On Lossless Quantum Data Compression with a Classical Helper

Rudolf Ahlswede and Ning Cai *

Abstract

After K. Boström and T. Felbinger observed that lossless quantum data compression does not exist unless decoders know the lengths of codewords, they introduced a classical noiseless channel to inform the decoder of a quantum source about the lengths of codewords.

In this paper we analyse their codes and present

1) a sufficient and necessary condition for the existence of such codes for given lists of lengths of codes

2) a characterization of the optimal compression rate for their codes.

However our main contribution is a more efficient way to use the classical channel. We propose a more general coding scheme. It turned out that the optimal compression can always be achieved by a code obtained by this scheme.

A *von Neumann entropy* lower bound to rates of our codes and a necessary and sufficient condition to achieve the bound are obtained. The gap between this lower bound and the compression rates is also well analysed.

For a special family of quantum sources we provide a sharper lower bound in terms of *Shannon entropy*.

Finally we propose some problems for further research.

Index terms: quantum source, lossless data compression, classical helper, quantum–variable–length codes, von Neumann entropy bound.

*This paper is supported in part by INTAS 00–738

1 Introduction

Since B. Schumacher extended Shannon's Source Coding Theorem to quantum sources in his well known work [12], the research on lossy quantum data compression got an impetus. However the extension of lossless data compression to quantum is impossible, because a length measurement performed at a codeword of a quantum variable–length code will destroy the codeword. This observation (Observation I for later reference) was made by many authors, e.g. [3], [14] and [2].

Consequently, one cannot compress quantum data by encoding them to a quantum variable–length code that can be decoded by the decoder, unless the decoder knows the length of sent codeword. In other words, there is no way to compress quantum data and decode them losslessly by using only a quantum source code.

Nevertheless quantum variable–length codes have been studied by several authors. In particular, Kraft's inequality has been established by B. Schumacher and M.P. Westmoreland in [14].

Some authors [3, 14] apply quantum variable–length codes to construct long codes in *lossy* quantum data compression. Along another line K. Boström and T. Felbinger [2] introduced a classical noiseless channel to inform the decoder about the lengths of codewords. The main goal of this paper is to discuss lossless quantum data compression in the presence of a classical, noiseless helper channel whose use is not restricted. We begin with definitions to prepare the discussion.

Let $\mathcal{H}$ be a Hilbert space of finite dimension d and let

$$\mathcal{B}(\mathcal{H}) = \{|i\rangle : i = 0, 1, 2, \ldots, d-1\} \tag{1}$$

be an orthonormal basis of $\mathcal{H}$.

Denote by $\mathcal{H}^{\otimes n}$ the nth tensor power of the Hilbert space $\mathcal{H}$. For $\ell = 1, 2, \ldots, \ell_{\max}$ let $\mathcal{H}^{\otimes \ell}$ be a set of pairwise orthogonal (sub)spaces (in a sufficiently large Hilbert space). Then we can define the direct sum

$$\mathcal{H}^{\oplus \ell_{\max}} = \mathcal{H} \oplus \mathcal{H}^{\otimes 2} \oplus \cdots \oplus \mathcal{H}^{\otimes \ell_{\max}}, \tag{2}$$

a Hilbert space of dimension $\sum_{\ell=1}^{\ell_{\max}} d^{\ell}$.

Throughout the paper we consider a quantum information source generating pure normal states $|s\rangle$ in a Hilbert space $\mathcal{S}$ of finite dimension d'. We often call $\mathcal{S}$ the source space. As was defined in [2], a lossless variable–length encoder $\mathcal{E}$ of maximal length $\ell_{\max}$ is a linear isometric operator from $\mathcal{S}$ to a subspace $\mathcal{C} \subset \mathcal{H}^{\oplus \ell_{\max}}$ of dimension d' i.e., for all $|s\rangle$, $|s'\rangle \in \mathcal{S}$ $\langle \mathcal{E}(s)|\mathcal{E}(s')\rangle = \langle s|s'\rangle$, where $|\mathcal{E}(s'')\rangle = \mathcal{E}(|s''\rangle)$. $\mathcal{C}$ is called the codeword space and the (normalized) vectors (i.e. states) in it are called codewords. To realize the coding procedure B. Schumacher and M. Westmoreland [14] introduced the zero–extended form of a code $\mathcal{C}$. This is the set of states in $\mathcal{H}^{\otimes \ell_{\max}}$ obtained by appending $|\ell_{\max} - \ell|$ $|0\rangle$'s at the codewords in $\mathcal{C} \cap \mathcal{H}^{\otimes \ell}$ for all $0 < \ell \leq \ell_{\max}$, i.e. the set

$$\left\{|\gamma^{\ell} 0^{\ell_{\max}-\ell}\rangle : |\gamma^{\ell}\rangle \in \mathcal{C} \cap \mathcal{H}^{\otimes \ell},\ \ell = 1, 2, \ldots, \ell_{\max}\right\}. \tag{3}$$

Similarly, to realize variable–length codes, K. Boström and T. Felbinger [2] put $|0^m1\rangle$ in front of codewords. As in the classical case, a code is called prefixfree if no codeword is a prefix of another codeword.

In Classical Information Theory the lengths of codewords in a variable–length code are determinate. For example, in the code $\{0, 10, 11\}$ the codewords $0, 10, 11$ have lengths $1, 2, 2$ respectively whereas the length of codewords in a quantum variable–length code are indeterminate because of superposition. Namely, for a vector $(a_1, a_2, \ldots, a_{\ell_{\max}}) \in \mathbb{C}^{\ell_{\max}}$, with $\sum_{\ell=1}^{\ell_{\max}} a_\ell^2 = 1$ and $|\gamma^\ell\rangle \in \mathcal{C} \cap \mathcal{H}^{\otimes\ell}$, $\sum_{\ell=1}^{\ell_{\max}} a_\ell|\gamma^\ell\rangle$ is a codeword because the encoder mapping is linear. Thus B. Schumacher and M. Westmoreland refer to these codes as "indeterminate length codes". One way to measure the lengths of codewords in this case is as follows ([14] and [2]). Let $\mathcal{H}^{\oplus\ell_{\max}}$ be the Hilbert space in (2) and let $\mathcal{P}_\ell$ be the projection of $\mathcal{H}^{\oplus\ell_{\max}}$ onto $\mathcal{H}^{\otimes\ell}$ for $\ell = 1, 2, \ldots, \ell_{\max}$. Then the observable $\mathcal{L} = \{\mathcal{P}_\ell\}$, where $\mathcal{P}_\ell$ corresponds to the outcome ℓ, is called the length observable. Thus with probability $tr(|w\rangle\langle w|\mathcal{P}_\ell) = \langle w|\mathcal{P}_\ell|w\rangle = |a_\ell|^2$ the outcoming length of a codeword $|w\rangle = \sum_{\ell=1}^{\ell_{\max}} a_\ell|\gamma^\ell\rangle$, $|\gamma^\ell\rangle \in \mathcal{H}^{\otimes\ell}$ is ℓ when one measures the codeword with $\mathcal{L}$. Let

$$\Lambda = \sum_{\ell=1}^{\ell_{\max}} \ell\mathcal{P}_\ell. \tag{4}$$

Then the expected outcoming length of a codeword $|w\rangle$ is

$$\overline{L}(|w\rangle) = tr(|w\rangle\langle w|\Lambda) = \langle w|\Lambda|w\rangle, \tag{5}$$

which we call the average length of codeword $|w\rangle$.

With this notation B. Schumacher and M. Westmoreland [14] presented a

Quantum Kraft Inequality: For all quantum uniquely decodable codes $\mathcal{C}$

$$\sum_{\ell=1}^{\ell_{\max}} \dim(\mathcal{C} \cap \mathcal{H}^{\otimes\ell})d^{-\ell} \leq 1, \tag{6}$$

where $d = \dim(\mathcal{H})$.

From this they deduced a von Neumann entropy bound.

An important parameter, the base length $L(|w\rangle)$ of a codeword $|w\rangle$ in a quantum variable length code, was introduced in [2]:

$$L(|w\rangle) = \max\{\ell : \langle w|\mathcal{P}_\ell|w\rangle \ > 0\}. \tag{7}$$

That is, $L(|w\rangle)$ is the largest ℓ such that $a_\ell \neq 0$ if $|w\rangle$ is a superposition $|w\rangle = \sum_\ell a_\ell|\gamma^\ell\rangle$, $|\gamma^\ell\rangle \in \mathcal{C} \cap \mathcal{H}^{\otimes\ell}$. It is clear that for all codewords $|w\rangle$

$$\overline{L}(|w\rangle) \leq L(|w\rangle). \tag{8}$$

In order to decode a quantum variable–length code without error, a decoder has to know the base length of the sent codeword. For this reason K. Boström and T. Felbinger introduced a classical channel in [2]. The following are the assumptions for their codes:

1) Visible Quantum Encoding: Suppose the encoder needs to encode the output states from the source space $\mathcal{S}$ of dimension d. He does this by a linear isometric operator from $\mathcal{S}$ to a subspace $\mathcal{C}$ of $\mathcal{H}^{\oplus \ell_{\max}}$ (c.f. Subsection 1.1). The encoding is visible, that is the encoder knows the output state of the quantum source and therefore the base length of the codeword to which the output state is encoded, say $\ell_{\mathcal{B}}$.

2) Classical Channel: Now the encoder knows $\ell_{\mathcal{B}}$ and has to inform the decoder about it. This is done via the classical channel.

We note that the classical channel in their model is only used to inform about the lengths of codewords.

In the next section we analyse the codes introduced in [2]. An important question for variable–length codes concerns the existence of codes for a given list of lengths of codewords. This question is answered in Classical Information Theory by Kraft's inequality. Since a classical channel is present, now the codes in [2] are not necessarily uniquely decodable and therefore (6) may not hold. K. Boström and T. Felbinger established a Kraft–type inequality in [2] with an additional term that depends on how to extend the quantum variable–length code to a uniquely decodable code and consequently on the structure of the particular code. Because of this dependence, the inequality seems not easily to be usable to verify the existence of codes for given lists of lengths of codewords. In Subsection 2.1 we first present a simple and more transparent necessary and sufficient condition for their existence. Based on the analysis of this condition, we then present a class of realizable codes which we call canonical codes, because their members have a nice form and all codes in the discussion are isomorphic to a member in that class. We must also point out that there is essentially no difference between our canonical codes and "the natural–prefix codes" in [2], because they can be matched by isomorphisms. This shows that one cannot do better. Probably the most important problem for data compression are characterizations of optimal compression rates. We solve this for the codes in [2] by Theorem 1 in Subsection 2.3. The characterization in Theorem 1 is then discussed. Section 2 concludes with simple observations to show that the visibility assumption for encoding is necessary.

However, our main goal in the paper is to find a more efficient way to use the classical helper than just to report the base lengths. This covers the rest of the paper.

We begin with a simple example in Section 3 to show that the codes in [2] may not be optimal in general. This motivates us to find a more general coding scheme. We introduce a class of codes constructed by the coding scheme, which we call quantum–classical or $q-c$ variable–length codes. It turns out that for all quantum sources the optimal compression rates can be achieved by a $q-c$ variable–length code.

In Section 4 we continue the discussion of $q-c$ variable–length codes by providing a lower bound to their code rates in terms of von Neumann entropy, and a sufficient and necessary condition for codes to achieve it in Theorem 2. We then analyse the gap between the von Neumann entropy bound and the optimal rates. Our analysis shows that the von Neumann entropy bound is seldom tight and that the gap in general may be arbitrarily large. This suggests as a challenging problem that of finding a new quantity better fitting lossless quantum data compression than von Neumann entropy. (Notice that almost all information quantities

successfully applied in Quantum Information Theory are in terms of von Neumann entropy!)

In Section 5 we present a sharper lower bound in terms of Shannon entropy for a special class of quantum sources. By applying it to memoryless quantum sources we conclude that the von Neumann entropy bound may not be tight even in the asymptotic sense in lossless compression of data from memoryless quantum sources. This shows that lossless quantum data compression is completely different from both classical data compression and lossy quantum data compression. It gives us a further reason to doubt that von Neumann entropy well fits lossless quantum data compression.

Finally we present a few problems for future research in Section 6.

2 Code Analysis Based on the Base Length

In this section we analyse the codes by Boström and Felbinger in [2]. First let us briefly review some results of [2]. Let $\mathcal{H}$ be a Hilbert space of dimension d with basis $\mathcal{B}(\mathcal{H})$ in (1). Then Boström and Felbinger suggested the following coding scheme under the assumptions 1) and 2) in Section 1.

(I) Choose an orthonormal basis of source space $\{\beta(j) : j = 1, 2, \ldots, d'\}$ and encode $\beta(i)$ to $|b(\beta(i))\rangle = |b_{r_i}(i)b_{r-1}(i)\ldots b_1(i)\rangle$, where $|b_m(i)\rangle \in \mathcal{B}(\mathcal{H})$ for $\mathcal{B}(\mathcal{H})$ in (1), $m = 1, 2, \ldots, r_i$, and $r_i = \lfloor \log_d i \rfloor + 1$ if the d–ary representation of i is $b_{r_i}(i)b_{r-1}(i)\ldots b_1(i)$. Then $|b(\beta(i))\rangle$ has determinate length r_i. By linearity a superposition $\sum_j c_j\beta(j)$, $\sum_j c_j^2 = 1$ $c_j \neq 0$ of the states in the basis $\mathcal{B}(\mathcal{H})$ is encoded to $\sum_j c_j|b(\beta(j))\rangle$, a codeword whose base length is equal to the maximum lengths of the $|b(j)\rangle$'s involved in the superposition.

(II) To realize the code physically, we choose an integer $n \geq \lfloor \log_d d' \rfloor + 1$. Pad $n - r$ $|0\rangle$'s in front of a codeword obtained in (I) to get a state in $\mathcal{H}^{\otimes n}$ if the codeword has base length r. Then we have a linear isometric encoding operator from $\mathcal{S}$ to a subspace of $\mathcal{H}^{\otimes n}$.

(III) The encoder encodes the output of the quantum source to a codeword in $\mathcal{H}^{\otimes n}$ by the encoding operator and informs the decoder about the base length r of the codeword through the classical channel. Then the encoder deletes the prefix of length $n - r$ of the codeword and sends the remaining part to the decoder. Notice that all deleted "symbols" are $|0\rangle$'s.

(IV) The decoder pads $n - r$ $|0\rangle$'s to the received quantum state and then uses the inverse of the encoding operator to decode if he receives a base length r from the classical channel.

In particular, Boström and Felbinger proposed the following algorithm for the discrete quantum source $\{(P(x), |x\rangle) : x \in \mathcal{X}\}$, which outputs the state $|x\rangle$, $x \in \mathcal{X}$ with probability $P(x)$, where $\mathcal{X}$ is a finite set,

(a) Choose a basis $\{|x_1\rangle, |x_2\rangle, \ldots, |x_{d'}\rangle\}$ recursively as follows

(a_1) Choose an $|x_1\rangle$ such that $P(x_1) = \max_{x \in \mathcal{X}_1} P(x)$ for $\mathcal{X}_1 = \mathcal{X}$.

(a_i) Having chosen $|x_1\rangle, \ldots, |x_{i-1}\rangle$, one first deletes all $|x'\rangle$ in the subspace spanned by $\{|x_1\rangle, \ldots, |x_{i-1}\rangle\}$ from $\{|x\rangle : x \in \mathcal{X}\}$ and obtains a subset $\{|x''\rangle : x'' \in \mathcal{X}_i\}$, $\mathcal{X}_i \subset \mathcal{X}$. Then one chooses an $|x_i\rangle$ in $\mathcal{X}_i$ such that $P(x_i) = \max_{x'' \in \mathcal{X}_i} P(x'')$.

($a_{d'}$) The procedure is stopped at a vector $|x_{d'}\rangle$ such that $\mathcal{X}_{d'+1} = \varnothing$.

(b) Gram–Schmidt Orthonormalization: Obtain an orthonormal basis $\{|\beta_i\rangle : i = 1, 2, \ldots, d'\}$ from $\{|x_i\rangle : i = 1, 2, \ldots, d'\}$ by Gram–Schmidt orthonormalization.

(c) Encoding: Suppose $\dim(\mathcal{H}) = d$, and let $z_d(i)$ be the d–ary representation of number i and $w_d^{d'}(i)$ be the d–ary sequence of length d obtained by padding $d - \lceil \log_d i \rceil$'s 0 in front of $z_d(i)$ for $i = 1, 2, \ldots, d'$. Then encode $|\beta_i\rangle$ to $|w_d^{d'}(i)\rangle$.

(d) Remove the redundancy and inform about the base length: Now assume a state $|x\rangle = \sum_{i=1}^{j} c_i |\beta_i\rangle$ for $c_j \neq 0$ as output. Then by the previous step and the linearity of the encoder, we know $|s\rangle$ is encoded to a codeword $\sum_{i=1}^{j} c_i |w_d^{d'}(i)\rangle$, a codeword starting with r zeros for $r = d - \lceil \log_d j \rceil$, say. Then the encoder, who knows $|s\rangle$ and consequently j, removes the r zeros to obtain a codeword of base length $\ell = \lceil \log_d j \rceil$, say, and inform the decoder about ℓ via a classical channel. Notice that the resulting codeword after removing the redudancy can be stored in a d–ary quantum register of length ℓ.

(e) Decoding: The decoder pads $d - \ell$ zeros in front of the received (quantum) codeword and recovers the state $|s\rangle$ by the inverse of the (isometric) encoder in Step (c).

A similar coding scheme in [2] is to pad $|0^m 1\rangle$'s instead of $|0^m\rangle$.

In [2] the possibility of lossless quantum data compression and its physical realization are widely discussed. The following two conclusions from the discussion in [2] are related to our current work:

— A classical helper channel is necessary for lossless quantum data compression.

— The quantum variable–length codes are not necessarily uniquely decodable because of the presence of the classical channel.

Let $\mathcal{B}(\mathcal{S}) = \{\beta(j) : j = 1, 2, \ldots, d'\}$ be the orthonormal basis in (I) of the coding scheme, whose members are encoded to codewords $|b(\beta(j))\rangle$, with determinate length $\ell(\beta(j))$ $j = 1, 2, \ldots, d'$. In general $\{\ell(\beta(j)) : \beta(j) \in \mathcal{B}(\mathcal{S})\}$ does not satisfy Kraft's inequality, by the second conclusion above. Consequently an example that exceeds the von Neumann entropy bound was found in [2]. The example exists because the lengths of codewords carry information. To obtain a Kraft–type inequality, Boström and Felbinger extended

a non–uniquely decodable code $\{b(\beta(j)) : \beta(j) \in \mathcal{B}(\mathcal{S})\}$ to a prefix code with lengths $\ell^*_{(\beta(j))} = \ell(\beta(j)) + \ell'(\beta(j))$, for which a Kraft–type inequality $\sum_{\beta(j)} d^{-\ell^*(\beta(j))} \leq 1$ holds.

There is no constraint on the classical rate in the model in [2]. Therefore, throughout this section, without loss of generality, we simply assume that the decoder knows base lengths of codewords.

2.1 Canonical Codes

It is well–known in classical Information Theory that there exists a uniquely decodable variable–length code with a list of lengths of codewords $L = \{\ell_j : j = 1, 2, \ldots, J\}$ iff the list L satisfies the Kraft inequality. Now we are looking for a condition for existence of a quantum variable–length code with a classical helper channel informing about the base lengths. In [2] a Kraft–type inequality is used. We notice that the inequality contains an additional term depending on how one extends the considered code to a prefix code. Actually there is a simpler and more general relation for these codes. To see this, first consider the classical case. We analogously assume that the decoder knows the lengths of codewords sent by the encoder via an additional noiseless channel and therefore the variable–length code is not necessarily uniquely decodable.

Obviously under this assumption there exists a variable–length code with N_ℓ codewords of length ℓ over an alphabet of size q iff $N_\ell \leq q^\ell$. To obtain such a code one needs simply to take arbitrary N_ℓ sequences of length ℓ for all ℓ with $q^\ell \geq N_\ell > 0$ as codewords. To see its analogue in the quantum version let us consider a quantum–variable length code $\mathcal{C}$ as defined in Subsection 1.1 for a complex Hilbert space $\mathcal{H}$ of dimension d. We first look at the sets of codewords in $\mathcal{C}$ with base length ℓ for all possible ℓ. We find that they are not subspaces, because linear combinations of codewords with base length ℓ may be codewords with base length smaller than ℓ. So we turn to the sets of codewords of base length not larger than ℓ for all $\ell \leq \ell_{\max}$ and denote them by $\mathcal{C}_\ell$. Obviously $\mathcal{C}_\ell \subset \mathcal{C}_{\ell'}$ for $\ell < \ell' \leq \ell_{\max}$ and the $\mathcal{C}_\ell$'s are linear subspaces. Actually by definition

$$\mathcal{C}_\ell = \mathcal{C} \cap \mathcal{H}^{\oplus \ell} \tag{9}$$

for $\mathcal{H}^{\oplus \ell} = \mathcal{H} \oplus \mathcal{H}^{\otimes 2} \oplus \cdots \oplus \mathcal{H}^{\otimes \ell}$ (c.f. (2) in Subsection 1.1). First we assume that there is no constraint to the code except for the linearity. Then we have the following conditions for the existence of codes:

$$\mathcal{C}_1 \subset \mathcal{C}_2 \subset \cdots \subset \mathcal{C}_{\ell_{\max}} \tag{10}$$

and

$$\dim \mathcal{C}_\ell \leq \dim \mathcal{H}^{\oplus \ell} = \sum_{i=1}^{\ell} d^i. \tag{11}$$

In particular, one can take a code such that equality holds in (11) for $\ell = 1, 2, \ldots, \ell_{\max}$. To store a codeword of base length ℓ, one needs a quantum register of length $\sum_{i=1}^{\ell} d^i$. This is not what we would like to have, because we expect that a codeword of base length ℓ could

be stored in a quantum register of length ℓ! It is therefore assumed in [2] that we have to constrain

$$\dim \mathcal{C}_\ell \leq d^\ell \tag{12}$$

instead of (11). It is not hard to see the existence of a code $\mathcal{C}$ satisfying the conditions (10) and (12). Indeed to obtain such a code one can simply take the set of $\ell_{\max}$ orthogonal normal states, say Q, such that $|Q \cap \mathcal{H}| = \dim \mathcal{C}_1$ and for $\ell = 2, 3, \ldots, \ell_{\max}$ $|Q \cap \mathcal{H}^{\otimes \ell}| = \dim \mathcal{C}_\ell - \dim \mathcal{C}_{\ell-1}$ and let $\mathcal{C} = \text{span}\{|t\rangle : |t\rangle \in Q\}$. (Note that by our assumption $\mathcal{H}^{\otimes \ell}$ and $\mathcal{H}^{\otimes \ell'}$ are orthogonal for $\ell \neq \ell'$.) Recall that our goal is to compress quantum data. We observe that to achieve good rates one has to choose codes for which equality holds in (12) for $\ell = 1, 2, \ldots, \ell_{\max} - 1$, that is,

$$\dim \mathcal{C}_\ell = d^\ell. \tag{13}$$

We conclude that there exists a quantum variable–length code defined in [2] iff (10) and (12) hold or equivalently for $\ell \leq \ell_{\max}$,

$$\sum_{i=1}^{\ell} N_{\ell_i} \leq d^\ell$$

where N_ℓ is the number of linearly independent codewords of base length ℓ.

To embed such a code $\mathcal{C}$ into a Hilbert space $\mathcal{H}^{\otimes \ell_{\max}}$ we choose an orthonormal basis of $\mathcal{H}$ in (1) and denote $\mathcal{D} = \{0, 1, \ldots, d-1\}$. Then we rewrite (1) as

$$\mathcal{B}(\mathcal{H}) = \{|\delta\rangle : \delta \in \mathcal{D}\}. \tag{14}$$

Let $\{|w_i\rangle : i = 1, 2, \ldots, d'\}$ be a basis of $\mathcal{C}$ such that its first d^ℓ members $\{|w_j\rangle : j = 1, 2, \ldots, d^\ell\}$ form a basis of $\mathcal{C}_\ell$. Then the linear isometric operator sending the basis of $\mathcal{C}_\ell$ for $\ell = 1, 2, \ldots, \ell_{\max}$ to $\{|\delta^\ell 0^{\ell_{\max}-\ell}\rangle : \delta^\ell \in \mathcal{D}^\ell\}$ embeds a code $\mathcal{C}$ into $\mathcal{H}^{\otimes \ell_{\max}}$. Let $\mathcal{C}'$ and $\mathcal{C}'_\ell$ be the images of $\mathcal{C}$ and $\mathcal{C}_\ell$ respectively. For simplicity, and without loss of generality, we assume $\mathcal{C}' = \mathcal{H}^{\otimes \ell_{\max}}$ and $\mathcal{C}'_\ell$ is the subspace spanned by

$$\triangle_\ell = \{|\delta^\ell 0^{\ell_{\max}-\ell}\rangle : \delta^\ell \in \mathcal{D}^\ell\}. \tag{15}$$

Thus a codeword $|c^{\ell_{\max}}\rangle \in \mathcal{H}^{\otimes \ell_{\max}}$ for a $c^{\ell_{\max}} = (c_1, c_2, \ldots, c_{\ell_{\max}})$ is the image of a codeword of base length ℓ_b iff $c_{\ell_b} \neq 0$ and $c_\ell = 0$ for $\ell = \ell_b + 1, \ldots, \ell_{\max}$. In this case, the encoder may remove the last $\ell_{\max} - \ell_b$ components (which are all zeros) from $c^{\ell_{\max}}$ and send the state $|c^{\ell_b}\rangle$ to the decoder after the embedding. To recover the output state the decoder appends $|0^{\ell_{\max}-\ell_b}\rangle$ to the state $|c^{\ell_b}\rangle$ and then obtains the output state from $|c^{\ell_b} 0^{\ell_{\max}-\ell_b}\rangle$ via the inverse operator of the encoding operator.

Summarizing the discussion, we obtain that lossless quantum data compression can always be done in the following way.

Coding Scheme 1:

(I) Choose an $\ell_{\max}$ such that $d^{\ell_{\max}-1} < d' \leq d^{\ell_{\max}}$, where d and d' are the dimensions of $\mathcal{H}$ and the source space $\mathcal{S}$ respectively, and the encoder $\mathcal{E}$ encodes the states in source space $\mathcal{S}$ to codewords in $\mathcal{H}^{\otimes \ell_{\max}}$ by a *properly* chosen unitary operator from $\mathcal{S}$ to $\mathcal{H}^{\otimes \ell_{\max}}$ (or its subspace in the case $d' \neq d^{\ell_{\max}}$).

(II) In the case, where the output state $|s\rangle$ is encoded to a codeword $|c^{\ell_b}0^{\ell_{\max}-\ell_b}\rangle = \mathcal{E}(|s\rangle)$ with $c^{\ell_b} = c_1, c_2, \ldots, c_{\ell_b}, c_{\ell_b} \neq 0$, the encoder sends $|c^{\ell_b}\rangle$ to the decoder via a (noiseless) quantum channel and sends ℓ_b to the decoder via the classical helping channel. [1]

(III) The decoder appends $|0^{\ell_{\max}-\ell_b}\rangle$ to $|c^{\ell_b}\rangle$ and then decodes $|c^{\ell_b}0^{\ell_{\max}-\ell_b}\rangle$ using the inverse of $\mathcal{E}$, the decoder $\mathcal{Y} = \mathcal{E}^{-1}$, to recover $|s\rangle = \mathcal{Y}(|c^{\ell_b}0^{\ell_{\max}-\ell_b}\rangle = \mathcal{E}^{-1}\big(\mathcal{E}(|s\rangle)\big)$.

We note that the Coding Scheme 1 is not different from "the natural–prefix codes" in [2] unless the padded $|0\rangle$ is moved to the end of codewords and so all optimal codes can be transmitted to codes in [2] by a unitary operator. Our discussion shows that the optimal codes can always be obtained by Coding Scheme 1. We call the codes obtained by the Coding Scheme 1 canonical codes.

2.2 Minimum Compression Rate

In the previous subsection we saw that the minimum achievable compression rate can be achieved by the canonical codes for all quantum sources. It is easy to see that the compression rate of a canonical code depends only on the encoding operator $\mathcal{E}$. In this subsection we focus on the encoding operator and reduce the problem further. To this end, consider a quantum source specified by a probability space $(\mathcal{S}, \mathcal{F}, P)$, where $\mathcal{S}$ is the source space, $\mathcal{F}$ is a σ–field, and the source outputs a state in an $F \in \mathcal{F}$ with probability $P(F)$. The source is not necessarily discrete and so the probability measure P is not necessarily discrete. We call a sequence of subspaces $L = \{L_\ell : \ell = 1, 2, \ldots, \ell_{\max} - 1\}$ of $\mathcal{S}$ for an $\ell_{\max}$ such that $d^{\ell_{\max}-1} < d' \leq d^{\ell_{\max}}$, where $d' = \dim \mathcal{S}$, d–nested if for all ℓ

$$\dim L_\ell = d^\ell \tag{16}$$

and

$$L_1 \subset L_2 \subset \cdots \subset L_{\ell_{\max}-1}. \tag{17}$$

Denote by $\mathcal{L}_d(\mathcal{S})$ the set of d–nested sequences of subspaces of $\mathcal{S}$. Then we have

Theorem 1. *The minimum achievable lossless compression rate of a quantum source specified by a probability space $(S, \mathcal{F}, P)$, via a quantum variable–length code with a classical helper channel informing about base lengths is*

$$R_0 \triangleq \ell_{\max} - \sup_{L \in \mathcal{L}_d(\mathcal{S})} \sum_{\ell=1}^{\ell_{\max}-1} P(L_\ell). \tag{18}$$

Proof: By the discussion in the previous subsection we know that the minimum achievable rate is achievable by canonical codes. So it is sufficient for us to show that R_0 is the minimum achievable rate by canonical codes. We denote the subspace of $\mathcal{H}^{\otimes \ell_{\max}}$ spanned by Δ_ℓ in (15)

[1] To simplify notation, we assume that the encoder has to send something via the quantum channel. So, if $\mathcal{E}(|s\rangle) = |0^{\ell_{\max}}\rangle$, the encoder sends $|0\rangle$ and 1 through quantum and classical channel.

by $\mathcal{C}'_\ell$. Consider a canonical code with encoding operator $\mathcal{E}$. Let $L_\ell(\mathcal{E})$ be the inverse image of the subspace $\mathcal{C}'_\ell$ under the encoding operator $\mathcal{E}$, that is,

$$L_\ell(\mathcal{E}) = \big\{|s\rangle \in \mathcal{S} : \mathcal{E}(|s\rangle) \in \mathcal{C}'_\ell\big\}. \tag{19}$$

Then by linearity of the encoding operator $\mathcal{E}$ $L_\ell(\mathcal{E})$ is a subspace of dimension d^ℓ of $\mathcal{S}$ and

$$L_1(\mathcal{E}) \subset L_2(\mathcal{E}) \subset \cdots \subset L_{\ell_{\max}-1}(\mathcal{E}). \tag{20}$$

That is, $L(\mathcal{E}) = \big\{L_\ell(\mathcal{E}) : \ell = 1, 2, \ldots, \ell_{\max} - 1\big\} \in \mathcal{L}_d(\mathcal{S})$. On the other hand, for any d–nested sequence of subspaces $L = \{L_\ell : \ell = 1, 2, \ldots, \ell_{\max} - 1\} \in \mathcal{L}_d(\mathcal{S})$, there exists a unitrary operator $\mathcal{E}$ sending L_ℓ to $\mathcal{C}'_\ell$ i.e. $L_\ell = L_\ell(\mathcal{E})$ in the sense of (19). Finally, because by definition a codeword has base length ℓ_b iff it is contained in the set $\mathcal{C}'_{\ell_b} \setminus \mathcal{C}'_{\ell_b-1}$ (note that $\mathcal{C}'_{\ell_b} \setminus \mathcal{C}'_{\ell_b-1}$ is not a subspace), we have that the average base length, or the compression rate of a canonical code with encoding operator $\mathcal{E}$ is

$$\begin{aligned}
R(\mathcal{E}) &= 1 \cdot P\big(L_1(\mathcal{E})\big) + \sum_{\ell=2}^{\ell_{\max}-1} \ell P\big(L_\ell(\mathcal{E}) \setminus L_{\ell-1}(\mathcal{E})\big) + \ell_{\max} P\big\{\mathcal{S} \setminus L_{\ell_{\max}-1}(\mathcal{E})\big\} \\
&\overset{(1)}{=} \sum_{\ell=1}^{\ell_{\max}-1} \ell P\big(L_\ell(\mathcal{E})\big) - \sum_{\ell=2}^{\ell_{\max}-1} \ell P\big(L_{\ell-1}(\mathcal{E})\big) + \ell_{\max}\big(1 - P(L_{\ell_{\max}-1}(\mathcal{E})\big) \\
&= \sum_{\ell=1}^{\ell_{\max}-1} \ell P\big(L_\ell(\mathcal{E})\big) - \sum_{\ell=1}^{\ell_{\max}-1} (\ell+1) P\big(L_\ell(\mathcal{E})\big) + \ell_{\max} \\
&= \ell_{\max} - \sum_{\ell=1}^{\ell_{\max}-1} P\big(L_\ell(\mathcal{E})\big),
\end{aligned} \tag{21}$$

where (1) holds because $L_{\ell-1}(\mathcal{E}) \subset L_\ell(\mathcal{E})$, $L_{\ell_{\max}-1}(\mathcal{E}) \subset \mathcal{S}$, $P(\mathcal{S}) = 1$ and so $P\big(L_\ell(\mathcal{E}) \setminus L_{\ell-1}(\mathcal{E})\big) = P\big(L_\ell(\mathcal{E})\big) - P\big(L_{\ell-1}(\mathcal{E})\big)$, $P\big(\mathcal{S} \setminus L_{\ell_{\max}-1}(\mathcal{E})\big) = 1 - P\big(L_{\ell_{\max}-1}(\mathcal{E})\big)$.

This completes our proof. □

The theorem reduces finding an optimal code to finding a d–nested sequence of subspaces maximizing $\sum_{\ell=1}^{\ell_{\max}-1} P\big(L_\ell(\mathcal{E})\big)$.

Two naive greedy algorithms to find the d–nested sequences of subspaces may be considered:

— First find a d dimensional space with maximal probability, say $L'_1(\mathcal{E})$, then take a d^2 dimensional subspace $L'_2(\mathcal{E})$ with maximal probability among all subspaces containing $L'_1(\mathcal{E}), \ldots$, and finally take a $d^{\ell_{\max}-1}$ dimensional subspace with maximal probability among the subspaces containing $L'_{\ell_{\max}-2}(\mathcal{E})$.

— One can also choose the subspaces going from bigger to smaller dimensions. First one chooses a $d^{\ell_{\max}-1}$ dimensional subspace with maximal probability as $L''_{\ell_{\max}-1}(\mathcal{S})$, then one chooses a $d^{\ell_{\max}-2}$ dimensional subspace of $L''_{\ell_{\max}-1}(\mathcal{S})$ with maximal probability as $L''_{\ell_{\max}-2}$, and so on.

The following example shows that in general neither of these methods guarantees that one obtains an optimal code.

It is not hard to see that the coding algorithm in [2] presented at beginning of this section may not be better than the first algorithm.

Example 1: Let $\mathcal{S}$ be a complex Hilbert space of dimension 8 with an orthonormal basis $\{|\beta_i\rangle : i = 0, 1, \ldots, 7\}$, let $\mathcal{H}$ be a complex Hilbert space of dimension 2 with orthonormal basis $\{|0\rangle, |1\rangle\}$, and let p be a real number in $\left(\frac{1}{2}, 1\right)$. We take m states $|\alpha_j\rangle$, $j = 0, 1, \ldots, m-1$ in the subspace spanned by $\{|\beta_i\rangle : i = 4, 5, 6, 7\}$ such that no four of them are in the same 3 dimensional subspace. Let P be a probability distribution with $P(|\alpha_j\rangle) = \frac{1}{m}p$, for $j = 0, 1, \ldots, m-1$ and $P(|\beta_i\rangle) = \frac{1}{4}(1-p)$ for $i = 0, 1, 2, 3$. Let us consider a code encoding $\mathcal{S}$ to $\mathcal{H}^{\otimes 3}$, in particular the following two codes.

Code A: The encoder sends two vectors in $|\beta_i\rangle$, $i = 0, 1, 2, 3$ e.g., say $|\beta_0\rangle$ and $|\beta_1\rangle$ to $|000\rangle$ and $|100\rangle$ respectively. That is, choose $L_1(\mathcal{E}) = \mathrm{span}(|\beta_0\rangle, |\beta_1\rangle)$, where $\mathrm{span}(\cdot)$ is the subspace spanned by the states (or the subset of states) in the bracket. Then choose $L_2(\mathcal{E}) = \mathrm{span}(|\beta_i\rangle : i = 0, 1, 2, 3)$. Thus $\ell_{\max} = 3$ and

$$P\big(L_1(\mathcal{E})\big) + P\big(L_2(\mathcal{E})\big) = \frac{1}{2}(1-p) + (1-p) = \frac{3}{2}(1-p). \tag{22}$$

Code B: Take any two vectors from the m vectors $|\alpha_j\rangle$, $j = 0, 1, \ldots, m-1$, say $|\alpha_0\rangle$ and $|\alpha_1\rangle$ and choose $L_1(\mathcal{E}) = \mathrm{span}(|\alpha_0\rangle, |\alpha_1\rangle)$ and $L_2(\mathcal{E}) = \mathrm{span}\{|\beta_i\rangle : i = 4, 5, 6, 7\}$. Thus $L_{\max} = 3$ and

$$P\big(L_1(\mathcal{E})\big) + P\big(L_2(\mathcal{E})\big) = \frac{2}{m}p + p = \frac{m+2}{m}p.$$

By comparing the right hand side of (22) and this formula, we conclude that code A (code B) is better iff $p < \frac{3m}{5m+4}$ $\left(p > \frac{3m}{5m+4}\right)$ as $\frac{3}{2}(1-p) > \frac{m+2}{m}p$ iff $p < \frac{3m}{5m+4}$.

At first we choose $p = \frac{3}{5}$, then for all m, the code B is better than code A, as $\frac{3m}{5m+4} < \frac{3}{5} = p$. However in the case $m > 6$, the subspace $\mathrm{span}\{|\beta_0\rangle, |\beta_1\rangle\}$ achieves the maximal probability of 2–dimensional subspaces, $\frac{1}{5}$ $\left(> \frac{6}{5m}\right)$, and so one obtains the code A if one performs the first greedy algorithm.

Secondly we choose $p = \frac{6}{11}$. Then for sufficiently large m, we have that $p = \frac{6}{11} < \frac{3m}{5m+4}$ since $\frac{3m}{5m+4} \to \frac{3}{5}$ as $m \to \infty$.

In this case the code A is better than code B. However for all m, one will obtain the code B when one performs the second greedy algorithm, because by the choice $p = \frac{6}{11}$, $\mathrm{span}\{|\beta_i\rangle : i = 4, 5, 6, 7\}$ achieves the maximal probability $\frac{6}{11}$ $\left(> \frac{5}{11}\right)$ of 4 dimensional subspaces.

Thus we conclude that no greedy algorithm is always able to find the optimal codes.

Recalling $\dim \mathcal{H} = d$, $\dim \mathcal{S} = d'$ and $d^{\ell_{\max}-1} < d' \leq d^{\ell_{\max}}$, we have that in order to embed $\mathcal{S}$ to the ℓth tensor power of $\mathcal{H}$, ℓ must be at least $\ell_{\max}$. In other words, the length of the code is $\ell_{\max}$ if one wants to encode the source space $\mathcal{S}$ to a block code with minimum length. So by Theorem 1 $\sum_{\ell=1}^{\ell_{\max}-1} P\big(L_\ell(\mathcal{E})\big)$ is the reduced part of length gained by using a variable length code. We say the source is compressible if it is positive and otherwise the source is

incompressible. Notice $0 \leq P\big(L_\ell(\mathcal{E})\big) \leq P\big(L_{\ell+1}(\mathcal{E})\big) \leq 1$. In the case that $P\big(L_{\ell^*}(\mathcal{E})\big) = 1$ for an $\ell^* \leq \ell_{\max} - 1$, we have that for R_0 in (18) $R_0 \leq \ell^* - \sum_{\ell=1}^{\ell^*-1} P\big(L_\ell(\mathcal{E})\big)$.

Intuitively this means one can encode $\mathcal{S}$ without error to the d^{ℓ^*}–dimensional subspace L_{ℓ^*}, which is isomorphic to $\mathcal{H}^{\oplus \ell^*}$ up to the difference of a null probability set. The other extremal case is $\sup_{L \in \mathcal{L}_d(\mathcal{S})} \sum_{\ell=1}^{\ell_{\max}-1} P(L_\ell) = 0$, which is equivalent to having the probabilities of all subspaces of $\mathcal{S}$, whose dimensions are not larger than $d^{\ell_{\max}-1}$, be zero. In this case the source is incompressible. So Theorem 1 has the following consequence.

Corollary 1. *A source $\mathcal{S}$ is incompressible iff the probabilities of all subspaces whose dimension are not larger than $d^{\ell_{\max}-1}$ are zero.*

Notice that here the compression rates and compressibility are based on the codes in [2]. In the following sections we discuss **more general and more efficient codes**. We shall see that for these codes Corollary 1 still holds. We conclude this section with an observation.

2.3 Base Length of Codewords are not Measurable

We know that there does not exist a blind quantum zero–error source code because by Observation I one can not perform a length measurement to learn the lengths of codewords. Then a question naturally arises: Would a blind encoder be possible under the assumption that the encoder has many states identifying the output of the source? We assume that these states are output by a "multiple identity" quantum source because one cannot copy an unknown state. Then the encoder may encode the identical states, send one of the identical codewords to the decoder and measure the rest to learn the base length of the codeword (when the decoder needs only one output state). Thus a blind code with classical helping channel would exist if the encoder might learn the base length from the outcome of measurement. But the following observation shows that in general one can never obtain the base length of an unknown codeword by a measurement.

Observation II: There is no measurement to measure the base length of unknown codewords without error.

The observation actually is an immediate consequence of the well–known fact that non–orthogonal states cannot be reliably distinguished (e.g. c.f. p87 of [11]). To see this let us consider two codewords $|w_1\rangle = |100\rangle$ and $|w_2\rangle = a|100\rangle + b|010\rangle$, $(a, b > 0\ a^2 + b^2 = 1)$ of a quantum binary canonical code of maximum length 3. Since $|w_1\rangle$ and $|w_2\rangle$ have different base lengths i.e. $L(|w_1\rangle) = 1 \neq 2 = L(|w_2\rangle)$, one could distinguish them correctly if there were a measurement to measure their base lengths without error. But in fact, it is not possible to distinguish them since they are not orthogonal.

Thus from the Observation II, we know that only a visible encoder can be considered even in the case, where a quantum source is able to output many indentical states simultaneously.

3 Lossless Quantum Data Compression with Classical Helper

We know that the following two assumptions are necessary for lossless quantum data compression.

(1) Visible quantum encoding: The encoder knows the output state of the quantum source.

(2) The classical helper: There is a classical channel connecting the encoder and the decoder such that the encoder can send classical information to the decoder.

Here we assume that the encoder and decoder each consist of two components, namely a quantum and a classical component and the quantum operators that are used in the quantum component of the encoder and the decoder are linear isometric. A classical variable–length code can be used in classical components. As the encoding is visible, the encoder may choose his quantum operators in quantum components and the codeword of the classical variable–length code according to the output of the quantum source. The decoder may choose his quantum operator according to the received classical codewords. Our constraint on the quantum operators may seem too strong compared to using to quantum operations, which are linear, completely positive and trace–preserving on density operators. But it turns out that one cannot do better than our coding scheme under our constraint even when one relaxes the constraint.

In [2] K. Boström and T. Felbinger use the classical channel to send the base lengths of the codewords. Our main motivation in the paper is to find a more efficient way to use the classical helper.

Example 2: Let $\dim \mathcal{H} = 2$, $\dim \mathcal{S} = 4$ and $\mathcal{S}_0$ and $\mathcal{S}_1$ be two orthogonal subspaces of $\mathcal{S}$ of dimension 2. P is a probability distribution over $\mathcal{S}$ such that $P(\mathcal{S}_1) = P(\mathcal{S}_2) = \frac{1}{2}$. Suppose the source outputs a state in $\mathcal{A} \subset \mathcal{S}$ with the probability $P(\mathcal{A})$. For an example of a "continuous" source one may assume P is uniformly distributed on $\mathcal{S}_0 \cup \mathcal{S}_1$ and for an example of a discrete quantum source one may assume P is uniformly distributed on a set of states

$$\{|u_i\rangle : i = 0, 1, 2, \ldots, m-1\} \cup \{|v_j\rangle : j = 0, 1, 2, \ldots, m-1\},$$

where $|u_i\rangle \in \mathcal{S}_0$, $|v_j\rangle \in \mathcal{S}_1$, and $m \geq 3$. But we shall see that the assumption for assigning the probabilities to the particular states makes no difference. Now $\ell_{\max} = 2$ and it is easy to see the maximum probability of 2–dimensional subspaces of $\mathcal{S}$ is $\frac{1}{2}$. So by Theorem 1, the best quantum compression rate with classical helping channel informing the base length is $\frac{3}{2}$. Additionally the encoder has to send one bit to the decoder to inform him about the base length.

In the source under consideration the probability is concentrated on $\mathcal{S}_0 \cup \mathcal{S}_1$. Therefore the encoder can compress the quantum source more efficiently. One can simply choose two arbitrary unitary operators U_0 and U_1, one mapping from $\mathcal{S}_0$ to $\mathcal{H}$ and the other from $\mathcal{S}_1$ to $\mathcal{H}$. In the case that a state $|s\rangle \in \mathcal{S}_i$ for $i = 0$ or 1, is output from the source, the encoder encodes it to $U_i|s\rangle$ by using operator U_i and sends i to the decoder via the classical channel.

Then the decoder who knows i decodes the quantum codeword by using U_i^{-1} and obtains $U_i^{-1}U_i|s\rangle = |s\rangle$. For this code the quantum compression rate is 1 and the encoder sends one bit via the classical channel. It is therefore a better code.

Using this idea, many such examples, including ones with more complicated sources can be found. The simple example above is sufficient to lead us to the following coding scheme. To simplify the notation we assume that the source is discrete. Let $\mathcal{H}$ and $\mathcal{S}$ be complex Hilbert spaces of dimensions d and d' respectively. Assume the quantum source outputs a state $|u\rangle \in \mathcal{S}$ with probability $P(u)$, where P is a probability distribution with a finite support $\mathcal{U}$. Without loss of generality we assume that $\mathcal{S} = \text{span}\{|u\rangle : u \in \mathcal{U}\}$, because otherwise we may replace $\mathcal{S}$ by $\text{span}\{|u\rangle : u \in \mathcal{U}\}$.

Coding Scheme 2:

(I) Partition $\mathcal{U}$ properly into $\{\mathcal{U}_j : j = 0, 1, \ldots, J-1\}$ for an integer J. For each j find the minimum ℓ_j such that there is an d^{ℓ_j}–dimensional subspace $\mathcal{S}_j$ of $\mathcal{S}$, containing $\text{span}\{|u\rangle : u \in \mathcal{U}_j\}$. We write $L_q(\mathcal{U}_j) = \ell_j$.

(II) For all $j \in \{0, 1, \ldots, J-1\}$, arbitrarily choose a unitary operator U_j from $\mathcal{S}_j$ to $\mathcal{H}^{\otimes \ell_j}$.

(III) Suppose a $|u\rangle \in \mathcal{S}$ is output by the quantum source and assume that $|u\rangle \in \mathcal{S}_j$. Then the encoder encodes $|u\rangle$ to a codeword $|w(u)\rangle \triangleq U_j|u\rangle \in \mathcal{H}^{\otimes L_q(\mathcal{U}_j)}$ by using the operator U_j. We say $|u\rangle$ is encoded to a quantum codeword $|w(u)\rangle$ of length $L_q(|w(u)\rangle) = L_q(\mathcal{U}_j)$. Then the encoder sends j by a classical variable–length code e.g., a Huffman code, for a classical source outputing $j \in \{0, 1, 2, \ldots, J-1\}$ with probability $Q(j) = P(\mathcal{U}_j)$, to the decoder via the classical channel.

(IV) Finally the decoder who has the quantum codeword $|w(u)\rangle = U_j|u\rangle$ and knows j from the classical channel, reconstructs the output state $|u\rangle$ by applying the operator U_j^{-1} to $|w(u)\rangle$.

The reader can generalize the scheme to the general source. The reader also may easily verify that the condition in Corollary 1 is still sufficient and necessary for compressibility by using the more general scheme below.

Also it is not hard to see that Coding Scheme 2 is the most general under the two assumptions (visible encoding and classical helper) at the beginning of this section i.e., there is no better code than the best codes constructed by the Coding Scheme 2.

To see this let us assume the encoder encodes the output states into classical messages in a finite set, say $\{0, 1, \ldots, J-1\}$, by a mapping φ and sends the value of φ via a classical channel. Let the inverse image of j under the mapping φ be $\varphi(j)^{-1} = \{|u\rangle : \varphi(|u\rangle) = j\}$ and set $\mathcal{U}_j = \{u : |u\rangle \in \varphi(j)^{-1}\}$. The encoder has to send the output states of the quantum source to tensor powers of Hilbert space $\mathcal{H}$, we allow the use of any linear mapping to do it, but two output states in the same inverse image $\varphi(j)^{-1}$ must be sent to different quantum codewords, because the compression is lossless. In other words the restriction of the "quantum" encoder must be injectivity. On the other hand the decoder has to treat the quantum codewords of all output states which are in $\varphi(j)^{-1}$ in the same tensor power space of $\mathcal{H}$, say $\mathcal{H}^{\otimes \ell}$ since his only classical knowledge is j.

Obviously $d^\ell \geq \dim\bigl[\mathrm{span}\{|u\rangle : u \in \mathcal{U}_j\}\bigr]$ must hold because the restriction on φ^{-1} is injective. Thus we obtain a code constructed by Coding Scheme 2 by letting $\mathcal{S}_j$ be the inverse image of the quantum encoder.

We call a code constructed by Coding Scheme 2 a quantum–classical or $q-c$ variable–length code, and its two components, quantum and classical component, respectively, and speak of lossless quantum data compression with a classical helper.

Next we point out that the compression in steps (II), (III) and (IV) of Coding Scheme 2 can not be improved except by choosing a better classical variable–length code in step (III). So the key step is the first step. That is, the quality of a compression based on the two assumptions at the beginning of this section depends only on how to partition $\mathcal{U}$ according to d, d' and the probability distribution P.

Coding Scheme 1 is a special case of Coding Scheme 2 obtained through assigning $\varphi^{-1}(j) = \{|u\rangle : u \in \mathcal{U}_\ell\} = L_\ell(\mathcal{E})$ as defined in the proof of Theorem 1. We note that there is no rate limit to the classical channel in [2]. We have to count the rate of the classical channel since otherwise the encoder may send the index u for all output states via the classical channel and the quantum part of compression is not needed at all.

We denote by $L_c(\mathcal{U}_j)$ the length of the codeword to which the classical message is encoded by the classical variable–length code in step (III) of the Coding Scheme 2 when $|u\rangle \in \mathcal{U}_j$. Then the classical and quantum components of the compression rate are

$$R_c = \sum_{j=0}^{J-1} P(\mathcal{U}_j) L_c(\mathcal{U}_j) \tag{23}$$

and

$$R_q = \sum_{j=0}^{J-1} P(\mathcal{U}_j) L_q(\mathcal{U}_j) \tag{24}$$

respectively. By the Shannon Source Coding Theorem [15], [4], [5], with the notation

$$Q \triangleq \bigl\{Q(j) = P(\mathcal{U}_j) : j = 0, 1, \ldots, J-1\bigr\},$$

(23) is bounded by

$$(\log a)^{-1} H(Q) \leq R_c < (\log a)^{-1} H(Q) + 1 \tag{25}$$

if an optimal classical variable–length code is used, where a is the size of the alphabet of the classical channel. To reduce the classical component of the rate, one has to reduce the Shannon entropy $H(Q)$, which amounts to reducing the number of subsets J. But the reduction of the number of subsets will increase the dimensions of subspaces $\mathcal{S}_j$, which increases the quantum component of the rate. Similarly reducing the quantum component of the rate will increase its classical component of the rate.

We note that for $j \neq j'$ the intersection of $\mathcal{S}_j \cap \mathcal{S}_{j'}$ is not necessarily $\{0\}$ and an output state $|u\rangle$ may be in different $\mathcal{S}_j$'s even in a good code for a source which is not so good (see section 4). When a state $|u\rangle \in \mathcal{S}_j \cap \mathcal{S}_{j'}$, $j \neq j'$, one can put $|u\rangle$ in $\mathcal{U}_j$ or $\mathcal{U}_{j'}$ without changing the list $\{L_q(\mathcal{U}_{j''}) : j'' = 0, 1, \ldots, J-1\}$ of the length of quantum codewords. To reduce

the classical components of the rate, one should put $|u\rangle$ into a $\mathcal{U}_j$ such that $|u\rangle \in \mathcal{S}_j$ and $P(\mathcal{U}_j)$ achieves $\max_{j':|u\rangle\in\mathcal{S}_{j'}} P(\mathcal{U}_{j'})$ to reduce the Shannon entropy $H(Q)$. On the other hand, to reduce the quantum component of the rate one should put $|u\rangle$ into a $\mathcal{U}_j$ such that $\mathcal{S}_j$ has the minimum dimension among the $\mathcal{S}_{j'}$'s containing $|u\rangle$. The two actions often tend to opposite directions as the lower dimensional space typically contains smaller subsets of $\mathcal{U}$ and smaller subsets often have smaller probabilities. In the model of [2] (c.f. Subsection 1.5.3 and Section 2) the classical component of the rate is not counted and we therefore always put $|u\rangle$ into a subspace of lowest possible dimension. This may increase the cost of using the classical channel.

4 Von Neumann Entropy Bound

In this section we consider a discrete quantum source i.e. the distribution of the source has a finite support set $\mathcal{U}$, and derive a lower bound on the compression rates of $q-c$ variable–length codes in term of von Neumann entropy. To simplify the notation, we assume that the codewords of classical components take values in a finite set $\mathcal{X}$ of cardinality $|\mathcal{X}| = \dim \mathcal{H} = d$. For a given $q-c$ variable–length code, we define $\{\mathcal{U}_j\}_{j=0}^{J-1}$ and $\{\mathcal{S}_j\}_{j=0}^{J-1}$ as in the previous section. Then we have

Theorem 2: *For any $q-c$ variable–length code,*

$$R_q + R_c \geq (\log d)^{-1} S(\rho), \tag{26}$$

where $S(\rho)$ is the von Neumann entropy of the state,

$$\rho \triangleq \sum_{u\in\mathcal{U}} P(u)|u\rangle\langle u|, \tag{27}$$

and equality holds iff the following conditions hold simultaneously.

(i) For the probability Q in (25) i.e. $Q(j) = P(\mathcal{U}_j)$,

$$R_c = (\log d)^{-1} H(Q). \tag{28}$$

(ii) For all $j \neq j'$

$$\mathcal{S}_j \perp \mathcal{S}_{j'}, \tag{29}$$

and

(iii) for all $j \in \{0, 1, \ldots, J-1\}$

$$P(\mathcal{U}_j)^{-1} \sum_{u\in\mathcal{U}_j} P(u)|u\rangle\langle u| = d_j'^{-1}\mathcal{P}_j, \tag{30}$$

where $d_j' = \dim \mathcal{S}_j$ and $\mathcal{P}_j$ is the projector onto the subspace $\mathcal{S}_j$.

Proof: Denote by $\mathcal{U}_j^* = \text{span}\{|u\rangle : u \in \mathcal{U}_j\}$, $d_j^* = \dim \mathcal{U}_j^*$, and $\rho_j = P(\mathcal{U}_j)^{-1} \sum_{u \in \mathcal{U}_j} P(u) |u\rangle\langle u| = \sum_{u \in \mathcal{U}_j} P(u|\mathcal{U}_j)|u\rangle\langle u|$. Then we have

$$d_j^* \leq d_j' = \dim \mathcal{S}_j \tag{31}$$

with equality iff

$$\mathcal{U}_j^* = \mathcal{S}_j. \tag{32}$$

Moreover by the property of von Neumann entropy that is maximized uniquely by the "uniform" state $I/\tilde{d}$, where I and $\tilde{d}$ are identity operator and the dimension of the Hilbert space containing the state, we obtain

$$S(\rho_j) \leq \log d_j^* \tag{33}$$

and equality holds iff

$$\rho_j = d_j^{*-1} \mathcal{P}_j^*, \tag{34}$$

where $\mathcal{P}_j^*$ is the projector onto the subspace $\mathcal{U}_j^*$. Then it follows from (31) – (34) that

$$S(\rho_j) \leq \log d_j' = \ell_j \log d \tag{35}$$

and equality holds iff (32) and (34) hold or in other words (30) holds. The equality in (35) holds because by the definitions of ℓ_j and $\mathcal{S}_j$ $\quad \dim \mathcal{S}_j = d^{\ell_j}$.

Next we notice that by the definition of ρ_j,

$$\rho = \sum_{j=0}^{J-1} Q(j) \rho_j, \tag{36}$$

where

$$Q(j) = P(\mathcal{U}_j). \tag{37}$$

We use a well–known inequality for von Neumann entropy e.g., see P. 518 [11] to obtain

$$S(\rho) \leq H(Q) + \sum_{j=0}^{J-1} Q(j) S(\rho_j), \tag{38}$$

with equality iff

$$\mathcal{U}_j^* \perp \mathcal{U}_{j'}^* \text{ for all } j \neq j'. \tag{39}$$

Finally we combine (23), (25), (35), (38) with the notation $L_q(|w(u)\rangle) = \ell_j$ for $|u\rangle \in \mathcal{U}_j$ and obtain

$$\begin{aligned} R_q + R_c &\overset{(1)}{\geq} (\log d)^{-1} H(Q) + \sum_{j=0}^{J-1} Q(j) L_q(\mathcal{U}_j) \overset{(2)}{\geq} (\log d)^{-1} H(Q) + \sum_{j=0}^{J-1} Q(j) (\log d)^{-1} S(\rho_j) \\ &= (\log d)^{-1} \left[H(Q) + \sum_{j=1}^{J-1} Q(j) S(\rho_j) \right] \overset{(3)}{\geq} (\log d)^{-1} S(\rho), \end{aligned} \tag{40}$$

where (1) holds by (23) and (25); (2) holds by (35) with $L_q(|w(u)\rangle) = \ell_j$ for $|u\rangle \in \mathcal{U}$; and (3) holds by (38). That is (26).

The equality holds in (26) iff the inequalities (1) – (3) in (40) hold with equalities, which is true iff (28), (30) and (39) hold or equivalently (28), (29) and (30) simultaneously hold. Thus the proof is complete. □

By the above theorem, we see that the von Neumann entropy as a lower bound of compression rate seldom is tight. It is also not hard to construct a quantum source such that the gap between the minimum achievable rate and von Neumann entropy is very large. This is completely different from the Shannon entropy as a lower bound to the classical compression rate. We therefore doubt whether von Neumann entropy fits lossless quantum data compression. In the next section we present more reasons for this, but first we look at gaps between the compression rates and von Neumann entropy.

We fix an arbitrary $q-c$ variable–length code and let

$$\Delta \triangleq (R_q + R_c) - (\log d)^{-1} S(\rho), \tag{41}$$

$$\Delta_c \triangleq R_c - (\log d)^{-1} H(Q), \text{ (for } Q \text{ in (37))}, \tag{42}$$

$$\Delta_{q,1}(j) \triangleq (\log d)^{-1} \log d'_j - (\log d)^{-1} \log d^*_j \text{ (for } d'_j, d^*_j \text{ in (31))} \tag{43}$$

$$\Delta_{q,1} \triangleq \sum_{j=0}^{J-1} Q(j) \Delta_{q,1}(j), \tag{44}$$

$$\Delta_{q,2} \triangleq (\log d)^{-1} \left(H(Q) + \sum_{j=0}^{J-1} Q(j) S(\rho_j) \right) - (\log d)^{-1} S(\rho), \text{ (for } \rho_j, \rho \text{ in (36))}, \tag{45}$$

$$\Delta_{q,3}(j) \triangleq (\log d)^{-1} \log d^*_j - (\log d)^{-1} S(\rho_j), \tag{46}$$

and

$$\Delta_{q,3} \triangleq \sum_{j=1}^{J-1} Q(j) \Delta_{q,3}(j). \tag{47}$$

Then by (24), (41) – (47) and

$$d'_j = d^{L_q(\mathcal{U}_j)} \tag{48}$$

we obtain that

$$\begin{aligned}\Delta &= \big[R_c - (\log d)^{-1}H(Q)\big] + \left[\sum_{j=0}^{J-1} Q(j)L_q(\mathcal{U}_j) - \sum_{j=0}^{J-1} Q(j)(\log d)^{-1}\log d_j^*\right] \\ &\quad + \left[(\log d)^{-1}\left(H(Q) + \sum_{j=0}^{J-1} Q(j)S(\rho_j)\right) - (\log d)^{-1}S(\rho)\right] \\ &\quad + \left[\sum_{j=0}^{J-1} Q(j)(\log d)^{-1}\log d_j^* - \sum_{j=0}^{J-1}(\log d)^{-1}Q(j)S(\rho_j)\right] \\ &= \Delta_c + \sum_{j=0}^{J-1} Q(j)\big[(\log d)^{-1}\log d_j' - (\log d)^{-1}d_j^*\big] + \Delta_{q,2} + \Delta_{q,3} \\ &= \Delta_c + \Delta_{q,1} + \Delta_{q,2} + \Delta_{q,3}. \end{aligned} \tag{49}$$

We now analyse the differences from an information theoretical point of view.

$\boldsymbol{\Delta_c}$ **:**The difference Δ_c in (42) is the gap between the rate of classical variable length and Shannon entropy. It cannot be avoided. It is not too serious because in the case that the classical component of the $q-c$ variable–length code is an optimal classical variable–length code, the gap satisfies $\Delta_c < 1$.

$\boldsymbol{\Delta_{q,1}}$ **:**The gap $\Delta_{q,1}$ in (42) and (43) is due to the fact that $\mathcal{U}_j^*$'s may not well match a tensor power of Hilbert space $\mathcal{H}$ and it vanishes when $d_j^* = \dim \mathcal{U}_j^*$ is a power of $d = \dim \mathcal{H}$ for all $j \in \{0, 1, \ldots, J\}$. Similar to classical coding, one looses rate when one encodes a set of messages whose size is not a power of d to a set of codewords of block length from an alphabet of cardinality d. Clearly it cannot be avoided, but this is not serious, because according to step (I) of Coding Scheme 2 one may always choose an $L_q(\mathcal{U}_j)$ such that $d^{L_q(\mathcal{U}_j)-1} < d_j^* \leq d^{L_q(\mathcal{U}_j)} = d_j'$ for all j and so the gap $\Delta_{q,1} < 1$.

$\boldsymbol{\Delta_{q,2}}$ **:**We have seen that $\Delta_{q,2}$ vanishes iff (29) holds for all $j \neq j'$. $\Delta_{q,2}$ may be very large if the $\mathcal{S}_j$'s are far away from being pairwise orthonormal. Denote by $\mathcal{X}(Q;\{\rho_j\})$ the Holevo quantity of quantum channel $\{\rho_j\}_{j=0}^{J-1}$ with classical input distribution Q, i.e. $\mathcal{X}(Q;\{\rho_j\}) = S(\rho) - \sum_{j=0}^{J-1} Q(j)S(\rho_j)$, where $\rho = \sum_{j=0}^{J-1} Q(j)\rho_j$.

Then $\Delta_{q,2} = (\log d)^{-1}\big[H(Q) - \mathcal{X}(Q(j);\{\rho_j\})\big]$. It is well–known that Holevo's bound [8] is a lower bound of the information about the input of the receiver of a quantum channel with classical input obtained by performing a measurement and $\max_P \mathcal{X}(P;\{\rho_j\})$ is the capacity of the classical quantum channel for classical information ([9] and [13]).

So roughly speaking $(\log d)\ \Delta_{q,2}$ plays a role of conditional entropy of a classical input of a quantum channel given the output, and it reflects the uncertainty of subspaces $\mathcal{S}_j$ containing a given state in $\mathcal{S}$. (The uncertainty is zero exactly when (29) holds.) The larger the gap $\Delta_{q,2}$ is, the less information about the classical component of a $q-c$ variable–length code is carried by the quantum component. In other words $\Delta_{q,2}$ is large, if the overlaps of the subspaces $\mathcal{S}_j$ are significant. In particular $\Delta_{q,2}$ never vanishes for the canonical codes in

Subsection 2.1 (and consequently for the code in Section VII of [2]). Even worse, one may construct a quantum source such that the gap $\Delta_{q,2}$ for an optimal $q-c$ variable–length code is larger than any given number. The following is an example:

Example 3: Let $\mathcal{U} = \{u_{i,k} : i = 0, 1, \ldots, I-1; k = 0, 1, \ldots, K-1\}$ be such that $\{|u\rangle : u \in \mathcal{U}\}$ is a basis of Hilbert space $\mathcal{S}$.

Let $\mathcal{S}_i^* = \text{span}\{|u_{i,k}\rangle : k = 0, 1, \ldots, K-1\}$. Suppose that a quantum source outputs a state $|u\rangle, u \in \mathcal{U}$, with probability $P(u) = \frac{1}{IK}$. We assume $\dim \mathcal{H} = d = 2$ and both I and K are powers of 2. We show in the next section that for an optimal code

$$R_q + R_c = H(P) = \log IK. \tag{50}$$

Now let $\{|u_{0,k}\rangle : k = 0, 1, \ldots, K-1\}$ be an orthonormal basis of the subspace $\mathcal{S}_0^*$, and denote $\rho_i = \sum_{k=0}^{K-1} \frac{1}{K} |u_{i,k}\rangle\langle u_{i,k}|$ for $i = 0, 1, \ldots, I-1$ and $\rho = \sum_{i=0}^{I-1} \frac{1}{I} \rho_i$. Then $S(\rho_0) = \log K$.

Next, fix the rate of classical component $R_c = \log I$. In order to obtain an optimal $q-c$ variable–length code, one has to choose $J = I$ and $\mathcal{S}_i = \mathcal{S}_i^*$ in Coding Scheme 2. Therefore by (45),

$$\Delta_{q,2} = \left(\log I + \sum_{i=0}^{I-1} \frac{1}{I} S(\rho_i) \right) - S(\rho). \tag{51}$$

However, if we choose $|u_{i,k}\rangle$ sufficiently close to $|u_{0,k}\rangle$ for $i = 1, 2, \ldots, I-1$ and $k = 0, 1, \ldots, K-1$ (in ℓ_2–distance or in trace distance $\frac{1}{2} tr \big| |u_{0,k}\rangle\langle u_{0,k}| - |u_{i,k}\rangle\langle u_{i,k}| \big|$), then ρ_i for $i = 1, 2, \ldots, I-1$ and ρ are sufficiently close to ρ_0 in trace distance.

Consequently by Fannes inequality [6] (the continuity of von Neumann entropy) $S(\rho)$ and $S(\rho_i)$ for $i = 1, 2, \ldots, I-1$ are sufficiently close to $S(\rho_0) = \log K$.

Thus by (50) (for $Q(i) = P(\mathcal{S}_i^*)$) $H(Q) + \sum_{i=0}^{I-1} Q(i) S(\rho_i)$ is sufficiently close to $R_q + R_c$ and by (51) $\Delta_{q,2}$ is sufficiently close to $\log I$.

$\Delta_{q,3}$: The gap $\Delta_{q,3}$ is even worse: One may make $S(\rho_j)$ in (46) arbitrarily small by choosing $\rho_j = (1-\varepsilon)|u_{j,0}\rangle\langle u_{j,0}| + \sum_{k=1}^{d_j^*-1} \frac{\varepsilon}{d_j^*-1} |u_{j,k}\rangle\langle u_{j,k}|$ for arbitrarily small ε and $|u_{j,k}\rangle$ for $j = 0, 1, 2, \ldots, J-1$ in Example 3 and so $\Delta_{q,3}$ may be arbitrarily close to the rate of quantum component R_q. Clearly the gap is the cost paid for the case that the decoder has no other knowledge except that which is received from the classical helper. As the quantum datum is locally incompressible, he has to treat ρ_j as the worst local state $d_j^{*-1}\mathcal{P}_j^*$ (in (34)). That is, the farther ρ_j is from $d_j^*\mathcal{P}_j$, the larger the gap $\Delta_{q,3}$ is.

Given the gaps $\Delta_{q,2}$ and $\Delta_{q,3}$, a question arises: Does von Neumann entropy fit lossless quantum data compression?

5 Linear Independent Discrete Quantum Sources: Return to Shannon Entropy

In this section we examine a special family of quantum sources. We call a discrete quantum source linearly independent, or LIDQS for short, if the support set of distribution P of the source is a set of linearly independent (not necessarily orthogonal) states $\{|u\rangle : u \in \mathcal{U}\}$ of the source space $\mathcal{S}$. As in the previous sections, we assume that the input alphabet of the classical channel has cardinality $d = \dim \mathcal{H}$ and denote it by $\mathcal{D} = \{0, 1, \ldots, d-1\}$. Let $\{|y\rangle : y \in \mathcal{D}\}$ be an orthonormal basis of $\mathcal{H}$, and let $\{\mathcal{U}_j\}_{j=0}^{J-1}$ and $\{\mathcal{S}_j\}_{j=0}^{J-1}$ be as in Coding Scheme 2 for a $q-c$ variable–length code for the LIDQS. Then

$$|\mathcal{U}_j| \leq \dim \mathcal{S}_j = \dim \mathcal{H}^{\otimes L_q(\mathcal{U}_j)} = d^{L_q(\mathcal{U}_j)} \tag{52}$$

since $\{|u\rangle : u \in \mathcal{U}_j\}$ is a set of independent states. Therefore we can map $|u\rangle : u \in \mathcal{U}_j$ to a sequence in $\mathcal{D}^{L_q(\mathcal{U}_j)}$ by an injector Ψ_j. Let j be encoded to a classical codeword $\varphi(j)$ in the classical variable–length code in the step (III) of Coding Scheme 2. For $u \in \mathcal{U}$ let

$$c(u) = \big(\varphi(j)\Psi_j(u)\big) \tag{53}$$

if $u \in \mathcal{U}_j$, and let $\ell_c(u)$ be the length of $c(u)$. Then $\{c(u) : u \in \mathcal{U}\}$ is a classically uniquely decodable variable–length encoding of the elements of $\mathcal{U}$ to sequences with alphabet $\mathcal{D}$. So by the classical source coding theorem (e.g. see [4] and [5]) we have that

$$\sum_{u\in\mathcal{U}} P(u)\ell_c(u) \geq (\log d)^{-1} H(P), \tag{54}$$

with equality iff for all $u \in \mathcal{U}$ (with $P(u) > 0$)

$$\ell_c(u) = -\log P(u). \tag{55}$$

By (53), we obtain

$$\ell_c(u) = L_c(\mathcal{U}_i) + L_q(\mathcal{U}_i), \tag{56}$$

if $u \in \mathcal{U}_i$ and so by (54)

$$R_q + R_c \geq (\log d)^{-1} H(P) \tag{57}$$

with equality iff for all $u \in \mathcal{U}_j$ and all j

$$-\log P(u) = L_c(\mathcal{U}_j) + L_q(\mathcal{U}_j). \tag{58}$$

Now we assume that (57) holds with equality and consequently so does (58). Recalling that $Q(j) = P(\mathcal{U}_j)$ in (25), under our assumption we obtain

$$R_q \leq (\log d^{-1})\big(H(P) - H(Q)\big) = -(\log d^{-1}) \left[\sum_{j=0}^{J-1} \sum_{u\in\mathcal{U}_j} P(u) \log \frac{P(\mathcal{U}_j)}{P(u)}\right], \tag{59}$$

by combining the first inequality in (25) with (57). (Note that under our assumption (57) is now an equality.) Moreover (58) implies that the value of probability $P(u)$ depends on u according to which $\mathcal{U}_j$ u lies in, or in other words for all $u \in \mathcal{U}_j$

$$P(u) = \frac{P(\mathcal{U}_j)}{|\mathcal{U}_j|}. \tag{60}$$

Thus by (24), (52), (59), and (60) we have that

$$R_q \leq -(\log d)^{-1} \sum_{j=0}^{J-1} P(\mathcal{U}_j) \log |\mathcal{U}_j| \leq \sum_{j=1}^{J-1} P(\mathcal{U}_j) L_q(\mathcal{U}_j) = R_q. \tag{61}$$

That means all inequalities in (59) and (61) must be equalities, which is equivalent to $L_c(\mathcal{U}_j) = -(\log d)^{-1} \log P(\mathcal{U}_j)$ for all j and $L_q(\mathcal{U}_j) = -(\log d)^{-1} \log |\mathcal{U}_j|$ for all j.

Proposition 1. For all $q - c$ variable–length codes for a given LIDQS,

$$R_q + R_c \geq (\log d)^{-1} H(P)$$

with equality iff for all j

$$L_c(\mathcal{U}_j) = -(\log d)^{-1} \log P(\mathcal{U}_j), L_q(\mathcal{U}_j) = -(\log d)^{-1} \log |\mathcal{U}_j|,$$

and for all $u \in \mathcal{U}_j$,

$$P(u) = \frac{P(\mathcal{U}_j)}{|\mathcal{U}_j|}.$$

Proof: We have shown the lower bound and the "only if" part. The "if" part can be shown by a simple calculation (omitted).

Since in general von Neumann entropy may be smaller than Shannon entropy, the proposition shows that for LIDQS the von Neumann entropy is in general not tight. It provides a sharper bound. The reason to use Shannon entropy for an LIDQS is that we can model it as a classical source coding problem. That is, $q - c$ variable–length coding for an LIDQS is equivalent to a classical source coding problem, where the encoder can send messages via two classical noiseless channels, one of which can send variable–length codes (with rate R_q) and another can only send block codes (with rate R_c).

Proposition 1 provides another reason for which one might doubt whether von Neumann entropy fits lossless quantum data compression, where what "fits" means is guided by Classical Information Theory. Shannon entropy fits both lossy data compression and lossless data compression well because for both compressions it equals the optimal rate for a discrete memoryless source.

The well–known Schumacher quantum data compression theorem [12] and the alternative fidelity version in [10] show that in an analogous sense von Neumann entropy fits lossy quantum data compression well. Now let us consider lossless quantum data compression. Let $\mathcal{U}$ be an index set of states and P be a probability distribution. Then a memoryless

quantum source outputs a sequence of states $|u^n\rangle$ for $u^n = (u_1, \ldots, u_n) \in \mathcal{U}^n$ with probability $P^n(u^n) = \prod_{i=1}^{n} P(u_i)$. By Proposition 1, we have that for LIDQS with $r_q \triangleq \frac{1}{n} \sum_{j=1}^{J-1} P(\mathcal{U}_j) L_q(\mathcal{U}_j)$ and $r_c \triangleq \frac{1}{n} \sum_{j=1}^{J-1} P(\mathcal{U}_j) L_q(\mathcal{U}_j)$,

$$r_q + r_c \geq H(P). \tag{62}$$

Obviously (62) is asymptotically achieved by choosing $\mathcal{U}_j = \{|u\rangle : u^n \in \mathcal{T}_p^n\}$ (where j runs over all n–types). Since the Shannon entropy may be larger than the von Neumann entropy, for lossless quantum data compression of memoryless quantum sources von Neumann entropy in general is not achievable. This is a basic difference between classical and quantum data compression.

6 Future Research

To conclude, we present some problems for future research. The first problem is to determine the achievable rates of $q-c$ variable–length codes for a discrete memoryless quantum source. It is not hard to show that (R_q, R_c) is achievable (for R_q, R_c in (62)) iff

$$R_q + R_c \geq \lim_{\delta \downarrow 0} \lim_{n \uparrow \infty} \frac{1}{n} \log \dim \operatorname{span}\big(\{|u^n\rangle : u^n \in \mathcal{T}_{P,\delta}^n\}\big), \tag{63}$$

where $\mathcal{T}_{P,\delta}^n$ is the set of δ–typical sequences (see [5]), and the limits on the right hand side of (63) exist. But this is a "non–single–letter" bound or non–computable bound in the terminology of Classical Information Theory. The problem is to find a tight single–letter bound. We shall discuss the problem in a separate paper.

In Sections 4 and 5 we have seen that von Neumann entropy in general does not well fit lossless quantum data compression. On the other hand, to the best of our knowledge, almost all information quantities applied in quantum information theory are in terms of or closely related to von Neumann entropy. A challenging problem is to find a quantity with good and simple properties which better fits lossless quantum data compression.

Another problem is to study the identification problem analogous to [1] for the classical helper. (The same question seems to be not appropiate for quantum components because of observations I and II).

Finally, it is worth investigating connections to [7], which we became aware of after the present work was done.

References

[1] R. Ahlswede, B. Balkenhol and C. Kleinewächter, Identification for sources, Preprint 00–120, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 2000.

[2] K. Boström and T. Felbinger, Lossless quantum data compression and variable–length coding, preprint, 2002. PRA 65 0323/3

[3] S.L. Braunstein, C.A. Fuchs and D. Gottesman, A quantum analog of Huffman coding, IEEE Trans. on Inform. Theory, 46, 1644–1649, 2000, also http://xxx.lanl.gov/abs/quant-ph/9805080,1998.

[4] T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley and Sons, New York, 1991.

[5] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, New York–San Francisco–London, 1981.

[6] M. Fanes, A continuity property of the entropy density for spin lattice systems, Commun. Math. Phys. 31, 291–294, 1973.

[7] P. Hayden, R. Josza and A. Winter, Trading quantum for classical resourses in quantum data compression, J. Math. Phy., 43 (9), 4404–4444, 2002, also http://xxx.lanl.gov.Ph/0204038, 2002.

[8] A.S. Holevo, Statistical problems in quantum physics, In Gisiro Maruyama and Jurii V. Prokhorov ed. Proceeding of 2nd Japan–USSR Sym. 104–119, Springer–Verlag Berlin, 1973.

[9] A.S. Holevo, The capacity of the quantum channel, with general signal states, IEEE Trans. Inf. Theory, 44 (1), 269–273, 1998.

[10] M.A. Nielsen, Quantum Information Theory, PHD thesis, Univ. of New Mexico, 1998.

[11] M.A. Nielsen and I.L. Chuang, Quantum computation and quantum information, Cambridge, 2000.

[12] B. Schumacher, Quantum coding, Phys. Rev. A. 51, 2738–2747, 1995.

[13] B. Schumacher and M.P. Westmoreland, Sending classical information via noisy quantum channels, Phys. Rev. A, 56 (1), 131–138, 1997.

[14] B. Schumacher, M.P. Westmoreland, Indeterminate–length quantum coding, Phys. Rev. A, 64 (4), art. no. 042304, Oct. 2001, also http://xxx.lanl.gov./abs/quant-ph/0011011,2000.

[15] C.E. Shannon, A mathematical theory of communication, Bell. Syst. Tech. J. 27, 379–423, 1948.