6 On Lossless Quantum Data Compression and Quantum Variable–Length Codes

Rudolf Ahlswede and Ning Cai

Universität Bielefeld
Universitätsstrasse 25
D-33615 Bielefeld
Germany

6.1 Introduction

We survey results in the recently (in the late 90's) emerged area described in the title. The focus is on the compression rates, i.e. the average length of codewords.

In Shannon's Foundation of Information Theory ([27]) perhaps the most basic contributions are Source Coding Theorems (lossy and lossless) and the Channel Coding Theorem. In the most natural and simple source model DMS the source outputs a sequence $X_1, X_2, \ldots$ of independent, identically distributed random variables taking finitely many values. The Lossy Source Coding Theorem says that this sequence can be compresed by block coding with arbitrarily small error probability at a rate $H(P)$, the entropy of the common distribution P of the X_i's. (Later Shannon gave an extension replacing the probability of error criterion by general fidelity criteria and an ingenious formula for the rate–distortion function replacing $H(P)$.)

The Lossless Source Coding Theorem states that for variable–length codes the optimal data compression rate for an arbitrary source with distribution P is between $H(P)$ and $H(P)+1$.

Whereas the beginning of Quantum Information Theory can be traced back for instance to Holevo's paper [15] from 1973 it started flourishing only in the midnineties. With Schumacher's Quantum Lossy Source Coding Theorem [24] one of Shannon's basic results could be carried over to the quantum world: a memoryless quantum source generating a sequence $|x_1, x_2, \ldots, x_n\rangle$ of pure states with probability $P^n(x^n)$, where $x^n = (x_1, \ldots, x_n)$ is a sequence of indices of the states from a finite index set $\mathcal{X}$, can be compressed at rate $S(P)$, the von Neumann entropy of the state $\sum_{x\in\mathcal{X}} P(x)|x><x|$ with arbitrary high fidelity.

Subsequent work on quantum lossy data compression can be found in [4], [5], [12], [13], [14], [18], [19], [22], and [28]. ([25] gives significant progess on channel coding; "single–letterisations" of the capacity formula are still not known.)

However, the extension of the Lossless Source Coding Theorem meets an obvious barrier: a measurement of the length of codewords will disturb the message. Thus quantum data cannot be compressed losslessly if only the quantum resource is available. This was pointed out by many authors, e.g. in [8], [9], [26].

Nevertheless some applications of quantum variable–length codes have been found. We report on these as follows:

In Section 6.2 we present basic concepts and results on quantum variable–length codes, mainly from [8] and [26].

In Section 6.3 it is explained how in [9] and [26] long quantum codes are build from quantum variable–length codes.

In Section 6.4 the model of Boström/Felbinger [8] (based on Bolström's [6], [7]) for a quantum source with a classical *helper*, a classical channel informing the decoder about the lengths of codewords, is described.

In Sections 6.5 and 6.6 from [3] our recent results for this model and also our more general model are given.

In Section 6.7 for the first time in this survey we discuss a model and a result for *mixed* states, which are due to Koashi/Imoto [21].

In Section 6.8, finally we include a recent model for *lossy* quantum data compression, because it is related to the helper aspects.

6.2 Codes, Lengths, Kraft Inequality and von Neumann Entropy Bound

It is obvious that there is no way to compress losslessly classical nor quantum data by using block codes. So we always mean that a variable–length or in other words an indeterminate–length code (refered to in [26]) is employed, when we speak of lossless data compression.

6.2.1 The Codes

Quantum variable–length codes are defined by different authors, e.g. [8], [9] and [26]. The following is the definition by Boström and Felbinger in [8].

Let $\mathcal{H}$ be a Hilbert space of finite dimension d with an orthonormal basis

$$\mathcal{B}(\mathcal{H}) = \{|i>: i = 0, 1, 2, \ldots, d-1\}. \tag{6.1}$$

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

$$\mathcal{H}^{\oplus \ell_{\max}} = \mathcal{H} \oplus \mathcal{H}^{\otimes 2} \oplus \cdots \oplus \mathcal{H}^{\otimes \ell_{\max}}, \tag{6.2}$$

a Hilbert space of dimension $\sum_{\ell=1}^{\ell_{\max}} d^{\ell}$. Now suppose we are given an information source space $\mathcal{S}$, a Hilbert space of finite dimension d', say. Then it was defined in [8] that a variable–length encoder $\mathcal{E}$ of maximal length $\ell_{\max}$ is a linear isometric operator $\mathcal{E}$ 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 codeword space and (normalized) vectors (i.e., states) in it are called codewords.

To realize coding schemes, Schumacher and Westmoreland introduce zero–extended forms (zef) in [26]. For a codeword $\gamma^{\ell} \in \mathcal{H}^{\otimes \ell}$, its zef $|\gamma^{\ell} 0^{\ell_{\max}-\ell}\rangle$ is obtained by appending $(\ell_{\max}-\ell)$'s

$|0\rangle$ to it. zef of a superposition of codewords $|\gamma^{\ell_i}> \in \mathcal{H}^{\otimes \ell_i}$ $i = 1, 2, \ldots, k$ (which in itself is a codeword) is the superposition of zefs of $|\gamma^{\ell_i}\rangle$'s. Similarly, to realize a coding scheme, $|0\rangle$'s are padded at the end in [9] and in front of codewords in [8].

6.2.2 Length Observable and Average Length of Codewords

In classical information theory the lengths of codewords in a variable–length code are determinate e.g., in the code $\{0, 10, 11\}$ the codewords $0, 10, 11$ have length $1, 2, 2$ 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 (cf. (6.2)) because the encoder mapping is linear. So Schumacher and Westmoreland prefer to refer to the codes as "indeterminate codes" in [26]. One way to measure the lengths of codewords in this case is as follows ([26] and [8]). Let $\mathcal{H}^{\oplus \ell_{\max}}$ be the Hilbert space in (6.2) 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 length observable. Thus with the 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 by $\mathcal{L}$. Let

$$\wedge = \sum_{\ell=1}^{\ell_{\max}} \ell \mathcal{P}_\ell. \tag{6.3}$$

The expected outcoming length, also called the average length, of a codeword $|w\rangle$ is

$$\overline{L}(|w\rangle) = tr(|w\rangle\langle w|\wedge) = \langle w| \wedge |w\rangle. \tag{6.4}$$

6.2.3 Kraft Inequality and von Neumann Entropy Bound

The quantum prefix codes i.e., a codes such that no codeword is a prefix of other codewords, are well studied by Schumacher and Westmoreland in [26]. In particular in similar ways as in Classical Information Theory they proved:

Quantum Kraft Inequality: For all quantum prefix codes $\mathcal{C}$ and $D = \dim(\mathcal{H})$

$$\sum_{\ell=1}^{\ell_{\max}} \dim(\mathcal{C} \cap \mathcal{H}^{\otimes \ell}) D^{-\ell} \leq 1. \tag{6.5}$$

Like in the classical case, it was shown in [26] that the quantum Kraft inequality holds for all uniquely decodable codes (cf. Section 6.3). For non–uniquely decodable codes, Boström and Felbinger in [8] extend them to prefix codes and then obtain a Kraft–type inequality with an additional term, which depends on the extension of the codes and therefore on the structure of the codes.

von Neumann Entropy Bound [26]: Consider a quantum source which outputs a state $|s\rangle$ with probability $P(s)$, and $\sigma = \sum_s P(s)|s\rangle\langle s|$. Then for all uniquely decodable quantum

codes, in particular for all quantum prefix codes in (6.4) the expected average length of codewords with respect to the probability P is lowerbounded by the von Neumann entropy $S(\sigma)$ of the state σ.

6.2.4 Base Length

An important parameter, the base length $L(|w\rangle)$ of a codeword $|w\rangle$ in a quantum variable–length code was introduced in [8]:

$$L(|w\rangle) = \max\{\ell : \langle w|\mathcal{P}_\ell|w\rangle \ > 0\}. \tag{6.6}$$

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 $|w\rangle$

$$\overline{L}(|w\rangle) \leq L(|w\rangle). \tag{6.7}$$

To store a codeword of base length ℓ, one needs a quantum register of length at least ℓ. So it is necessary for the decoder to know the base length of codewords.

6.3 Construct Long Codes from Variable–length Codes

When they observed that it is impossible to losslessly compress quantum data if only quantum resources are available (cf. next section), Braustein and Fuchs in [9] suggested to apply quantum variable–length codes to construct a long block code in the following way. First connect N codewords of a quantum variable–length code and then truncate the obtained codeword and keep the first $N(\tilde{L}+\delta)$ components, where $\tilde{L}$ is the expectation of the average lengths with respect to the source distribution. Then a block code of length $N(\tilde{L}+\delta)$ with high fidelity is obtained.

Algorithms for the purposes of storage and communication are also presented in [9]. It is shown that, in both cases, the computational complexity using quantum variable–length codes to construct long block codes is remarkably lower than the best known algorithms.

Constructing block codes from quantum variable–length codes is systematically analysed in [26]. A transformation, called condensation, is introduced. A code is said to be *condensable* if for all N there exists a unitary operator U (depending on N) such that for all $\gamma^{\ell_i} \in \mathcal{H}^{\otimes\ell_i}$, $i = 1, 2, \ldots, N$

$$U|\gamma^{\ell_1}0^{\ell_{\max}-\ell_1}\rangle|\gamma^{\ell_2}0^{\ell_{\max}-\ell_2}\rangle \ldots |\gamma^{\ell_N}0^{\ell_{\max}-\ell_N}\rangle = |\Psi^{\sum_{i=1}^N \ell_i}0^{N\ell_{\max}-\sum_{i=1}^N \ell_i}\rangle \tag{6.8}$$

for a $|\Psi^{\sum_{i=1}^N \ell_i}\rangle \in \mathcal{H}^{\otimes\sum_{i=1}^N \ell_i}$, and the process is called *condensation*. The code is called *simply* condensable and the condensation is said to be simple if for all γ^{ℓ_i} $i = 1, 2, \ldots, N$, $|\Psi^{\sum_{i=1}^N \ell_i}\rangle$ in (6.8) is $|\gamma^{\ell_1}\gamma^{\ell_2}\ldots\gamma^{\ell_N}\rangle$. Then a prefix code is simply condensable. Obviously simply condensable codes are analogous to uniquely decodable codes in Classical Information Theory. So we also address simply condensable codes as uniquely decodable codes. Since unitary transformations are isoperimetric, a condensable code essentially is treated as a uniquely decodable code. In [26] quantum Kraft inequality and von Neumann entropy bound

(see Subsection 6.2.3) for condensable codes are established. Based on them it is shown in [26] that the rate of asymptotically optimal codes of high fidelity constructed by condensation equals von Neumann entropy for pure state sources generating $|s\rangle$ with probability $P(s)$. More efficient ways to use quantum variable–length codes to build long block codes are also presented in [26].

6.4 Lossless Quantum Data Compression, if the Decoder is Informed about the Base Lengths

In this and the next two sections we consider lossless quantum data compression for pure state quantum sources. We emphasize again

Observation I: A length measurement performed at a codeword of a quantum variable–length code will destroy the codewords.

That means, there must be a way to inform the decoder about the base lengths of codewords in a procedure of lossless quantum data compression since to decode correctly the decoder must know the base length of the decoded codeword. Moreover it is noticed in [3]

Observation II: In general, there is no way to measure the base length of unknown codewords *without error*.

So to inform the decoder about the lengths of codewords the encoder should know the output of the quantum source i.e., the output should be visible (by the encoder).

Boström and Felbinger in [8] suggest to code the message in the following way.

1) Quantum Source Output Visible by Encoder: 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 6.2.1). The output 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. Thus the decoder may store and decode the codewords correctly.

We point out here that the classical channel in their model only is used to inform about the lengths of codewords. Under this assumption the authors of [8] proposed the following coding scheme 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 $P(x_1) = \max\limits_{x \in \mathcal{X}_1} P(x)$ 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 $P(x_i) = \max\limits_{x'' \in \mathcal{X}_i} P(x'')$.

($a_{d'}$) 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).

6.5 Code Analysis Based on the Base Length

In our recent work [3] we systematically analyse the code of Boström and Felbinger [8], which is defined by the coding scheme 1), 2) in the previous section. To realize the coding scheme we expect that a codeword of base length ℓ can be stored in a quantum register of length ℓ. So we constrain the code $\mathcal{C}$ such that $\mathcal{C} \cap \mathcal{H}^{\otimes \ell}$ can be embedded in $\mathcal{H}^{\otimes \ell}$. Under this constraint we obtain a sufficient and necessary condition for the existence of codes of Section 6.4. For such a code $\mathcal{C}$ we denote by $\mathcal{C}_\ell$ the set of codewords of base lengths at most ℓ and by N_ℓ the number of codewords of base length ℓ. Then $\mathcal{C}_\ell$ is a linear subspace and the code exists iff for $\ell = 1, 2, \ldots, \ell_{\max}$

$$\mathcal{C}_1 \subset \mathcal{C}_2 \subset \cdots \subset \mathcal{C}_{\ell_{\max}}, \tag{6.9}$$

$$\dim \mathcal{C}_\ell \leq d^\ell \tag{6.10}$$

or equivalently

$$\sum_{i=1}^{\ell} N_\ell \leq d^\ell. \tag{6.11}$$

To realize the coding scheme we may obtain its ref by appending $|0\rangle$'s to its codewords. Then we obtain the canonical codes introduced in [3].

Moreover we determine the optimal compression rate for an arbitrary quantum pure state source.

Let $\mathcal{S}$ be a Hilbert space of dimension d', which will serve as a source space. $\mathcal{F}$ is a σ–field on $\mathcal{S}$ and P is a probability distribution over $\mathcal{F}$, which is not necessary discrete. Suppose a quantum source outputs pure states in $F \in \mathcal{F}$ with probability $P(F)$.

We call a sequence of subspaces $L = \{L_\ell : \ell = 1, 2, \ldots, \ell_{\max} - 1\}$ 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

$$\dim L_\ell = d^\ell, \tag{6.12}$$

$$L_1 \subset L_2 \subset \cdots \subset L_{\ell_{\max}-1}. \tag{6.13}$$

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

Theorem 6.1. (Ahlswede, Cai [3]) *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 helping channel informing about base lengths i.e., the codes in Section 6.4, is*

$$R_0 \triangleq \ell_{\max} - \sup_{L \in \mathcal{L}_d(\mathcal{S})} \sum_{\ell=1}^{\ell_{\max}-1} P(L_\ell). \tag{6.14}$$

6.6 Lossless Quantum Data Compression with a Classical Helper

We have seen in Section 6.4 that in the codes introduced by Boström and Felbinger, the classical channel only transmits the base lengths of codewords. As the lengths of codewords actually carry information we naturally ask ourselves "Why do'nt we use the classical channel to send other information?" The following example in [3] shows that we can do better.

Example: 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 a "continuous" source one may assume P is uniformly distributed on $\mathcal{S}_0 \cup \mathcal{S}_1$ and for the 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 that the maximum probability of 2–dimensional subspaces of $\mathcal{S}$ is $\frac{1}{2}$. So by Theorem 6.1 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.

As in the current source the probability is concentrated on $\mathcal{S}_0 \cup \mathcal{S}_1$ the encoder has a more clever way to compress the quantum source. He can just simply choose arbitrary two unitrary 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 now 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 a better code.

This simple example motivated us to look for a more efficient way to use the classical helper. By Observations I and II in Section 6.4 the following assumptions are necessary.

(1) Visible 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.

Under these assumptions we have the following coding scheme.

We let $\mathcal{H}$ and $\mathcal{S}$ be complex Hilbert spaces of dimensions d and d' respectively and P be a probability distribution with support set $\mathcal{U} \subset \mathcal{S}$.

Suppose a quantum source outputs a state $|u\rangle \in \mathcal{S}$ with probability $P(u)$. 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:

(I) Partition $\mathcal{U}$ properly into $\{\mathcal{U}_j : j = 0, 1, \ldots, J-1\}$ for an integers J. For each j find the minimum ℓ_j such that there is an $d\ell_j$–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 unitrary 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 $|w(u)\rangle \triangleq \mathcal{U}_j|u\rangle \in \mathcal{H}^{\otimes L_q(\mathcal{U}_j)}$ 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 classical variable–length code e.g., 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 $|w(u)\rangle$.

It is not hard to see that this coding scheme is most general under the two assumptions, there is no better code than the best codes constructed by this coding scheme.

The key step is how to choose the partition in (I) and it is actually the most difficult part in the coding scheme.

We call a code constructed by coding scheme a quantum–classical variable–length code, or shortly a $q-c$ varibale–length code and its two components, quantum and classical components respectively and speak of lossless quantum data compression with classical helper.

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 when $|u\rangle \in \mathcal{U}_j$. Then the classical and quantum components of the compression rate are

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

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

respectively. By Shannon's Lossless Source Coding Theorem ([27]; also in [10], [11]), with the notation $Q \triangleq \big\{Q(j) = P(\mathcal{U}_j) : j = 0, 1, \ldots, J-1\big\}$, R_c is bounded by

$$(\log a)^{-1} H(Q) \leq R_c < (\log a)^{-1} H(Q) + 1, \tag{6.15}$$

where a is the size of the alphabet of the classical code and H is Shannon's entropy.

To simplify notation, in the sequel we assume the size of the classical alphabet to be $a = d = \dim \mathcal{H}$. Then we have

Theorem 6.2. (Ahlswede, Cai [3]) *For any $q - c$ variable–length code,*

$$R_q + R_c \geq (\log d)^{-1} S(\rho) \tag{6.16}$$

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

$$\rho \triangleq \sum_{u \in \mathcal{U}} P(u) |u><u|, \tag{6.17}$$

and equality holds iff the following conditions hold simultaneously.

(i) For the probability Q in (6.15), i.e. $Q(j) = P(\mathcal{U}_j)$,

$$R_c = (\log d)^{-1} H(Q). \tag{6.18}$$

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

$$\mathcal{S}_j \perp \mathcal{S}_{j'}, \tag{6.19}$$

and

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

$$P(\mathcal{U}_j)^{-1} \sum_{u \in \mathcal{U}_j} P(u) |u\rangle\langle u| = d_j'^{-1} \mathcal{P}_j, \tag{6.20}$$

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

When the support set of the source distribution is a set of independent pure states, we have a sharper bound.

Proposition 6.1 (Ahlswede, Cai [3]). *Let $|u\rangle$, $u \in \mathcal{U}$, be a set of independent pure states, let P be a probability distribution on $\mathcal{U}$ and let a quantum source output $|u\rangle$ with probability $P(u)$, where $\mathcal{U}$ is a finite index set. Then for all $q - c$ variable–length codes for the source*

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

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

We conclude this section with a few *Problems*, which we pose on lossless quantum data compression for pure state sources:

1. In [3] we showed that the gap between von Neumann entropy and the optimal compression rates in Theorem 6.2 may be arbitrary large. On the other hand by our knowledge successfully used quantum information measures are all in terms of von Neumann entropy. So we ask "Is there a quantity better to fit lossless quantum data compression than von Neumann entropy?"

2. For an arbitrary discrete memoryless quantum pure state source determine the optimal compression rate. For this problem we know that von Neumann entropy and Shannon entropy are lower and upper bounds and in general neither bound is tight.

3. Study other models of quantum data compression e.g., the quantum version of the identification problem treated in [2], which was introduced in the context of [1].

6.7 Lossless Quantum Data Compression for Mixed State Sources

We report here the work of Koashi and Imoto [21]. This is their model:

— A quantum source outputs for $i = 1, \ldots, I$ mixed states ρ_i in a Hilbert space $\mathcal{H}'_A$ with probability p_i.

— The measurement of lengths of codewords is performed in an auxiliary quantum system $\mathcal{H}_E$ so that it will (by assumption) not disturb the message.

More precisely the encoding–decoding operator is specified by a unitary operator U acting on $\mathcal{H}'_A \otimes \mathcal{H}_E$ such that for $i = 1, 2, \ldots, I$

$$tr_E\big[U(\rho_i \otimes \Sigma_E)U'\big] = \rho_i, \tag{6.21}$$

where $\mathcal{H}_E$ is an auxiliary system initially prepared in a pure state Σ_E. They assume that there is an observable $\mathcal{L}$ acting on $\mathcal{H}_E$, which corresponds to the lengths of codewords such that the expected length of codewords for $\rho = \sum_i P_i\rho_i$,

$$\bar{\bar{L}} = tr_E\big\{\mathcal{L}\ell_{\gamma_A}\big[U(\rho \otimes \Sigma_E)U'\big]\big\}. \tag{6.22}$$

The coding theorem is based on their previous work [20], where it was shown that for a set $\{\rho_i\}_i$ if mixed states a probability distribution $\{P_i\}_i$ and $\rho = \sum_i \rho_i P_i$, there is a unique decomposition of the support set $\mathcal{H}_A$ of ρ such that

$$\mathcal{H}_A = \oplus_\ell \mathcal{H}_J^{(\ell)} \otimes \mathcal{H}_K^{(\ell)} \tag{6.23}$$

and for all i

$$\rho_i = \oplus_\ell q^{(i,\ell)} \rho_J^{(i,\ell)} \otimes \rho_K^{(\ell)}, \tag{6.24}$$

where $\rho_J^{(i,\ell)}$ and $\rho_K^{(\ell)}$ are normalized density operators acting on $\mathcal{H}_J^{(\ell)}$ and $\mathcal{H}_K^{(\ell)}$, respectively, $q^{(i,\ell)}$ is the probability for the states to be in the subspace $\mathcal{H}_J^{(\ell)} \otimes \mathcal{H}_K^{(\ell)}$, $\rho_K^{(\ell)}$ is independent of i and $\{\rho_J^{(i,\ell)}\}_i$ cannot be expressed in a simultaneously block–diagonalized form. For a

quantum source outputing mixed state ρ_i with probability P_i, we denote $P(\ell) = \sum_i P_i q^{(i,\ell)}$, $I_c = -\sum_\ell P(\ell)\log P(\ell) = H(P)$ and $P_{NC} = \sum_\ell P(\ell)\log\dim\mathcal{H}_J^{(\ell)}$.

Let $\overline{R}$ be the optimal compression rate for this model. Then the coding theorem in [21] says

$$I_c + D_{NC} \leq \overline{R} \leq I_c + D_{NC} + 2. \tag{6.25}$$

6.8 A Result on Tradeoff between Quantum and Classical Resources in Lossy Quantum Data Compression

In this last section we report a result on lossy quantum data compression due to Hayden, Jozsa, and Winter in [14], because it relates to the helper model, best briefly, because it concerns the lossy case whereas this survey primarily adresses the lossless case.

This is the model:

— A quantum DMS outputs a pure state $|u\rangle$, $u \in \mathcal{U}$ with probability $Q(u)$ and consequently outputs a sequence $|u^n\rangle = |u_1u_2\ldots u_n\rangle$, $u^n \in \mathcal{U}^n$ with probability $Q^n(u^n) = \prod_{i=1}^n Q(u_i)$. In other words an ensembles $\mathcal{E} = \{|u\rangle, P(u)\}$ u is given.

— Assume that the encoder can send messages to the decoder via a classical channel at rate R bits per signa,.

— Then the trade–off function $Q^*(R)$ is defined as the asymptotically optimal compression rate (qbits per signal) with an arbitrarily high fidelity under the above assumptions.

To compute $Q^*(R)$ the authors Hayden, Jozsa, and Winter decompose the ensemble $\mathcal{E} = \{|u\rangle, P(u)\}$ into at most $|\mathcal{U}|+1$ ensembles $\mathcal{E}_j = \{|u\rangle, W(u|j)\}$ with weight $P(j)$ and their union $\cup_j P(j)\mathcal{E}_j$ reporduces $\mathcal{E}$. This is equivalent to decompositing the probability distribution Q by an input distribution P over $\{0,1,\ldots,|\mathcal{U}|\}$ and using a classical channel $W : \{0,1,\ldots,|\mathcal{U}|\} \to \mathcal{U}$ such that for all $u \in \mathcal{U}$ $Q(u) = \sum_j P(j)W(u|j)$. Let $\mathcal{D}(R)$ be the set of decompositions with $I(P;W) = R$ and $\overline{S}(P) = \sum_j P(j)S(\mathcal{E}_j)$, where I is Shannon's mutual information and S is von Neumann entropy. Then

Theorem 6.3. (Hayden, Jozsa, Winter [13])

$$Q^*(R) = \min_{(P,W)\in\mathcal{D}(R)} \overline{S}(R).$$

Finally we remark that in addition to the difference between the models in Sections 4 – 6 and this section with respect to the property lossless versus property lossy, another difference is that we deal with general quantum sources in Sections 4 – 6 and with quantum DMS in this section.

Moreover, in Sections 6.4 and 6.5 the decoder allows only to send lengths of codewords via the classical channel and there is not such a restriction in sections 6.6 and 6.7.

References

[1] R. Ahlswede, General theory of information transfer, Preprint 97–118, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld.

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

[3] R. Ahlswede and N. Cai, On Lossless quantum data compression with a classical helper, submitted to IEEE Trans. Inf. Theory.

[4] H. Bornum, C.M. Caves, C.A. Fuchs, R. Jozsa, and B. Schumacher, On quantum coding for ensembles of mixed states, http://xxx.lanl.gov/abs/quant-ph/0008024v1,2000.

[5] H. Barnum, C.A. Fuchs, R. Jozsa, and B. Schumacher, General fidelity limit for quantum channels, Phys. Rev. A 54, 4707, 1996.

[6] K. Boström, Concept of a quantum information theory of many letters, http://xxx.lanl.gov/abs/quant-ph/0009052, 2000.

[7] K. Boström, Lossless quantum coding in many–letter space, http://xxx.lanl.gov/abs/quant-ph/0009073, 2000.

[8] K. Boström and T. Felbinger, Lossless quantum data compression and variable–length coding, preprint, 2002.

[9] S.L. Braunstein and C.A. Fuchs, A quantum analog of Huffman coding. http://xxx.lanl.gov/abs/quant-ph/9805080,1998.

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

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

[12] I. Devetak and T. Berger, Quantum rate–distortion theory for memoryless sources, IEEE Trans. Inform. Theory, Vol. 48, 1580–1589, 2000.

[13] M. Hayashi, Exponents of quantum fixed–length pure state source coding, http://xxx.lanl.giv/abs/quant-ph/0202002, 2002.

[14] P. Hayden, R. Josza and A. Winter, Trading quantum for classical resourses in quantum data compression, http://xxx,Canl.gov.Ph/0204038, 2002.

[15] 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.

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

[17] D.A. Huffman, A method for the construction of minimum redundancy codes, Proc. IRE 40, 1098–1101, 1952.

[18] R. Jozsa and B. Schumacher, A new proof of the quantum noiseless theorem, Mod. Opt. 14, 2343, 1994.

[19] M. Koashi and N. Imoto, Compressibility of mixed–state signals, http://xxx.lanl.giv/abs/quant-ph/0103128v1, 2001.

[20] M. Koashi and N. Imoto, What is possible without disturbing partially quantum states, http://xxx.lanl.gov/abs/quant-ph/01011444v2, 2001.

[21] M. Koashi and N. Imoto, Quantum information is incompressible without errors, http://xxx.lanl.gov/abs/quant-ph/0203045v1, 2002.

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

[23] M.A. Nielsen and I.L. Chuang, Quantum Computation and Quantum Information, Cambridge, 2000.

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

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

[26] B. Schumacher, M.P. Westmoreland, Indeterminate–length quantum coding, http://xxx.lanl.gov./abs/quant-ph/0011011, 2000.

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

[28] A. Winter, Coding theorems of quantum information theory, PhD. Thesis, Univ. Bielefeld, 1999.