Lossless quantum data compression and variable-length coding

Kim Bostroem and Timo Felbinger
University of Potsdam, Potsdam, Germany
(Received 8 May 2001; published 20 February 2002)

In order to compress quantum messages without loss of information it is necessary to allow the length of the encoded messages to vary. We develop a general framework for variable-length quantum messages in close analogy to the classical case and show that lossless compression is only possible if the message to be compressed is known to the sender. The lossless compression of an ensemble of messages is bounded from below by its von-Neumann entropy. We show that it is possible to reduce the number of qbits passing through a quantum channel even below the von Neumann entropy by adding a classical side channel. We give an explicit communication protocol that realizes lossless and instantaneous quantum data compression and apply it to a simple example. This protocol can be used for both online quantum communication and storage of quantum data.

DOI: 10.1103/PhysRevA.65.032313 PACS number(s): 03.67.Hk

I. INTRODUCTION

Any physical system can be considered as a carrier of information because the state of that system could, in principle, have been intentionally manipulated to represent a *message*. The state of a system composed from distinguishable subsystems forms a message of a certain length, where each subsystem represents one letter. In quantum information theory, the systems are quantum and the system states represent quantum messages. A message is compressed if it is mapped to a shorter message and if this map is reversible, then no information has been lost. Schumacher was the first to present a method for quantum data compression [1]. It is based on the concept of encoding only a *typical subspace* spanned by the typical sequences emitted by a memoryless source. Since then there have been further investigations [2–8], but all the considered compression methods are only faithful in the limit of large block lengths. Now we ask: Is it possible to compress quantum messages without *any* loss of information? To answer this question some basic concepts of quantum information theory have to be revisited. In particular, the requirement of a fixed block length for quantum messages has to be abandoned and must be replaced by a more general theory of quantum messages, which enables a flexible and easy treatment of quantum codes involving code words of variable length. At first, we develop a general framework in close analogy to the classical case, based on previous work by one of us [9,10]. A different approach to variable-length quantum messages (appearing as a special case in our formalism) has been worked out by Braunstein *et al.* [6] and Schumacher and Westmoreland [8]. We define a measure of information quantifying the effort of communication. Compression then means reducing this effort. We argue that prefix codes are practically not very useful for quantum coding and suggest a different method involving an additional classical side channel. With the help of this channel, certain problems of instantaneous quantum communication can be avoided and, moreover, the quantum channel can be used with higher efficiency. At last, we present a communication protocol that enables lossless and instantaneous quantum data compression and we demonstrate its efficiency by an explicit example. Let us start with reviewing the fundamental notion of a *code* [17].

II. CODES

Basically, when you have a set of things and you want to give them a name, then this is a coding task. There is a code for bank accounts, telephone devices, and inhabitants of a country, there even is a code for living beings: the genetic code. Language is a code for thoughts, which are in turn codes for abstract ideas or concrete objects of human experience. A code gives *meaning* to a message, it relates objects to their description. Objects are encoded into messages composed from a basic alphabet. The number of letters that is needed to describe a particular object is a good measure of the *information content* given to the object by the code. This is the key to data compression, which we will study in the following with a focus on quantum codes.

Classically, a *code* is a map $c:\Omega\to M$ from a set of objects, Ω, to a set of messages, M (see Fig. 1). It is the messages that can be communicated and not the objects themselves, so communication is always based on a code. *Messages* (or *strings*) are sequences of letters taken from an *alphabet* $\mathcal{A}$ and are denoted by $x^n := x_1 \cdots x_n$, $x_i \in \mathcal{A}$. The

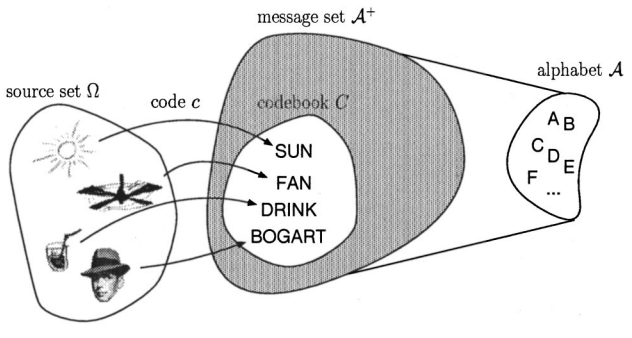

FIG. 1. A classical code is a map from a set of source objects into a set of code words composed from an alphabet. An ensemble of source objects is mapped to an ensemble of code words. For variable-length codes, the length of the code words is allowed to vary.

empty message is denoted by $x^0 := \varnothing$. All messages of length n form the set

$$\mathcal{A}^n := \{x^n | x_i \in \mathcal{A}\} \tag{1}$$

and the empty message forms the set $\mathcal{A}^0 := \{\varnothing\}$. All strings of finite length form the set of general messages over the alphabet $\mathcal{A}$,

$$\mathcal{A}^+ := \bigcup_{n=0}^{\infty} \mathcal{A}^n. \tag{2}$$

Every subset $M \subset \mathcal{A}^+$ is a *message set*. Now we can precisely define a classical *k-ary code* as a map $c: \Omega \to \mathcal{A}^+$ with $k := |\mathcal{A}|$. The set $C = c(\Omega)$ is the *code book* and each member of C is a *code word*. Being a subset of $\mathcal{A}^+$, a code book is also a message set (just like a nightingale is also a bird). If $C \subset \mathcal{A}^n$ for some $n \in \mathbb{N}$, then c is called a *block code*, otherwise a *variable-length code*. There is another important classification: *lossless* and *lossy* codes. A code is lossless (or *uniquely decodable* or *nonsingular*), if there are distinct code words for distinct objects, i.e., $\forall\ x,\ y \in \Omega: x \neq y \Rightarrow c(x) \neq c(y)$. In case of a lossy code, some objects are mapped to the same encoding. Lossy codes are used when it is more important to reduce the size of the message than to ensure the correct decoding (a fine example is the MP3 code for sound data). For a given probability distribution on Ω, lossy codes can also be useful if the *fidelity F*, i.e., the probability of correct decoding, is close to 1. For lossless codes the fidelity is exactly 1. In this paper, we only consider lossless codes.

A. The general message space

The transition from classical to quantum information is simple. We just allow the elements of a source set Ω to be in superposition. Precisely, we interpret Ω as an orthonormal basis for a Hilbert space $\mathcal{V}$ and consider every normalized vector of $\mathcal{V}$ as a valid object. Then $\mathcal{V}$ is the *linear span* of Ω and we write $\mathcal{V} = \mathrm{Span}(\Omega)$ with $\dim \mathcal{V} = |\Omega|$. The same goes for the messages. We interpret a message set M as an orthonormal basis for a message space $\mathcal{M} = \mathrm{Span}(M)$ with $\dim \mathcal{M} = |M|$ and consider each element of $\mathcal{M}$ as a valid message. The map $c: \mathcal{V} \to \mathcal{M}$ then represents a *quantum code* with the space $\mathcal{C} = c(\mathcal{V})$ being the *code space* and the elements of $\mathcal{C}$ being the code words. In order to preserve linearity, the code must be a linear map and in order to preserve norm, the code must be an isometric map. In the literature, often the code space $\mathcal{C}$ rather than the map c is called a code [this is a bit like calling $f(x)$ a function]. However, by saying ''code'' we will refer to the map c here, in full analogy to the classical case. Now let us find the general message space corresponding to the classical general message set $\mathcal{A}^+$. Interpret the letters of a *quantum alphabet* $\mathcal{Q}$ as an orthonormal basis for a *letter space* $\mathcal{H} := \mathrm{Span}(\mathcal{Q})$. A letter space $\mathcal{H}$ with $k = \dim \mathcal{H} = |\mathcal{Q}|$ is called a *k-ary space*. Quantum letters are composed into messages by tensor multiplication, giving *product messages* $|x^n\rangle := |x_1\rangle \otimes \cdots \otimes |x_n\rangle$ that form the set $\mathcal{Q}^n := \{|x^n\rangle \,||\, |x_i\rangle \in \mathcal{Q}\}$ and span the block space $\mathcal{H}^{\otimes n} := \mathrm{Span}(\mathcal{Q}^n)$, giving

$$\mathcal{H}^{\otimes n} = \bigotimes_{i=1}^{n} \mathcal{H} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}. \tag{3}$$

The space $\mathcal{H}^{\otimes n}$ is the quantum analog to the set $\mathcal{A}^n$ of classical block messages given by Eq. (1), and contains arbitrary superpositions of product messages, which are called *entangled* messages. Because superposition and entanglement have no classical interpretation, quantum information is truly different from classical information. The empty message, denoted by $|x^0\rangle \equiv |\varnothing\rangle$, forms the set $\mathcal{Q}^0 = \{|\varnothing\rangle\}$ and spans the one-dimensional space $\mathcal{H}^{\otimes 0} := \mathrm{Span}(\mathcal{Q}^0)$. Elements of $\mathcal{H}^{\otimes n}$ for some $n \in \mathbb{N}$ are called *block* messages. The set of all product messages composed from $\mathcal{Q}$ is denoted by $\mathcal{Q}^+ := \cup_{n=0}^{\infty} \mathcal{Q}^n$. Now the general message space $\mathcal{H}^{\oplus}$ induced by $\mathcal{H}$ can be defined by $\mathcal{H}^{\oplus} := \mathrm{Span}(\mathcal{Q}^+)$, giving

$$\mathcal{H}^{\oplus} = \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n} = \mathcal{H}^{\otimes 0} \oplus \mathcal{H} \oplus \mathcal{H}^{\otimes 2} \oplus \cdots. \tag{4}$$

The space $\mathcal{H}^{\oplus}$ is the quantum analog to the set $\mathcal{A}^+$ of general classical messages given by Eq. (2). $\mathcal{H}^{\oplus}$ is a separable Hilbert space with the countable basis $\mathcal{Q}^+$. The space $\mathcal{H}^{\oplus}$ is similar to the Fock space in many-particle theory, except that the particles are letters here, which must be distinguishable, so there is no symmetrization or antisymmetrization. The general message space contains also superpositions of messages of distinct length, for example,

$$\frac{1}{\sqrt{2}}(|101\rangle + |11100\rangle) \in \mathcal{H}^{\oplus} \tag{5}$$

if $|0\rangle, |1\rangle \in \mathcal{H}$. Any block space $\mathcal{H}^{\otimes n}$ is a subspace of $\mathcal{H}^{\oplus}$ and is orthogonal to any other block space $\mathcal{H}^{\otimes m}$ with $n \neq m$. Elements with components of distinct length are called *variable-length messages* (or *indeterminate-length messages*) to distinguish them from block messages. Any subspace $\mathcal{M} \subset \mathcal{H}^{\oplus}$ is called a *message space* and its elements are *quantum messages*.

B. Length operator

Define the length operator in $\mathcal{H}^{\oplus}$ measuring the length of a message as

$$\hat{L} := \sum_{n=0}^{\infty} n \Pi_n, \tag{6}$$

where Π_n is the projector on the block space $\mathcal{H}^{\otimes n} \subset \mathcal{H}^{\oplus}$, given by

$$\Pi_n = \sum_{x^n \in Q^n} |x^n\rangle\langle x^n|. \tag{7}$$

As $\hat{L}$ is a quantum observable, the length of a message $|x\rangle \in \mathcal{H}^{\oplus}$ is generally not sharply defined. Rather, the measure-

ment of $\hat{L}$ generally disturbs the message by projecting it on a block space of the corresponding length. The expected length of a message $|x\rangle \in \mathcal{H}^{\oplus}$ is given by

$$L(x) := \langle x|\hat{L}|x\rangle. \tag{8}$$

However, in $\mathcal{H}^{\oplus}$ there are also messages whose expected length is infinite. Classical analogs are probability distributions with nonexisting moments, e.g., the Lorentz distribution. Block messages are eigenvectors of $\hat{L}$, that is, $\hat{L}|x\rangle = n|x\rangle$ for all $|x\rangle \in \mathcal{H}^{\otimes n}$.

The generalization to statistical ensembles is straightforward. Consider an ensemble $\Sigma = \{p, \mathcal{X}\}$ of variable-length messages $|x\rangle \in \mathcal{X} \subset \mathcal{H}^{\varphi}$ occurring with probability $p(x) > 0$ $\forall$ $|x\rangle \in \mathcal{X}$ such that $\Sigma_{x\in\mathcal{X}}\, p(x) = 1$. Then there is a density operator

$$\sigma = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|, \tag{9}$$

called *a statistical quantum message*, representing the ensemble Σ. The set of all such density operators is denoted by $\mathcal{S}(\mathcal{H}^{\oplus})$. Vice versa, however, for a given density operator $\sigma \in \mathcal{S}(\mathcal{H}^{\oplus})$ there is, in general, a noncountable set of corresponding ensembles. In terms of information theory, σ cannot be regarded as a lossless code for the ensemble Σ. There is more information in the ensemble than in the corresponding density operator. As we will see, this additional *a priori* knowledge is in fact needed to make lossless compression possible.

The expected length of an ensemble Σ or of the corresponding statistical message $\sigma \in \mathcal{S}(\mathcal{H}^{\oplus})$ is defined as

$$L(\Sigma) = L(\sigma) := \mathrm{Tr}\{\sigma\hat{L}\} = \sum_{x\in X} p(x)L(x). \tag{10}$$

C. Base length

The expected length of a quantum message $|x\rangle$, given by Eq. (8), will, in general, not be the outcome of a length measurement. Every length measurement results in one of the length eigenvalues supported by $|x\rangle$ and generally disturbs the message. If there is a maximum value resulting from a length measurement of a state $|x\rangle$, namely, the length of the longest component of $|x\rangle$, then let us call it the *base length* of $|x\rangle$, defined as

$$\underline{L}(x) := \max\{n \in \mathbb{N} | \langle x|\Pi_n|x\rangle > 0\}. \tag{11}$$

For example, the quantum message

$$|x\rangle = \frac{1}{\sqrt{2}}(|\mathrm{abra}\rangle + |\mathrm{cadabra}\rangle) \tag{12}$$

has base length 7. Since the base length of a state is the size of its longest component, we have

$$\underline{L}(x) \geqslant L(x). \tag{13}$$

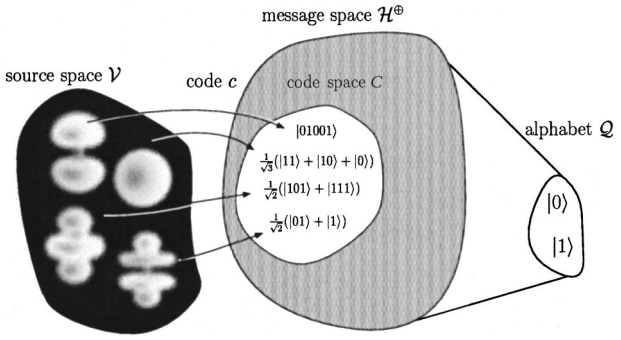

FIG. 2. A quantum code is a linear isometric map from a source space of quantum objects into a code space of code words composed from a quantum alphabet. Superpositions of source objects are encoded into superpositions of code words. An ensemble of source objects is mapped to an ensemble of code words. For a variable-length quantum code, the length of the code words is allowed to vary. Superpositions of code words of distinct length lead to code words of indeterminate length. The *base length* of a code word is defined as the length of the longest component.

It is important to note that the base length is not an observable. It is only available if the message $|x\rangle$ is *a priori* known.

D. Quantum code

Now we can precisely define a *k-ary quantum code* to be a linear isometric map $c: \mathcal{V} \to \mathcal{H}^{\oplus}$, where $\mathcal{V}$ is a Hilbert space and $\mathcal{H}^{\oplus}$ is the general message space induced by a letter space $\mathcal{H}$ of dimension k. The image of $\mathcal{V}$ under c is the *code space* $\mathcal{C} = c(\mathcal{V})$ (see Fig. 2). Being a quantum analog to the code book, $\mathcal{C}$ is the space of valid code words. The code c is uniquely specified by the transformation rule

$$|\omega\rangle \overset{c}{\mapsto} |\gamma\rangle, \tag{14}$$

where $|\omega\rangle$ are elements of a fixed orthonormal basis $\mathcal{B}_{\mathcal{V}}$ of $\mathcal{V}$ and $|\gamma\rangle = |c(\omega)\rangle$ are elements of an orthonormal basis $\mathcal{B}_{\mathcal{C}}$ of $\mathcal{C}$. Since c is an *isometric* map, i.e., $\langle\omega|\omega'\rangle = \langle c(\omega)|c(\omega')\rangle$, this implies that $|c(\omega)\rangle \neq |c(\omega')\rangle$ for all $|\omega\rangle \neq |\omega'\rangle$ in $\mathcal{V}$, so c is a lossless code with an inverse c^{-1}. The quantum code c can be represented by the isometric operator

$$C := \sum_{\omega\in\mathcal{B}_{\mathcal{V}}} |c(\omega)\rangle\langle\omega| = \sum_{\gamma\in\mathcal{B}_{\mathcal{C}}} |\gamma\rangle\langle c^{-1}(\gamma)| \tag{15}$$

called the encoder of c. Since c is lossless, there is an inverse operator

$$D := C^{-1} = \sum_{\gamma\in\mathcal{B}_{\mathcal{V}}} |\omega\rangle\langle c(\omega)| = \sum_{\gamma\in\mathcal{B}_{C}} |c^{-1}(\gamma)\rangle\langle\gamma| \tag{16}$$

called the *decoder*. In practice, the source space $\mathcal{V}$ and the code space $\mathcal{C}$ are often subspaces of one and the same physical space $\mathcal{R}$. Since C is an isometric operator between $\mathcal{V}$ and $\mathcal{C}$, there is a (nonunique) *unitary extension* U_C on $\mathcal{R}$ with

$$U_C|x\rangle = C|x\rangle, \quad \forall \ |x\rangle \in \mathcal{V} \subset \mathcal{R}, \tag{17}$$

$$U_C^\dagger|y\rangle = C^{-1}|y\rangle, \quad \forall \ |y\rangle \in \mathcal{C} \subset \mathcal{R}. \tag{18}$$

However, using C and distinguishing between $\mathcal{V}$ and $\mathcal{C}$ is more convenient and more general. Codes with $\mathcal{C} \subset \mathcal{H}^{\otimes n}$ for some $n \in \mathbb{N}$ are called block codes, otherwise variable-length codes.

III. REALIZING VARIABLE-LENGTH MESSAGES

Variable-length messages could, in principle, directly be realized by a quantum system whose particle number is not conserved, for instance, an electromagnetic field. Each photon may carry letter information by its field mode, while the number of photons may represent the length of the message. The photons can be ordered either using their spacetime position (e.g., single photons running through a wire) or some internal state with many degrees of freedom (e.g., a photon with frequency ω_2 can be defined to ''follow'' a photon with frequency $\omega_1 < \omega_2$). The Hilbert space representing such a system of distinguishable particles with nonconserved particle number simply is the message space $\mathcal{H}^\oplus$. In case we have only a system at hand where the number of particles is conserved, we can also realize variable-length messages by embedding them into block spaces.

It is a good idea to distinguish between the message space, which is a purely abstract space, from its physical realization. Let us call the physical realization of a message space $\mathcal{M}$ the operational space $\tilde{\mathcal{M}}$. Between $\mathcal{M}$ and $\tilde{\mathcal{M}}$, there is an isometric map, so dim $\mathcal{M}$ = dim $\tilde{\mathcal{M}}$. This is expressed by $\mathcal{M} \cong \tilde{\mathcal{M}}$. The operational space $\tilde{\mathcal{M}}$ is the space of physical states of a system representing valid code words of $\mathcal{M}$. Often the operational space is a subspace of the total space of all physical states of the system. Denoting the total physical space by $\mathcal{R}$ we have

$$\mathcal{M} \cong \tilde{\mathcal{M}} \subset \mathcal{R}. \tag{19}$$

A. Bounded message spaces

The general message space $\mathcal{H}^\oplus$ is the ''mother'' of all message spaces induced by the letter space $\mathcal{H}$. It contains just about *every* quantum message that can be composed using letters from $\mathcal{H}$ and the laws of quantum mechanics. However, it is an abstract space, i.e., independent from a particular physical implementation. It would be good to know if such a space can also physically be realized. It is clear that if you have a finite system you can only realize a *finite-dimensional subspace* of the general message space, whose dimension is infinite. So let us start with the physical realization of the r-bounded message space

$$\mathcal{H}^{\oplus r} := \bigoplus_{n=0}^{r} \mathcal{H}^{\otimes n}, \tag{20}$$

containing all superpositions of messages of maximal length r.

Say you have a physical space $\mathcal{R} = \mathcal{D}^{\otimes s}$ representing a register (see Fig. 3) consisting of s systems with dim $\mathcal{D} = k$. Each subspace $\mathcal{D}$ represents one *quantum digit* in the register. In the case $k=2$ the quantum digits are *quantum bits*, in short ''qubits.'' The physical space R represents the space of all physical states of the register, while the message space $\mathcal{H}^{\oplus r}$ represents the space of valid code words that can be held by the register and it is isomorphic to a subspace $\tilde{\mathcal{H}}^{\oplus r}$ of the physical space $\mathcal{R}$. Let dim $\mathcal{H} = k$, then you must choose s such that

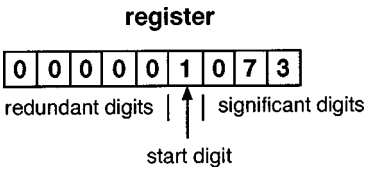

FIG. 3. Realizing a general variable-length message.

$$\dim(\mathcal{H}_i^{\oplus r}) \leq \dim(\mathcal{D}^{\otimes s}), \tag{21}$$

$$\Rightarrow \sum_{n=0}^{r} k^n = \frac{k^{r+1}-1}{k-1} \leq k^s, \tag{22}$$

$$\Rightarrow s \geq r+1. \tag{23}$$

Thus you need a register of at least $(r+1)$ digits to realize the message space $\mathcal{H}^{\oplus r}$. Choose the smallest possible register space $\mathcal{R} = \mathcal{D}^{\otimes (r+1)}$. Since at most r digits are carrying information, one digit can be used to indicate either the beginning or the end of the message. Now you can conveniently use *k-ary representations of natural numbers* as code words. Each natural number i has a unique k-ary representation $Z_k(i)$. For instance, $Z_2(3) = 11$ and $Z_{16}(243) = E3$. All k-ary representations have a *neutral prefix* ''0'' that can precede the representation without changing its value, e.g., $000011 \cong 11$. For a natural number $n > 0$, define $Z_k^n(i)$ as the n-extended k-ary representation of i by

$$Z_k^n(i) := \underbrace{0 \cdots 0}_{n} Z_k(i), \quad 0 \leq i \leq k^r - 1. \tag{24}$$

For example, $Z_2^6(3) = 000011$ and $Z_{16}^6(243) = 0000E3$. Let us define that the message starts after the first appearance of ''1,'' e.g., $000102540 \cong 02540$. Now define orthonormal vectors

$$|e_i^n\rangle := |\underbrace{0 \cdots 0}_{r-n} 1 Z_k^n(i)\rangle \in \mathcal{R}, \tag{25}$$

where $n > 0$ and $0 \leq i \leq k^n - 1$. The n digits of $Z_k^n(i)$ are called *significant digits*. The empty message corresponds to the unit vector

$$|\varnothing\rangle := |e_0^0\rangle := |0 \cdots 01\rangle. \tag{26}$$

Obviously, $|\varnothing\rangle$ has no significant digits. Next, define orthonormal basis sets

$$\tilde{\mathcal{B}}^n := \{|e_0^n\rangle, \ldots, |e_{k^n-1}^n\rangle\}, \quad 0 \leq n \leq r, \tag{27}$$

which span the operational block spaces

$$\tilde{\mathcal{H}}^{\otimes n} = \mathrm{Span}(\tilde{\mathcal{B}}^n). \tag{28}$$

Note that $\tilde{\mathcal{H}}^{\otimes n}$ is truly different from $\mathcal{H}^{\otimes n}$, because $\tilde{\mathcal{H}}^{\otimes n}$ has dimension k^{r+1}, while $\tilde{\mathcal{H}}^{\otimes n}$ has dimension k^n. Next, define an orthonormal basis

$$\tilde{\mathcal{B}}^+ := \bigcup_{n=0}^{r} \tilde{\mathcal{B}}^n \tag{29}$$

and construct the operational space $\tilde{\mathcal{H}}^{\oplus r} \subset \mathcal{R}$ as

$$\tilde{\mathcal{H}}^{\oplus r} := \mathrm{Span}(\tilde{\mathcal{B}}^+). \tag{30}$$

Altogether, the physical space $\mathcal{R} = \mathcal{D}^{\otimes (r+1)}$ is the space of all physical states of the register, while the operational space $\tilde{\mathcal{H}}^{\oplus r} \subset \mathcal{R}$ is the space of those register states that represent valid code words, and it is isomorphic to the abstract message space $\mathcal{H}^{\oplus r}$.

A general message is represented by the vector

$$|x\rangle = \sum_{n=0}^{r} \sum_{i=0}^{k^n-1} x_{n,i} |e_i^n\rangle \tag{31}$$

with $\Sigma_{n=0}^{r} \Sigma_{i=0}^{k^n-1} |x_{n,i}|^2 = 1$. The length operator introduced in Sec. II B is here of the form

$$\hat{L} := \sum_{n=0}^{r} n \Pi_n, \tag{32}$$

because there are at most r digits to constitute a message. Now we need to know how the projectors Π_n are constructed in the operational space $\tilde{\mathcal{H}}^{\oplus r}$. For a register state containing a message of sharply defined length, the length eigenvalue n is given by the *number of significant digits* in that register,

$$\hat{L} |e_i^n\rangle := n |e_i^n\rangle, \tag{33}$$

for $0 \leqslant i \leqslant k^n - 1$. Each projector is then defined by

$$\Pi_n := \sum_{i=0}^{k^n-1} |e_i^n\rangle\langle e_i^n| \tag{34}$$

and projects onto the space $\mathcal{H}^{\otimes n} \subset \mathcal{R}$. Note that the *physical length* of each message is always given by the fixed size $(r+1)$ of the register. Only the *significant length* of a message, i.e., the number of digits that constitute a message contained in the register, is, in general, not sharply defined. Note further that the particular form of the length operator depends on the realization of the message space.

In the limit of large r we have $\lim_{r\to\infty} \mathcal{H}^{\oplus r} = \mathcal{H}^{\oplus}$, but that space can no longer be embedded into a physical space $\mathcal{R} = \mathcal{D}^{\otimes\infty} := \lim_{n\to\infty} \mathcal{D}^{\otimes n}$, since the latter is no longer a separable Hilbert space. However, we can think of r as very large, such that working in $\mathcal{H}^{\oplus}$ just means working with a quantum computer having enough memory.

B. Realizing more message spaces

A code is a map $c: \mathcal{V} \to \mathcal{H}^{\oplus}$ from source states in $\mathcal{V}$ to code words in $\mathcal{H}^{\oplus}$. The space $\mathcal{C} = c(\mathcal{V})$ of all code words is the code space and as a subspace of the general message space $\mathcal{H}^{\oplus}$ it is just a special message space. In order to implement a particular code c, it is, in practice, sufficient to realize only the corresponding code space $\mathcal{C}$ by a physical system. Let us realize some important code spaces now. However, we will not discuss the very important class of *error-correcting* code spaces here, since this would go beyond the scope of this paper.

1. Block spaces

An important message space is the block space $\mathcal{H}^{\otimes n}$, which contains messages of fixed length n. Block spaces are *the* message spaces of standard quantum information theory. They can directly be realized by a register $\mathcal{R} = \mathcal{H}^{\otimes n}$ of n digits, e.g., n two-level systems representing one qubit each.

2. Prefix spaces

Another interesting message space is the space of prefix code words of maximal length r. Such a space contains only superpositions of prefix code words. A set of code words is *prefix* (or *prefix-free*), if no code word is the prefix of another code word. For example, the set $P_3 = \{0,10,110,111\}$ is a set of binary prefix code words of maximal length 3. Prefix code words have one significant advantage: Prefix code words are *instantaneous*, that is, sequences of prefix code words do not need a word separator. The separator can be added while reading the sequence from left to right. A sequence from P_3 can be separated as

$$110111010110 \mapsto 110,111,0,10,110. \tag{35}$$

However, there is also a drawback: Prefix code words are, in general, not as short as possible. This is a consequence of the fact that there are, in general, less prefix code words than possible code words. For example, if you want to encode four different objects, you can use the prefix set P_3 above with maximal length 3. If you renounce the prefix property you can use the set {0, 1, 01, 10} with maximal length 2.

A prefix space $\mathcal{P}_r$ of maximal length r is given by the linear span of prefix code words of maximal length r. For the set P_3, the corresponding prefix space is $\mathcal{P}_3 = \mathrm{Span}\{|0\rangle, |10\rangle, |110\rangle, |111\rangle\}$. The prefix space $\mathcal{P}_r \subset \mathcal{H}^{\oplus r}$ can physically be realized by a subspace $\tilde{\mathcal{P}}_r$ of the register space $\mathcal{R} = \mathcal{D}^{\otimes r}$ spanned by the prefix code words that have been extended by zeros at the end to fit them into the register. For example, $\tilde{\mathcal{P}}_3 = \mathrm{Span}\{|000\rangle, |100\rangle, |110\rangle, |111\rangle\} \subset \mathcal{D}^{\otimes 3}$ is a physical realization of the prefix space $\mathcal{P}_3$. The length operator measures the significant length of the code words, given by the length of the corresponding prefix code words.

Schumacher and Westmoreland [8] as well as Braunstein *et al.* [6] used prefix spaces for their implementation of variable-length quantum coding. However, we will show later on that the significant advantage of prefix code words in fact vanishes in the quantum case, whereas the disadvantage remains.

FIG. 4. Realizing variable-length messages by neutral-prefix code words.

3. Neutral-prefix space

A specific code space will be of interest, namely, the space of *neutral-prefix* code words, which we define as follows. The k-ary representation of a natural number i is denoted by $Z_k(i)$ (see Sec. III A). The empty message $\varnothing$ is represented by $Z_k(0)=\varnothing$. Define an orthonormal basis

$$\mathcal{B}_r := \{|Z_k(0)|,\ldots,\rangle Z_k(k^r-1)\rangle\} \tag{36}$$

of variable-length messages of maximal length r. The length of each basis message $|Z_k(i)\rangle$ is given by

$$|Z_k(i)| = \lceil \log_k(i+1) \rceil, \tag{37}$$

where $\lceil x \rceil$ denotes the smallest integer $\lceil \geqslant \rceil x$. These basis messages span the r-bounded neutral-prefix space

$$\mathcal{N}_r := \mathrm{Span}(\mathcal{B}_r). \tag{38}$$

Note that $\mathcal{N}_r$ is not equal to the r-bounded message space $\mathcal{H}^{\oplus r}$ as you can see by comparing the dimension $\dim \mathcal{N}_r = k^r$ with $\dim \mathcal{H}^{\oplus r} = (k^{r+1}-1)/(k-1)$. $\mathcal{N}_r$ is smaller than $\mathcal{H}^{\oplus r}$ because not all messages of $\mathcal{H}^{\oplus r}$ are contained in $\mathcal{N}_r$. For example, the message $|01\rangle$ is in $\mathcal{H}^{\oplus r}$ but not in $\mathcal{N}_r$, hence we have

$$\mathcal{N}_r \subset \mathcal{H}^{\oplus r}. \tag{39}$$

Now we want to find a physical realization of $\mathcal{N}_r$. This turns out to be quite easy (see Fig. 4). As already noted in Sec. III A, the k-ary representation $Z_k(i)$ of any natural number i can be extended by adding leading zeros to the r-extended k-ary representation $Z_k^r(i) := 0\cdots 0 Z_k(i)$. Take a register $\mathcal{R} = \mathcal{D}^{\otimes r}$ of r digits with $\mathcal{D} = \mathbb{C}^k$. Then the set

$$\mathcal{B}_{\mathcal{R}} := \{|Z_k^r(0)\rangle,\ldots,|Z_k^r(k^r-1)\rangle\} \tag{40}$$

is an orthonormal basis for the register space $\mathcal{R}$. At the same time it can be regarded as an orthonormal basis for the operational space $\tilde{\mathcal{N}}_r$ representing the neutral-prefix space $\mathcal{N}_r$. While the physical length of each code word is constantly r, the significant length is measured by the length operator

$$\hat{L} := \sum_{n=0}^{r} n \Pi_n, \tag{41}$$

with mutually orthogonal projectors

$$\Pi_n := \sum_{i:|Z_k(i)|=n} |Z_k^r(i)\rangle\langle Z_k^r(i)|. \tag{42}$$

Note that the so-defined length operator looks different from that defined in Sec. III A. While $\hat{L}$ is always of the same form (32), the projectors Π_n are different because the operational spaces are different.

The empty message can be defined by

$$|\varnothing\rangle := |Z_k^r(0)\rangle = |0\cdots 0\rangle. \tag{43}$$

A general message in $\tilde{\mathcal{N}}_r$ is given by

$$|x\rangle = \sum_{i=0}^{k^r-1} x_i |Z_k^r(i)\rangle. \tag{44}$$

We have realized the neutral-prefix space $\mathcal{N}_r$ by exhausting the entire register space $\mathcal{R}$ so that the quantum resources are optimally used. In other words all messages in $\mathcal{N}_r$ are as short as possible. Remember that the physical realization of $\mathcal{H}^{\oplus r}$ requires one additional digit to represent the beginning or the end of a message. This digit does not contain any message information, it is sort of wasted. For quantum coding, the additional digit may really count, since it would have to be added each time a code word is stored or transmitted. Also the prefix space considered in Sec. III B 2 contains messages that are not as short as possible. You can encode a space $\mathcal{V}$ of dimension $\dim \mathcal{V}=4$ by a prefix space spanned by $\{|000\rangle,|100\rangle,|110\rangle,|111\rangle\}$ with corresponding lengths {1, 2, 3, 3}, but then you need a register of three qubits. In contrast to that, $\mathcal{V}$ can be encoded by a neutral-prefix space spanned by the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ with corresponding lengths {0, 1, 2, 2}, and you need a register of only two qubits. In the operational space $\tilde{\mathcal{N}}_r$, the basis messages reveal their length information by simply discarding leading zeros. That way, not all variable-length messages can be realized, but we save one register digit, so $\mathcal{N}_r$ is a good candidate for variable-length quantum coding.

IV. DATA COMPRESSION

A. Classical data compression

Intuitively, compression is achieved when the effort to store or communicate the code words is minimized. But how can we precisely define that ''effort?'' The key idea is the concept of a *raw code*. One can always construct a code for Ω by inventing a new letter for each single object. Such a classical raw code is a code $c:\Omega\to\mathcal{A}$ for some alphabet $\mathcal{A}$ of the same size as Ω. The chinese writing is a fairly good illustration of a raw code. There are up to 50 000 letters representing a manifold of abstract and concrete things, e.g., the ''noise of a running horse.'' The length of the code is minimized to 1, but the encoding and decoding machines will need a large memory to remember all the letters. Obviously, a raw code does not compress at all, so it is a good idea to set the effort of communication in relation to the raw information content of Ω (similar to the notion in [14], p. 71 and interestingly similar also to the Boltzmann entropy of a microcanonical ensemble), defined by

$$I_0(\Omega) := \log_2|\Omega|. \tag{45}$$

$I_0(\Omega)$ represents the number of binary digits (bits) needed to enumerate the elements of Ω. This motivates the following definition. The *code information content* of an individual object in an arbitrary set Ω for a given k-ary code $c:\Omega\to\mathcal{A}^+$ is defined as

$$I_c(x) := \log_2(k) L_c(x), \quad x \in \Omega, \tag{46}$$

where $L_c(x)$ denotes the length of the code word $c(x) \in \mathcal{A}^+$. $I_c(x)$ represents the number of bits needed to describe the object x by the code c. For a raw code $c:\Omega\to\mathcal{A}$, definition (46) gives the raw information content for every object $x \in \Omega$. A few remarks about the code information are as follows.

(1) The code information is defined for things, not for strings. Of course, things may sometimes also be strings. If so, one can define the *direct information* of a string x^n over an alphabet $\mathcal{A}$ as

$$I(x^n) := n \log_2|\mathcal{A}|. \tag{47}$$

(2) The code information I_c is code dependent, reflecting the philosophy that there is no information contained in an object without a code giving it some meaning. The code word "XWF$%&$ FggHz((" may be a random sequence of letters or may in a certain code represent the first digits of π or in another code the beginning of a Mozart symphony.

Now let there be a probability distribution p on Ω. We can define the code information of the ensemble $\Sigma=\{p,\Omega\}$ as the average of Eq. (46),

$$I_e(\Sigma) := \log_2 k \sum_{x\in\Omega} p(x) L_c(x). \tag{48}$$

Compression means reducing the code information of the ensemble. We can define the compression rate achieved by a code c on the ensemble Σ by

$$R_c(\Sigma) := \frac{I_c(\Sigma)}{I_0(\Omega)}. \tag{49}$$

A code $c:\Omega\to C$ is compressive on Σ if and only if

$$R_c(\Sigma)<1, \quad \text{i.e.,} \quad I_c(\Sigma)<I_0(\Omega). \tag{50}$$

B. Quantum data compression

Now that we have a classical definition of compression, the next step is to translate these concepts to the quantum case. Again, the key is the raw information, i.e., the size of a noncompressed message, so let us look for its quantum analog. The raw information (45) of a set Ω is $I_0(\Omega)=\log_2|\Omega|$ because we need $|\Omega|$ distinct letters to encode each element of Ω by a raw code. Interpreting Ω as an orthonormal basis for a Hilbert space $\mathcal{V}$, the raw information of $\mathcal{V}$ is also $\log_2|\Omega|$, because we still need $|\Omega|$ distinguishable letters to represent each element of the space $\mathcal{V}$. Since $|\Omega|=\dim\mathcal{V}$, we define the quantum raw information of a spacer $\mathcal{V}$ as

$$I_0(\mathcal{V}) := \log_2(\dim\mathcal{V}). \tag{51}$$

So the quantum raw information I_0 corresponding to a space $\mathcal{V}$ equals the fixed number of qubits needed to represent all states in $\mathcal{V}$.

Now, for a given k-ary code $c:\mathcal{V}\to\mathcal{H}^{\oplus}$ represented by an encoder C, the code information operator can be defined as

$$\hat{I}_c := \log_2(k)\hat{L}_c, \tag{52}$$

where $\hat{L}_c := C^{-1}\hat{L}C$ is the length operator measuring the length of the code word for a source vector in $\mathcal{V}$. If the code is based on a qubit alphabet, $\hat{I}_c$ measures the number of qubits forming the code message, hence the measuring unit of $\hat{I}_c$ is "1 qubit." In analogy to Eq. (47), we define the direct information operator acting on the message space $\mathcal{H}^{\oplus}$ by

$$\hat{I} := \log_2(k)\hat{L}. \tag{53}$$

In short, the code information operator is defined in an arbitrary Hilbert space $\mathcal{V}$ and depends on a quantum code $c:\mathcal{V}\to\mathcal{H}^{\oplus}$, while the direct information operator is defined in a message space $\mathcal{H}^{\oplus}$ without referring to a quantum code. For a given code, the relation between both operators is

$$\hat{I}_c = C^{-1}\hat{I}C. \tag{54}$$

Now one compresses a code word by removing redundant quantum digits. The number of quantum digits carrying information is given by the base length of the code word. All other digits are redundant and can be removed without loss of information. This motivates the definition of the *code information* of a state $|x\rangle\in\mathcal{V}$ respecting a code c by

$$\underline{I}_c(x) := \log_2(k)\underline{L}_c(x), \tag{55}$$

where $\underline{L}_c(x)=\underline{L}(c(x))$ is the base length of the code word for $|x\rangle$. $\underline{I}_c(x)$ represents the number of qubits needed to describe the state $|x\rangle$ by the code c. This value must be distinguished from the *expected* number of qubits $I_c(x) = \langle x|\hat{I}_c|x\rangle$, which is found by performing a length measurement on the code word for $|x\rangle$. In the classical case, this difference vanishes.

Now one wants to encode an ensemble $\Sigma=\{p,\mathcal{X}\}$ of states $|x\rangle\in\mathcal{X}$ that span the source space $\mathcal{V}$. Each individual message $|x\rangle$ can be compressed to $\underline{I}_c(x)$ qubits, so the entire ensemble Σ will on the average be compressed to the code information

$$\underline{I}_c(\Sigma) := \log_2 k \sum_{x\in\mathcal{X}'} p(x)\underline{L}_c(x). \tag{56}$$

The compression rate can then be defined by

$$R_c(\Sigma) := \frac{\underline{I}_c(\Sigma)}{I_0(\mathcal{V})}. \tag{57}$$

A code c is compressive on the ensemble Σ, if and only if

$$R_c(\Sigma)<1, \quad \text{i.e.,} \quad \underline{I}_c(\Sigma)<I_0(\mathcal{V}). \tag{58}$$

Note that these definitions only apply to lossless codes. The lossy case is not considered here.

V. NO-GO THEOREMS

Of course, lossy compression is always possible. But let us look for some statements about lossless codes. The first three of the following no-go theorems are also known in classical information theory and are easily transferred to the quantum case by general reasoning. However, we show them by applying the tools developed in this paper. The last theorem is genuinely quantum with no classical analog.

A. No lossless compression by block codes

A code is a block code if all code words have the same length, else it is a variable-length code. Unfortunately, lossless block codes do not compress. Take an arbitrary ensemble $\Sigma=\{p,\mathcal{X}\}$ with $\mathcal{X}\subset\mathcal{V}$ and any lossless k-ary block code $c:\mathcal{V}\to\mathcal{H}^{\otimes n}$. Let $\mathcal{B}_{\mathcal{V}}$ and $\mathcal{B}_n$ be orthonormal basis sets of $\mathcal{V}$ and $\mathcal{H}^{\otimes n}$, respectively. In order to find for every basis vector $|\omega\rangle\in\mathcal{B}_{\mathcal{V}}$ a code basis vector $|c(\omega)\rangle\in\mathcal{B}_n$, the code must fulfill $\dim\mathcal{V}\leqslant\dim\mathcal{H}^{\otimes n}=k^n$. For every $|x\rangle\in\mathcal{X}$, the corresponding code word $|c(x)\rangle$ has sharp length $L(x)=n$, hence

$$\underline{I}_c(\Sigma)=\log_2 k\sum_{x\in\mathcal{X}}p(x)L_c(x)=\log_2(k)n=\log_2(k^n) \tag{59}$$

$$\geqslant\log_2(\dim\,\mathcal{V})=I_0(\mathcal{V}), \tag{60}$$

which violates condition (58). This implies that there is no lossless compressing block code. By choosing mutually orthogonal source states one can derive the analog statement for the classical case.

For long strings emitted by a memoryless source, block codes can achieve almost lossless compression by encoding only typical subspaces. The quantum code performing this type of lossy compression is known as the Schumacher code [3]. The only way to compress messages without loss of information is by use of a variable-length code. In order to achieve compression, more frequent objects must be encoded by shorter messages, less frequent objects by longer messages, so that the average length of the codes is minimized. This is the general rule of lossless data compression.

B. No lossless compression by changing the alphabet

Trying to achieve compression by using a different alphabet does not work.

A code $c:\mathcal{H}_A^{\otimes n}\to\mathcal{H}_B^{\otimes m}$ that transforms messages over some letter space $\mathcal{H}_A$ into messages over some letter space $\mathcal{H}_B$ is lossless only if $\dim\mathcal{H}_A^{\otimes n}\leqslant\dim\mathcal{H}_B^{\otimes n}$, which implies that

$$I_0(\mathcal{V})=n\log_2(\dim\mathcal{H}_A) \tag{61}$$

$$\leqslant m\log_2(\dim\mathcal{H}_B)=\underline{I}_c(x) \tag{62}$$

for every $|x\rangle\in\mathcal{H}_A$. So for every ensemble $\Sigma=\{p,\mathcal{X}\}$ of messages $|x\rangle\in\mathcal{H}_A^m$, we have $\underline{I}_c(\Sigma)=\underline{I}_c(x)\geqslant I_0(\mathcal{V})$, which violates condition (58). By choosing mutually orthogonal source states, one can derive the analog statement for the classical case. The present paper would probably *look* much shorter when written in Chinese symbols. However the effort of communication that is expressed by the code information I_c would not be reduced.

C. No universal lossless compression

We have seen that it is not possible to compress messages without loss of information by using a block code or by using a different letter space. Now we will see that no code can compress *all* messages without loss of information.

Say you have a space $\mathcal{H}^{\otimes n}$ of block messages of fixed length r and you want to compress all of them by use of a variable-length code $c:\mathcal{H}^{\otimes r}\to\mathcal{H}^{\oplus s}$ with $s<r$. The code can only be lossless if

$$\dim\mathcal{H}^{\otimes r}\leqslant\dim\mathcal{H}^{\oplus s}. \tag{63}$$

But since $\dim\mathcal{H}^{\otimes r}=k^r$ and $\dim\mathcal{H}^{\oplus s}=(k^{s+1}-1)/(k-1)$, we have

$$k^r\leqslant\frac{k^{s+1}-1}{k-1} \tag{64}$$

$$\Rightarrow k^{r+1}\leqslant k^{s+1}+k-1, \tag{65}$$

which is wrong for $r\geqslant s$ and $k>1$, so you cannot compress all block messages of a given length. Now say you have a space $\mathcal{H}^{\oplus r}$ of variable-length messages with maximal length r. Assume that there is a universal lossless code c that reduces the length of all messages in $\mathcal{H}^{\oplus r}$. The code can only be lossless if $\dim\mathcal{H}^{\oplus r}\leqslant\dim\mathcal{H}^{\oplus s}$, which is obviously wrong for $r>s$, so you cannot compress all variable-length messages with a given maximal length. Concluding, there is no universal lossless compression that reduces the size of all messages. Some messages are unavoidably lengthened by a lossless code. By choosing mutually orthogonal source states, one can derive the analog statement for the classical case.

D. No lossless compression of unknown messages

Now we come to a no-go theorem that is typically quantum. In quantum mechanics there is a profound difference between a *known* and an *unknown* state. For example, a known state can be cloned (by simply preparing another copy of it), whereas an unknown state cannot be cloned.

Assume that there is a lossless quantum compression algorithm $c:\mathcal{H}^{\otimes r}\to\mathcal{H}^{\oplus s}$ that compresses messages of fixed length r to variable-length messages of maximal length s. As we have seen in the preceding section, a lossless code cannot compress *all* messages, so $s>r$. Now there is an oracle that hands you an arbitrary message $|x\rangle=\Sigma_{i=1}^n x_i|\omega_i\rangle$, where the $|\omega_i\rangle\in\mathcal{H}^{\oplus r}$ are mutually orthogonal states. The algorithm encodes the message $|x\rangle$ into $|c(x)\rangle=\Sigma_{i=1}^n x_i|c(\omega_i)\rangle$. Even if

all the code-word components $|c(\omega_i)\rangle$ have determinate length $L_c(\omega_i)$, the total code word $|c(x)\rangle$ has, in general, indeterminate length. If you want to remove redundant digits without loss of information, you must know at least an upper bound for its base length, i.e., the length of its longest component. Since you do not know the source message $|x\rangle$, you do not know the base length of its encoding $|c(x)\rangle$, so you have to assume the maximal length s. Since $s>r$, no compression is achieved. The same argument applies to quantum-compression algorithms $c:\mathcal{H}^{\oplus r}\rightarrow\mathcal{H}^{\oplus s}$ compressing variable-length messages of maximal length r to variable-length messages of maximal length s.

We conclude that lossless compression of unknown quantum messages is, in general, impossible. This statement is not true for the classical case. A classical message is not disturbed by a length measurement, so it can, in principle, be compressed without loss of information. It would have been nice to compress a quantum hard disk without loss of information just like a classical hard disk, but this cannot, in general, be accomplished.

Now that we have found a lot of impossible things to do with quantum messages, it is time to look for the possible things.

VI. LOSSLESS COMPRESSING CODES

The intention of using compressing codes is to minimize the effort of communication between two parties: one who *prepares, encodes, compresses, and sends* the messages and one who *receives, decompresses, decodes*, and possibly *reads* them. So it is time for Alice and Bob to enter the scene. Alice is preparing source messages $|x\rangle\in\mathcal{V}$ and encodes them into code words $|c(x)\rangle\in\mathcal{H}^{\oplus r}$ by applying the encoder C. She compresses the code words by removing redundant quantum digits and sends the result to Bob, who receives them and decompresses them by appending quantum digits. After that he can decode the messages by applying the decoder D and read them or use them as an input for further computations. The communication has been lossless if the decoded message equals the source message. Note that it is not required for Bob to *read* the message he received. In fact, if Bob wants to use the message as an input for a quantum computer, he even must not do that, else he will potentially lose information. We require Alice to know which source messages she prepares, otherwise no lossless compression is possible, as we have seen in the preceding section.

A. Why prefix quantum codes are not very useful

In classical information theory, prefix codes are favored for lossless coding. The reason is that they are *instantaneous*, which means that they carry their own length information (see Sec. III B 2). Prefix code words can be sent or stored without a separating signal between them. The decoder can add word separators (commas) while reading the sequence from left to right. Whenever a string of letters yields a valid code word, the decoder can add a comma and proceed. After all, a continuous stream of letters is separated into valid code words.

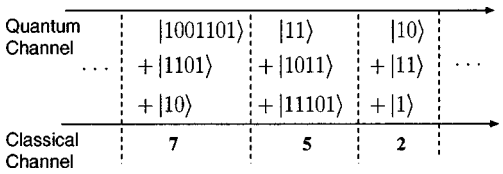

FIG. 5. Storing length information in a classical side channel.

Prefix code words can be separated while reading the sequence, but in the quantum case this is potentially a very bad thing to do. Reading a stream of quantum letters means, in general, disturbing the message all the time. Therefore, the length information is generally not available. Furthermore, prefix code words are, in general, longer than nonprefix code words, because there are less prefix code words of a given maximal length than possible code words. Hence, by using prefix code words, qubits are wasted to encode length information, which is unavailable anyway. We conclude that prefix quantum codes are practically not very useful for lossless coding.

B. A classical side channel

One could try to encode length information in a different quantum channel, as proposed by Braunstein *et al.* [6] (unnecessarily they used prefix code words anyhow). But that does not fix the problem. Whatever one does, reading out length information about different components of a variable-length code word equals a length measurement and hence means disturbing the message. There should be *some* way to make sure where the code words have to be separated, else the message cannot be decoded at all. Here is an idea: Use a *classical side channel* to inform the receiver where the code words have to be separated. This has two significant advantages: (i) If the length information equals the base length of the code word, the message is not disturbed and can be losslessly transmitted and decoded; (ii) abandoning the prefix condition, shorter code words can be chosen, such that the quantum channel is used with higher efficiency.

Let us give an example (see Fig. 5). Alice wants to send a message $|x_1\rangle$, which is encoded into the code word $|c(x_1)\rangle=(1/\sqrt{3})(|1001101\rangle+|1101\rangle+|10\rangle)$. The base length of $|c(x_1)\rangle$ is 7, so she submits that information through the classical channel. Dependent on which realization of variable-length messages Alice and Bob have agreed to use, Alice sends enough qubits (at least 7) representing the code word $|c(x_1)\rangle$ through the quantum channel. The next code word is $|c(x_2)\rangle=(1/\sqrt{3})(|11\rangle+|1011\rangle+|11101\rangle)$. The base length of $|c(x_2)\rangle$ is 5, so Alice sends the length information "5" through the classical channel and enough qubits (at least 5) representing the code word $|c(x_2)\rangle$ through the quantum channel. She proceeds like that with all following messages. On Bob's side, there is a continuous stream of qubits coming through the quantum channel and a continuous stream of classical bits coming through the classical channel. Bob can read out the classical length information, separate the qubits into the specified blocks and apply the decoder to each code word. In this way, Bob obtains all source messages without loss of information.

C. How much compression?

1. Lower bound

How much compression can maximally be achieved by using the method sketched in Sec. VI B? Say Alice has an ensemble $\Sigma=\{p,\mathcal{X}\}$ of $m=|\mathcal{X}|$ messages, $|x_i\rangle\in\mathcal{X}$, $i=1,\ldots,m$ that she wants to encode by k-ary code words. The source space $\mathcal{V}$ is spanned by the elements of $\mathcal{X}$, thus $\mathcal{V}:=\mathrm{Span}(\mathcal{X})$ and has dimension $d:=\dim\mathcal{V}$. Alice fixes a basis set $\mathcal{B}_\mathcal{V}$ of d orthonormal vectors $|\omega_i\rangle$, $i=1,\ldots,d$. The ensemble Σ corresponds to the message matrix

$$\sigma:=\sum_{i=1}^{m} p(x_i)|x_i\rangle\langle x_i|=\sum_{i,j=1}^{d}\sigma_{ij}|\omega_i\rangle\langle\omega_j|, \tag{66}$$

with $\sigma_{ij}:=\langle\omega_i|\sigma|\omega_j\rangle$ and $\Sigma_{i=1}^{d}\sigma_{ii}=1$. The source messages are encoded by the isometric map $c:\mathcal{V}\to\mathcal{H}^{\oplus}$, defined by

$$|\omega_i\rangle\overset{c}{\mapsto}|c(\omega_i)\rangle,\quad i=1,\ldots,d. \tag{67}$$

The code space is k-ary, which means that $k=\dim\mathcal{H}$. Let each code word $|c(\omega_i)\rangle$ have determinate length $L_c(\omega_i)$ such that the code-length operator $\hat{L}_c$ on $\mathcal{V}$ is orthogonal in the basis $\mathcal{B}_\mathcal{V}$ and reads

$$\hat{L}_c=\sum_{i=1}^{d} L_c(\omega_i)|\omega_i\rangle\langle\omega_i|. \tag{68}$$

The code words $|c(\omega_i)\rangle$ are not necessarily prefix, because Alice can encode the length information about each code word in a classical side channel. In order for the transmission to be lossless, she has to transmit the base length $\underline{L}_c(x_i)$ of each code word corresponding to the source message $|x_i\rangle$. The base length is at least as long as the expected code length of the code word, hence

$$\underline{L}_c(x_i)\geqslant\langle x_i|\hat{L}_c|x_i\rangle. \tag{69}$$

Now we are interested in the average base length, since this determines the compression rate. The average base length is bounded from below by

$$\underline{L}_c(\Sigma)=\sum_{i=1}^{m} p(x_i)\underline{L}_c(x_i) \tag{70}$$

$$\geqslant\sum_{i=1}^{m} p(x_i)\langle x_i|\hat{L}_c|x_i\rangle=\mathrm{Tr}\{\sigma\hat{L}_c\} \tag{71}$$

$$=\sum_{i=1}^{m}\sigma_{ii}L_c(\omega_i). \tag{72}$$

Now we perform the following trick. As already stated, nonprefix code words can be chosen shorter than (or at most as long as) prefix code words. Consider an arbitrary prefix code c', then

$$L_{c'}(\omega_i)=L_c(\omega_i)+l_{c'}(\omega_i)\geqslant L_c(\omega_i), \tag{73}$$

where $l_{c'}(\omega_i)\geqslant 0$ is the length difference between the prefix and the nonprefix code word for $|\omega_i\rangle$. Prefix codes, just like all uniquely decodable symbol codes, have to fulfill the Kraft inequality [11,12]

$$\sum_{i=1}^{d} k^{-L_{c'}(\omega_i)}\leqslant 1. \tag{74}$$

Since the code-length operator $\hat{L}_{c'}$ is orthogonal in the basis $\mathcal{B}_\mathcal{V}$, we can express the above condition by the quantum Kraft inequality

$$\mathrm{Tr}_\mathcal{V}\{k^{-\hat{L}_{c'}}\}\leqslant 1, \tag{75}$$

where $\hat{L}_{c'}:=\hat{L}_c+\hat{l}_{c'}$ and

$$\hat{l}_{c'}:=\sum_{i=1}^{d} l_{c'}(\omega_i)|\omega_i\rangle\langle\omega_i|. \tag{76}$$

The quantum Kraft inequality was derived for the first time by Schumacher and Westmoreland [8]. Here, the quantum Kraft inequality requires that

$$Q:=\sum_{i=1}^{d} k^{-L_c(\omega_i)-l_{c'}(\omega_i)}\leqslant 1. \tag{77}$$

Now define implicit probabilities

$$q(\omega_i):=\frac{1}{Q}k^{-L_c(\omega_i)-l_{c'}(\omega_i)}, \tag{78}$$

which can be rewritten as

$$L_c(\omega_i)=-\log_k q(\omega_i)-\log_k Q-l'(\omega_i). \tag{79}$$

Summing over σ_{ii} yields

$$\sum_{i=1}^{d}\sigma_{ii}L_c(\omega_i)=-\sum_{i=1}^{d}\sigma_{ii}\log_k q(\omega_i)-\log_k Q-l', \tag{80}$$

where

$$l':=\sum_{i=1}^{d}\sigma_{ii}l_{c'}(\omega_i)=\mathrm{Tr}\{\sigma\hat{l}_{c'}\} \tag{81}$$

is the average additional length. The inequality (72) can now be expressed by

$$\underline{L}_c(\Sigma)\geqslant-\sum_{i=1}^{d}\sigma_{ii}\log_k q(\omega_i)-\log_k Q-l'. \tag{82}$$

Gibbs' inequality implies that

$$\underline{L}_c(\Sigma)\geqslant-\sum_{i=1}^{d}\sigma_{ii}\log_k\sigma_{ii}-\log_k Q-l'. \tag{83}$$

The von Neumann entropy of the message matrix σ cannot decrease by a nonselective projective measurement in the basis $\mathcal{B}_\mathcal{V}$, hence

$$S(\sigma) \leqslant S(\sigma'), \tag{84}$$

where

$$\sigma' := \sum_{i=1}^{d} |\omega_i\rangle\langle\omega_i|\sigma|\omega_i\rangle\langle\omega_i| = \sum_{i=1}^{d} \sigma_{ii}|\omega_i\rangle\langle\omega_i|. \tag{85}$$

Since

$$S(\sigma') = -\sum_{i=1}^{d} \sigma_{ii} \log_2 \sigma_{ii} = -\log_2 k \sum_{i=1}^{d} \sigma_{ii} \log_k \sigma_{ii}, \tag{86}$$

relation (84) states that

$$-\sum_{i=1}^{d} \sigma_{ii} \log_k \sigma_{ii} \geqslant \frac{1}{\log_2 k} S(\sigma). \tag{87}$$

Using Eq. (87) together with the Kraft inequality $Q \leqslant 1$, relation (83) transforms into

$$\log_2(k)\{\underline{L}_c(\Sigma) + l'\} \geqslant S(\sigma) - \log_k Q \geqslant S(\sigma). \tag{88}$$

Recalling the definition of the code information (56) and defining the length information that can be drawn into the classical side channel by

$$I' := \log_2(k) l', \tag{89}$$

we finally arrive at the lower-bound relation

$$\underline{I}_c(\Sigma) + I' \geqslant S(\sigma). \tag{90}$$

If c is a uniquely decodable symbol code, e.g., a prefix code, we have $I' = 0$. Inequality (90) states that the ensemble Σ can be losslessly compressed not below $S(\sigma)$ qubits. However, by drawing length information into a classical side channel, it is possible to reduce the average number of qubits passing through the quantum channel *below* the von Neumann entropy. We will give an example later on where this really happens.

2. Upper bound

Let us look for an upper bound for the compression that can be achieved. In order to encode every source vector in $\mathcal{V}$ by a k-ary code, we need at most

$$\underline{L}_c(x) \leqslant \lceil \log_k(\dim \mathcal{V}) \rceil \leqslant \log_k(\dim \mathcal{V}) + 1 \tag{91}$$

digits. Using $\log_a x = \log_a b \cdot \log_b x$, we have

$$\underline{I}_c(\Sigma) \leqslant \log_2(\dim \mathcal{V}) + \log_2 k. \tag{92}$$

This upper bound is neither very tight nor is it related to the von Neumann entropy. However, our efforts to find a more interesting upper bound were not successful. It remains an open question to find such a bound and hence a quantum-mechanical generalization to Shannon's theorem [15],

$$H(\Sigma) \leqslant I_c(\Sigma) \leqslant H(\Sigma) + \log_2 k, \tag{93}$$

which looks more familiar for $k = 2$, such that $\log_2 k = 1$ and $I_c(\Sigma) = L_c(\Sigma)$.

D. Quantum Morse codes

One way to avoid a classical side channel is to leave a *pause* between the quantum code words, which equals an additional orthogonal "comma state." Such a code is a quantum analog to the Morse code, where the code words are also separated by a pause, in order to avoid prefix code words. Of course, the code words *plus* the pause are prefix. Due to the close analogy one could speak of quantum Morse codes. Here, the information I' needed for the comma state is independent of the statistics, because the comma state must be sent after each letter code word, no matter which one. In contrast to that, I' is, in general, dependent on the statistics. If one transmits the length of each code word through a classical side channel, one can use a Huffman code to find shorter code words for more frequent length values. This is done in the following compression scheme.

VII. A LOSSLESS COMPRESSION SCHEME

Let us construct an explicit coding scheme that realizes lossless quantum compression.

A. Preparations

Alice and Bob have a quantum computer on both sides of the channel. They both allocate a register of r k-ary quantum digits, whose physical space is given by $\mathcal{R} = \mathcal{D}^{\otimes r}$ with $\mathcal{D} = \mathbb{C}^k$. They agree to use neutral-prefix code words (see Sec. III B 3) to implement variable-length coding, hence the message space is $\mathcal{N}_r$ of dimension k^r and is physically realized by the operational space $\tilde{\mathcal{N}}_r = \mathcal{R}$. Alice is preparing source messages $|x_i\rangle$, $i = 1, \ldots, m$ from a set $\mathcal{X}$. The space spanned by these messages is the source space $\mathcal{V} = \mathrm{Span}(\mathcal{X})$. Alice prepares each message $|x\rangle \in \mathcal{X}$ with probability $p(x)$, which gives the ensemble $\Sigma := \{p, \mathcal{X}\}$. She encodes the source messages into variable-length code words $|c(x)\rangle \in \mathcal{N}_r$ of maximal length r. If the dimension of $\mathcal{V}$ is given by $d := \dim \mathcal{V}$, then the length of the register must fulfill

$$r \geqslant \lceil \log_k d \rceil. \tag{94}$$

If the set $\mathcal{X}$ is linearly dependent, Alice creates a set $\tilde{\mathcal{X}} = \mathcal{X}$, removes the most probable message from $\tilde{\mathcal{X}}$ and puts it into a list **M**. Next, she removes again the most probable message from $\tilde{\mathcal{X}}$, appends it to the list **M** and checks if the list is now linearly dependent. If so, she removes the last element from **M** again. Then she proceeds with removing the next probable message from $\tilde{\mathcal{X}}$ and appending it to **M**, checking for linear dependence, and so on. In the end she obtains a list

$$\mathbf{M} = (|x_1\rangle, \ldots, |x_d\rangle) \tag{95}$$

of linearly independent source messages from $\mathcal{X}$, ordered by decreasing probability, such that $p(x_i) \geqslant p(x_j)$ for $i \leqslant j$. She performs a Gram-Schmidt orthonormalization on the list **M**, giving a list **B** of orthornormal vectors $|\omega_i\rangle$, defined by

$$|\omega_1\rangle := |x_1\rangle, \tag{96}$$

$$|\omega_i\rangle := N_i \left[1 - \sum_{j=1}^{i-1} |\omega_j\rangle\langle\omega_j| \right] |x_i\rangle, \tag{97}$$

with $i=2,\ldots,d$ and suitable normalization constants N_i. The elements of **B** form an orthonormal basis $\mathcal{B}_\mathcal{V}$ for the source space $\mathcal{V}$. Now she assigns code words

$$|c(\omega_i)\rangle := |Z_k^r(i-1)\rangle, \quad i=1,\ldots,d \tag{98}$$

of increasing significant length

$$L_c(\omega_i) = \lceil \log_k(i) \rceil. \tag{99}$$

Note that the first code word is the empty message $|\varnothing\rangle = |Z_k^r(0)\rangle = |0\cdots 0\rangle$, which does not have to be sent through the quantum channel at all. Instead, nothing is sent through the quantum channel and a signal representing ''length 0'' is sent through the classical channel. Alice implements the encoder

$$C := \sum_{i=1}^{d} |c(\omega_i)\rangle\langle\omega_i| \tag{100}$$

by a gate array on $\mathcal{R}$. Then she calculates the base lengths of the code words

$$\underline{L}_c(x) = \max_{i=1,\ldots,d} \{ L_c(\omega_i) \,|\, |\langle\omega_i|x\rangle|^2 > 0 \} \tag{101}$$

for every message $|x\rangle \in \mathcal{X}$ and writes them into a table. The classical information is compressed using Huffman coding of the set of distinct base-length values $\mathcal{L} = \{L_c(\omega_1),\ldots,L_c(\omega_d)\}$. Alice constructs the Huffman code word to each length $l \in \mathcal{L}$ appearing with probability

$$p_l = \sum_{x:\underline{L}_c(x)=l} p(x), \tag{102}$$

and writes them into a table. At last, Alice builds a gate array realizing the decoder $D = C^{-1}$ and gives it to Bob. For the classical channel she hands the table with the Huffman code words for the distinct lengths to Bob. Now everything is prepared and the communication can begin.

B. Communication protocol

Alice prepares the message $|x\rangle \in \mathcal{X}$ and applies the encoder C to obtain $|c(x)\rangle$. She looks up the corresponding code base length $L_c(x)$ in the table. If $L_c(x) < r$, she truncates the message to $L_c(x)$ digits by removing $r - L_c(x)$ leading digits. She sends the $L_c(x)$ digits through the quantum channel and the length information $L_c(x)$ through the classical channel. Then she proceeds with the next message.

For any message $|x\rangle$ Alice sends, Bob receives the length information $L_c(x)$ through the classical channel and $L_c(x)$ quantum digits through the quantum channel. He adds $r - L_c(x)$ digits in the state $|0\rangle$ at the beginning of the received code word. He then applies the decoder D and obtains the original message $|x\rangle$ with perfect fidelity. Note that Alice can send *any* message from the source message space $\mathcal{V}$, the protocol will ensure a lossless communication of the message. For such arbitrary messages, however, compression will, in general, not be achieved, since the protocol is only adapted to the particular ensemble Σ. Also, Bob can as well store all received quantum digits on his quantum hard disk and the received length information on his classical hard disk, and go to bed. The next day, he can scan the classical hard disk for length information and separate and decode the corresponding code words on the quantum hard disk. The protocol works as well for online communication as for data storage.

C. An explicit example

Alice and Bob want to communicate vectors of a four-dimensional Hilbert space $\mathcal{V} = \mathrm{Span}\{|0\rangle,|1\rangle,|2\rangle,|3\rangle\}$, where we use the row notation in the following. Alice decides to use the (linearly dependent) source message set

$$\mathcal{X} = \{|a\rangle,|b\rangle,|c\rangle,|d\rangle,|e\rangle,|f\rangle,|g\rangle,|h\rangle,|i\rangle,|j\rangle\}, \tag{103}$$

whose elements are given by

$$|a\rangle = \frac{1}{2}1,1,1,1, \tag{104}$$

$$|b\rangle = \frac{1}{\sqrt{5}}(1,2,1,1), \tag{105}$$

$$|c\rangle = \frac{1}{\sqrt{6}}(1,3,1,1), \tag{106}$$

$$|d\rangle = \frac{1}{\sqrt{7}}(1,4,1,1), \tag{107}$$

$$|e\rangle = \frac{1}{\sqrt{2}}(1,0,1,0), \tag{108}$$

$$|f\rangle = \frac{1}{\sqrt{3}}(2,0,1,0), \tag{109}$$

$$|g\rangle = \frac{1}{2}(3,0,1,0), \tag{110}$$

$$|h\rangle = \frac{1}{\sqrt{2}}(0,1,0,1), \tag{111}$$

$$|i\rangle = \frac{1}{\sqrt{3}}(0,2,0,1), \tag{112}$$

$$|j\rangle = \frac{1}{2}(0,3,0,1) \tag{113}$$

and which are used with the probabilities

$$p(a)=0.6, \quad p(b)=p(c)=p(d)=0.1, \tag{114}$$

$$p(e)=\cdots=p(j)=\frac{0.3}{3}. \tag{115}$$

The Shannon entropy of the ensemble $\Sigma=\{p,\mathcal{X}\}$ is

$$H(\Sigma)=2.029\,45, \tag{116}$$

and the classical raw information (45) reads

$$I_0(\mathcal{X})=\log_2|\mathcal{X}|=3.321\,93, \tag{117}$$

which gives an optimal classical compression rate of $R=H/I_0=0.610\,924$. If Bob knows Alice's list of possible messages, then this rate could in the optimal case be achieved by pure classical communication. However, Bob does not know the list and classical communication is not the task here. The message matrix $\sigma=\Sigma_{x\in\mathcal{X}}p(x)|x\rangle\langle x|$ given by

$$\sigma=\begin{pmatrix} 0.214\,549 & 0.224\,624 & 0.197\,882 & 0.177\,882 \\ 0.224\,624 & 0.403\,02 & 0.224\,624 & 0.244\,624 \\ 0.197\,882 & 0.224\,624 & 0.191\,216 & 0.177\,882 \\ 0.177\,882 & 0.244\,624 & 0.177\,882 & 0.191\,216 \end{pmatrix} \tag{118}$$

has von Neumann entropy

$$S(\sigma)=0.571\,241. \tag{119}$$

The orthogonalization procedure yields the basis $\mathcal{B}_{\mathcal{V}}=\{|\omega_i\rangle\}$ with

$$|\omega_1\rangle=(0.5,0.5,0.5,0.5), \tag{120}$$

$$|\omega_2\rangle=(-0.288\,675,0.866\,025,-0.288\,675,-0.288\,675), \tag{121}$$

$$|\omega_3\rangle=(0.408\,248,0,0.408\,248,-0.816\,497), \tag{122}$$

$$|\omega_4\rangle=(0.707\,107,0,-0.707\,107,0). \tag{123}$$

Let the quantum channel be binary, i.e., let $k=2$. The code words are constructed along $|c(\omega_i)\rangle=|Z_2(i-1)\rangle$, yielding the variable-length states

$$|c(\omega_1)\rangle=|\varnothing\rangle, \tag{124}$$

$$|c(\omega_2)\rangle=|1\rangle, \tag{125}$$

$$|c(\omega_3)\rangle=|10\rangle, \tag{126}$$

$$|c(\omega_4)\rangle=|11\rangle \tag{127}$$

that span the code space $\mathcal{C}$. In a neutral-prefix code they are realized by the two-qubit states

$$|\tilde{c}(\omega_1)\rangle=|00\rangle, \tag{128}$$

$$|\tilde{c}(\omega_2)\rangle=|01\rangle, \tag{129}$$

$$|\tilde{c}(\omega_3)\rangle=|10\rangle, \tag{130}$$

$$|\tilde{c}(\omega_4)\rangle=|11\rangle \tag{131}$$

that span the operational code space $\tilde{\mathcal{C}}$, which is a subspace of the physical space $\mathcal{R}=\mathbb{C}^2\otimes\mathbb{C}^2$. Alice realizes the encoder $C:\mathcal{V}\rightarrow\tilde{\mathcal{C}}$, $C=\Sigma_i|\overline{c}(\omega_i)\rangle\langle\omega_i|$ given by

$$C=\begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ -0.288\,675 & 0.866\,025 & -0.288\,675 & -0.288\,675 \\ 0.408\,248 & 0 & 0.408\,248 & -0.816\,497 \\ 0.707\,107 & 0 & -0.707\,107 & 0 \end{pmatrix} \tag{132}$$

and the decoder $D=C^{-1}$ given by

$$D=\begin{pmatrix} 0.5 & 0.408\,248 & -0.288\,675 & 0.707\,107 \\ 0.5 & 0 & 0.866\,025 & 0 \\ 0.5 & 0.408\,248 & -0.288\,675 & -0.707\,107 \\ 0.5 & -0.816\,497 & -0.288\,675 & 0 \end{pmatrix} \tag{133}$$

by gate arrays and gives the decoder to Bob. The encoded alphabet is obtained by $|c(x)\rangle=C|x\rangle$. Alice writes the base lengths of the code words

$$\underline{L}_c(a)=0, \quad \underline{L}_c(b)=\underline{L}_c(c)=\underline{L}_c(d)=1, \tag{134}$$

$$\underline{L}_c(e)=\cdots=\underline{L}_c(j)=2 \tag{135}$$

in a table and calculates the corresponding probabilities

$$p_0=0.6, \quad p_1=0.3, \quad p_2=0.1. \tag{136}$$

She constructs Huffman code words for each length

$$c_0=1, \quad c_1=01, \quad c_2=00, \tag{137}$$

such that the average bit length is

$$L'=\sum_{l=0}^{2} p_l l=1.4, \tag{138}$$

which is the optimal value next to the Shannon entropy of the length ensemble,

$$I'=-\sum_{l=0}^{2} p_l \log_2 p_l=1.295\,46. \tag{139}$$

Alice hands the table with the Huffman code words to Bob and tells him that he must listen to the classical channel, decode the arriving Huffman code words into numbers, receive packages of qubits, whose size corresponds to the decoded numbers, and add to each package enough leading qubits in the state $|0\rangle$ to end up with two qubits. Then he must apply the decoder D to each extended package and he will get Alice's original messages.

Say, Alice wants to send the message $|a\rangle$. She prepares $|a\rangle$ and applies the encoder C to obtain the code word $|00\rangle$. She looks up the corresponding base length $\underline{L}_c(a)=0$ and truncates the code word to $\underline{L}_c(a)=0$ qubits. In this case there are no qubits left at all, so she sends nothing through the quantum channel and the Huffman code word for ''length 0'' through the classical channel. Bob receives the classical length information ''0'' and knows that nothing comes through the quantum channel and that in this case he has to prepare two qubits in the state $|00\rangle$. He applies the decoder D and obtains Alice's original message $|a\rangle$. In order to send message $|b\rangle$, Alice truncates the code word to $L_c(b)=1$ qubit and obtains $(1/\sqrt{2})(|0\rangle+|1\rangle)$. She sends the qubit through the quantum channel together with the classical signal ''length 1.'' Bob receives the length message and knows that he has to take the next qubit from the quantum channel and that he has to add one leading qubit in the state $|0\rangle$. He applies D and obtains Alice's original message $|b\rangle$. The whole procedure works instantaneously and without loss of information. We have implemented the above example by a MATHEMATICA™ program, and numerical simulations show that the procedure works fine and the specified compression of quantum data is achieved. (You can find the program and the package at [16]).

Let us look for the compression that has been achieved.

(a) The quantum code information, i.e., the average number of qubits being sent through the quantum channel,

$$\underline{I}_c=\sum_{x\in\mathcal{X}} p(x)\underline{L}_c(x)=0.5, \tag{140}$$

falls below the von Neumann entropy

$$\underline{L}_c<S=0.571\,241. \tag{141}$$

Such a behavior has already been suspected in Sec. VI C 1.

(b) The quantum raw information, i.e., the size of the noncompressed messages, is given by

$$\underline{I}_c<I_0=\log_2(\dim\,\mathcal{V})=2, \tag{142}$$

hence the compression rate on the quantum channel reads

$$R_c=\frac{\underline{I}_c}{I_0}=0.25. \tag{143}$$

In other words, the number of qubits passing through the quantum channel is reduced by 75%. Sending 100 messages without compression requires 200 qubits. Using the compression scheme, Alice typically sends 50 qubits.

(c) The sum of both quantum and classical information,

$$I_{\mathrm{tot}}=\underline{I}_c+I'=1.795\,46, \tag{144}$$

is smaller than the Shannon entropy (116) of the original ensemble Σ,

$$I_{\mathrm{tot}}<H=2.029\,45. \tag{145}$$

Thus it is better to use the quantum-compression scheme than to simply *tell* Bob on the phone, which state he must prepare. As already suspected, I_{tot} is still greater than the von Neumann entropy (119),

$$I_{\mathrm{tot}}>S=0.571\,241. \tag{146}$$

The classical part of the compression depends on the algorithm. Only in the ideal case, the information can be compressed down to the Shannon entropy of the length ensemble given by I'. Using the Huffman scheme, the average length $L'=1.4$ represents the information that is effectively sent through the classical channel, such that the total *effective* information is given by

$$I_{\mathrm{eff}}=\underline{I}_c+L'=1.9. \tag{147}$$

(d) The total compression rate of both channels reads

$$R_{\mathrm{tot}}=\frac{\underline{I}_c+I'}{I_0}=0.897\,731<1, \tag{148}$$

where it is assumed that the information on the classical channel can be compressed down to its Shannon entropy I'. Using the Huffman scheme (as we have done in our example), the information on the classical channel can only be compressed to $L'>I'$, such that the *effective* total compression rate is given by

$$R_{\mathrm{eff}}=\frac{\underline{I}_c+L'}{I_0}=0.95<1. \tag{149}$$

Thus in any case there is an overall compression. For higher-dimensional source spaces (hence more letters), the compression is expected to get better (provided the letter distribution is not too uniform). However, the numerical effort for higher-dimensional letter spaces increases very fast and we want to keep the example as simple as possible.

VIII. CONCLUDING REMARKS

We have developed a general framework for variable-length quantum messages and have defined an observable measuring the quantum-information content of individual states by the number of qubits needed to represent the state by a given code. We derived some basic statements about lossless compression. In particular, we have demonstrated that a quantum message can only be compressed without loss of information if the source message is *a priori* known to the sender. On these grounds, we have worked out a lossless and instantaneous quantum data-compression protocol. One can object that there is no use in compressing quantum states that are already *known* to the sender, because then Alice could as well tell Bob classically which of the quantum states she wants to communicate. However, such a pure classical communication would require Bob to have a list of possible messages Alice may send. Moreover, for *arbitrary* quantum messages from the source space, Alice would need infinitely many bits to communicate them through a classical channel to Bob. In contrast to that, in our communication scheme

Alice *can* send arbitrary messages from the source message space, but she must know which message she is going to send to get the base length. Bob needs only the decoder and the user instructions for the classical channel, then he can reobtain Alice's original messages with perfect fidelity. The protocol can individually be adapted to a given message ensemble such that compression is achieved for that ensemble.

IX. OPEN QUESTIONS

It would be satisfying to find an *optimal* compressing lossless quantum code with a tight upper bound related to the von Neumann entropy. This would represent a quantum analog to Shannon's relation (93). There might be interesting applications to quantum cryptography. By combining the methods of quantum cryptography with the methods of lossless compression, the efficiency of secure data transfer may possibly be increased. Furthermore, it would be interesting to see how the framework of variable-length messages applies to quantum computation, since the data stored in the register of a quantum computer could also be regarded as a variable-length quantum message. One could also think about variable-length quantum error-correcting codes. We hope that the presented work stimulates some more discussion and theoretical research on variable-length quantum coding and its applications.

ACKNOWLEDGMENTS

We would like to thank Martin Wilkens, Jens Eisert, and Alexander Albus for inspiring discussions and supervisory advice. This work is supported by the EU project EQUIP, the International Max Planck Research School IMPRS, and the Deutsche Forschungsgemeinschaft DFG.

[1] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[2] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[3] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).
[4] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. A **54**, 4707 (1996).
[5] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, e-print quant-ph/9805017.
[6] S. L. Braunstein, C. A. Fuchs, D. Gottesman, and H.-K. Lo, e-print quant-ph/9805080.
[7] I. L. Chuang and D. S. Modha, IEEE Trans. Inf. Theory **46**, 1104 (2000).
[8] B. Schumacher and M. D. Westmoreland, e-print quant-ph/0011014.
[9] K. Bostroem, e-print quant-ph/0009052.
[10] K. Bostroem, e-print quant-ph/0009073.
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[12] D. J. C. MacKay, Information *Theory, Inference, and Learning algorithms*, URL: http://wol.ra.phy.cam.ac.uk/mackay/itprnn/book.html
[13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2001).
[14] J. Preskill, *Lecture Notes.*, URL: http://www.theory.caltech.edu/people/preskill/ph219/
[15] C. E. Shannon and W. Weaver, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[16] T. Felbinger, QMATRIX-A MATHEMATICA Package for Quantum Information, URL: http://www.quantum.physik.uni-potsdam.de/Timo_Felbinger/qmatrix
[17] Some notions and definitions already exist, some are based on our own reasoning. When we find an already existing definition equal or similar to the desired one, we use it and in case it is not a standard definition, we give an explicit reference. For a profound review on classical information theory, see [11,12]; for a profound review on quantum information theory, see [13,14].