho)) = \sum_{i,j}\ell_i \rho_j \langle e_i|\rho_j\rangle^2$ over the set of basis of $\mathbb{C}^d$. Let us introduce the doubly stochastic matrix D with entries $D_{i,j} \equiv |\langle e_i|\rho_j\rangle|^2$, *i.e.*, $D_{i,j} \geq 0$ and $\sum_j D_{i,j} = \sum^i D_{i,j} = 1$ for all i and j. So the minimization problem consists in minimizing $\overrightarrow{\ell}\ D\ \overrightarrow{\rho}^t$ over the set of doubly stochastic matrices, where $\overrightarrow{\ell} = [\ell_1 \ldots \ell_d]$ and $\overrightarrow{\rho} = [\rho_1 \ldots \rho_d]$. From the Birkhoff theorem[28,29], one can write $D = \sum_k \pi_k \Pi_k$ as a convex combination of permutations matrices Π_k. Thus, $\overrightarrow{\ell}\ D\ \overrightarrow{\rho}^t = \sum_k \pi_k \overrightarrow{\ell} \Pi_k \overrightarrow{\rho}^t \geq \overrightarrow{\ell} \Pi_{k'} \overrightarrow{\rho}^t$ for some k', so that $D = \Pi_{k'}$. Although one does not know such permutation, this implies that each element of $\{|e_i\rangle\}$ coincides with only one of $\{|\rho_j\rangle\}$. On the other hand, one can skip the search of $\Pi_{k'}$ since one has now to minimize the averaged length with respect to the set of lengths $\{\ell_i\}$ subject to the classical Kraft-McMillan inequality. Indeed, without loss of generality, the permutation can be incorporated in the lengths by replacing $\overrightarrow{\ell} \to \overrightarrow{\ell}\ \Pi_{k'}$. Therefore, one has $|e_i\rangle = |\rho_i\rangle$ and thus $\ell(C(\rho)) = \sum_i \ell_i \rho_i$ is the classical average length of the classical code c. Finally, one has to find the classical optimal code c, whose solution is well known in the literature given by the Huffman code[21]. $\square$

Proof of theorem 3.

Proof. Let us first introduce the density operator

$$\sigma \equiv \frac{U^\dagger k^{-\Lambda} U}{\beta} \qquad \text{with} \qquad \beta \equiv \text{Tr}(U^\dagger k^{-\Lambda} U) \tag{29}$$

acting on $\mathcal{H}_\mathcal{S}$. Let $C(\rho) = U\rho U^\dagger$ with U an arbitrary encoding scheme of the form (1). Then, noting that $\Lambda = -\log_k k^{-\Lambda}$, and thus that $U^\dagger \Lambda U = -\log_k(U^\dagger k^{-\Lambda} U)$, it is straightforward to show that

$$\ell(C(\rho)) = S(\rho) + S(\rho\|\sigma) - \log_k\beta, \tag{30}$$

where $S(\rho\|\sigma) = \text{Tr}[\rho(\log_k\rho - \log_k\sigma)]$ is the quantum relative entropy. The quantum relative entropy being definite positive, and from $\log_k\beta \leq 0$ due to the quantum Kraft-McMillan inequality, it follows that

$$\ell(C(\rho)) \geq S(\rho), \tag{31}$$

for any encoding scheme U, in particular for the optimum one.

In order to proof the upper bound, let us consider the quantum Shannon code $U^{\text{Sh}} = \sum_{i=1}^d |c(i)\rangle\langle\rho_i|$ of ρ, where the lengths of the codewords $\{c(i)\}$ are $\{\ell_i = \lceil -\log_k\rho_i\rceil\}$. Notice that this code satisfies the quantum Kraft-McMillan inequality (5) by construction. Then, from $\lceil -\log_k\rho_i\rceil < -\log_k\rho_i + 1$, we have

$$\ell(C^{\text{Sh}}(\rho)) = \sum_{i=1}^d \rho_i \lceil -\log_k\rho_i\rceil < S(\rho) + 1. \tag{32}$$

The upper bound in (14) immediately follows from this inequality and from $\ell(C^{\text{opt}}(\rho)) \leq \ell(C^{\text{Sh}}(\rho))$ (by definition of the optimal code). $\square$

Proof of theorem 4.

Proof. Let $U^{\text{Sh}} = \sum_{i=1}^d |c(i)\rangle\langle\tau_i|$ be the quantum Shannon code of τ. It is straightforward to show that $\ell(C^{\text{Sh}}(\rho)) = \sum_{i=1}^d \langle\tau_i|\rho|\tau_i\rangle\lceil -\log_k\tau_i\rceil$. The bounds result thus directly from $-\log_k\tau_i \leq \lceil -\log_k\tau_i\rceil < -\log_k\tau_i + 1$ and $-\sum_{i=1}^d \langle\tau_i|\rho|\tau_i\rangle\log_k\tau_i = -\text{Tr}(\rho\log_k\tau) = S(\rho) + S(\rho\|\tau)$. $\square$

Proof of theorem 5.

Proof. For a given lossless quantum encoding scheme $U = \sum_i |c(i)\rangle e_i\langle e_i|$ it is straightforward to see that $\ell_t(C(\rho)) = \frac{1}{t}\log_k(\sum_{i,j} k^{t\ell_i}\rho_j|\langle e_i|\rho_j\rangle|^2)$. Noting that minimizing $\ell_t(C(\rho))$ is equivalent to minimizing $\sum_{i,j} k^{t\ell_i}\rho_j|\langle e_i|\rho_j\rangle|^2$, the proof is the very same than that of theorem 2, where $\overrightarrow{\ell}$ is replaced by $[k^{t\ell_1} \ldots k^{t\ell_d}]$ and where the classical optimal code turns to be $\{c_t^{\text{opt}}(i)\}$. This last one can be computed by the algorithms proposed in[7,24,25]. $\square$

Proof of theorem 6.

Proof. The proof is similar to that of theorem 3. Let $\mathbb{C}(\rho) = U\rho U^{\dagger}$ with U an arbitrary encoding scheme of the form (1). Then, noting that $\rho U^{\dagger} k^{t\Lambda} U = \rho (U^{\dagger} k^{-\Lambda} U)^{-t} = \rho_t^{1+t} \sigma^{-t} \beta^t \left[\text{Tr}\left(\rho^{\frac{1}{1+t}} \right) \right]^{-(1+t)}$, where σ and β are defined in (29) and ρ_t in (23), we immediately obtain

$$\ell_t(C(\rho)) = S_{\frac{1}{1+t}}(\rho) + S_{1+t}(\rho_t \| \sigma) - \log_k \beta, \tag{33}$$

where $S_\alpha(\rho \| \sigma) = \frac{1}{\alpha - 1} \log_k \text{Tr} \rho^\alpha \sigma^{1-\alpha}$ is the quantum Rényi divergence (see *e.g.*[30]). The quantum Rényi divergence being definite positive, and from $\log_k \beta \leq 0$ due to the quantum Kraft-McMillan inequality, it follows that

$$\ell_t(C(\rho)) \geq S_{\frac{1}{1+t}}(\rho), \tag{34}$$

for any encoding scheme U, in particular for the optimal one.

In order to prove the upper bound, let us now consider the quantum Shannon code $U_t^{\text{Sh}} = \sum_{i=1}^d c(i) \langle \rho_i |$ of the escort density operator ρ_t, where the lengths of the codewords $\{c(i)\}$ are $\{\ell_i = \lceil -\log_k \rho_{t_i} \rceil\}$ being ρ_{t_i} the escort probabilities, eigenvalues of ρ_t. Notice that this code satisfies the quantum Kraft-McMillan inequality (5) by construction. Then, from $\lceil -\log_k \rho_{t_i} \rceil < -\log_k \rho_{t_i} + 1$, we have

$$\ell_t(C_t^{\text{Sh}}(\rho)) = \frac{1}{t} \log_k \left(\sum_{i=1}^d \rho_i k^{t \lceil -\log_k \rho_{t_i} \rceil} \right) < S_{\frac{1}{1+t}}(\rho) + 1. \tag{35}$$

Because $\ell_t(C_t^{\text{opt}}(\rho)) \leq \ell_t(C_t^{\text{Sh}}(\rho))$ by definition of the optimal code, the upper bound in (24) immediately follows from this inequality. $\quad\square$

Proof of theorem 7.

Proof. Let $U_t^{\text{Sh}} = \sum_{i=1}^d |c(i)\rangle \langle \tau_i |$ be the quantum Shannon code of the escort density operator τ_t. It is straightforward to show that $\ell_t(C_t^{\text{Sh}}(\rho)) = \frac{1}{t} \log_k(\sum_{i=1}^d \langle \tau_i | \rho | \tau_i \rangle k^{t \lceil -\log_k \tau_{t_i} \rceil})$. The bounds result thus directly from $-\log_k \tau_{t_i} \leq \lceil -\log_k \tau_{t_i} \rceil < -\log_k \tau_{t_i} + 1$ together with $\sum_{i=1}^d \langle \tau_i | \rho | \tau_i \rangle \tau_{t_i}^{-t} = \text{Tr}(\rho \tau_t^{-t}) = \left[\text{Tr}\left(\rho^{\frac{1}{1+t}} \right) \right]^{1+t} \text{Tr}(\rho_t^{1+t} \tau_t^{-t})$. $\quad\square$

Proof of theorem 8

Proof. Notice that for this code, $U_t^{\text{Sh}} = \sum_{i=1}^d |c(i)\rangle \langle \rho_i |$, the classical codewords lengths can be expressed as

$$\ell_i = \left\lceil -\log_k \rho_{t_i} \right\rceil = \left\lceil \frac{1}{1+t}(-\log_k \rho_i) + \frac{t}{1+t} S_{\frac{1}{1+t}}(\rho) \right\rceil. \tag{36}$$

Thus, we have that the average length is given by

$$\ell(C_t^{\text{Sh}}(\rho)) = \sum_{i=1}^d \rho_i \left\lceil -\log_k \rho_{t_i} \right\rceil = \left\lceil \frac{1}{1+t} S(\rho) + \frac{t}{1+t} S_{\frac{1}{1+t}}(\rho) \right\rceil, \tag{37}$$

so that the lower and upper bounds in (27) are directly obtained.

References

1. Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb00917.x (1948).
2. Cover, T. M. & Thomas, J. A. *Elements of information theory* (John Wiley & Sons, 2006).
3. Kraft, L. J. *A device for quantizing, grouping, and coding amplitude-modulated pulses*. Ph.D. thesis, Massachusetts Institute of Technology (1949).
4. McMillan, B. Two inequalities implied by unique decipherability. *IRE Transactions on Information Theory* **2**, 115–116, https://doi.org/10.1109/TIT.1956.1056818 (1956).
5. Campbell, L. A coding theorem and Rényi's entropy. Information and Control 8, 423–429, https://doi.org/10.1016/S0019-9958(65)90332-3, http://www.sciencedirect.com/science/article/pii/S0019995865903323 (1965).
6. Yamano, T. Information theory based on nonadditive information content. *Phys. Rev. E* **63**, 046105, https://doi.org/10.1103/PhysRevE.63.046105 (2001).
7. Baer, M. B. Source coding for quasiarithmetic penalties. *IEEE Transactions on Information Theory* **52**, 4380–4393, https://doi.org/10.1109/TIT.2006.881728 (2006).
8. Bercher, J.-F. Source coding with escort distributions and Rényi entropy bounds. *Physics Letters A* **373**, 3235–3238. //www.sciencedirect.com/science/article/pii/S037596010900838X, https://doi.org/10.1016/j.physleta.2009.07.015 (2009).
9. Chapeau-Blondeau, F., Delahaies, A. & Rousseau, D. Source coding with Tsallis entropy. *Electron. Lett.* **47**, 187–188, https://doi.org/10.1049/el.2010.2792 (2011).
10. Rényi, A. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 1, 547–561 (1961).
11. Campbell, L. L. Definition of entropy by means of a coding problem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **6**, 113–118 https://doi.org/10.1007/BF00537132 (1966).

12. Schumacher, B. & Westmoreland, M. D. Indeterminate-length quantum coding. *Phys. Rev. A* **64**, 042304, https://doi.org/10.1103/PhysRevA.64.042304 (2001).
13. Bostroem, K. & Felbinger, T. Lossless quantum data compression and variable-length coding. *Phy. Rev. A* **65**, 032313, https://doi.org/10.1103/Phys-RevA.65.032313 (2002).
14. Koashi, M. & Imoto, N. Quantum information is incompressible without errors. *Phy. Rev. Lett.* **89**, 097904, https://doi.org/10.1103/PhysRevLett.89.097904 (2002).
15. Ahlswede, R. & Cai, N. On lossless quantum data compression and quantum variable-length codes. In Leuchs, G. & Beth, T. (eds.) *Quantum Information Processing, chap*. 6, 66–78 (Wiley-VCH, 2003).
16. Ahlswede, R. & Cai, N. On lossless quantum data compression with a classical helper. *IEEE Transactions on Information Theory* **50**, 1208–1219, https://doi.org/10.1109/TIT.2004.828071 (2004).
17. Müller, M. & Rogers, C. Quantum bit strings and prefix-free Hilbert spaces. *arXiv preprint arXiv*:0804.0022 https://arxiv.org/abs/0804.0022 (2008).
18. Müller, M., Rogers, C. & Nagarajan, R. Lossless quantum prefix compression for communication channels that are always open. *Phy. Rev. A* **79**, 012302, https://doi.org/10.1103/PhysRevA.79.012302 (2009).
19. Hayashi, M. Universal approximation of multi-copy states and universal quantum lossless data compression. *Communications in Mathematical Physics* **293**, 171–183, https://doi.org/10.1007/s00220-009-0909-y (2010).
20. Hayashi, M. *A Group Theoretic Approach to Quantum Information* (Springer, Berlin, 2017).
21. Huffman, D. A. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE* **40**, 1098–1101, https://doi.org/10.1109/JRPROC.1952.273898 (1952).
22. Schumacher, B. & Westmoreland, M. D. Relative entropy in quantum information theory. *Contemp. Math.* **305**, 265–290 (2002).
23. Kaltchenko, A. Reexamination of quantum data compression and relative entropy. *Phy. Rev. A* **78**, 022311, https://doi.org/10.1103/PhysRevA.78.022311 (2008).
24. Parker, D. S. Conditions for optimality of the Huffman algorithms. *SIAM Journal on Computing* **9**, 470–489, https://doi.org/10.1137/0209035 (1980).
25. Humblet, P. A. Generalization of the Huffman coding to minimize the probability of buffer overflow. *IEEE Transactions on Inf. Theory* **27**, 230–232, https://doi.org/10.1109/TIT.1981.1056322 (1981).
26. Hayashi, M. Exponents of quantum fixed-length pure-state source coding. *Phys. Rev. A* **66**, 032321, https://doi.org/10.1103/PhysRevA.66.032321 (2002).
27. Hayashi, M. *Quantum Information Theory* (Springer, Berlin, 2017).
28. Birkhoff, G. Three observations on linear algebra. *Universidad Nacional de Tucumán. Revista. Serie A, matématica y física teórica* **5**, 147–151 (1946).
29. Bhatia, R. *Matrix Analysis* (Springer Verlag, New-York, 1997).
30. Petz, D. Quasi-entropies for finite quantum systems. *Reports on Math. Phy.* **23**, 57–65, https://doi.org/10.1016/0034-4877(86)90067-4 (1986).

Acknowledgements

Author Contributions

G.M.B. and S.Z. conceived the idea behind this work. G.B. and F.H. contributed with the proofs and discussions. All the authors wrote and reviewed the manuscript.

Additional Information