OPTIMAL LOWER BOUND FOR LOSSLESS QUANTUM BLOCK ENCODING

GEORGE ANDROULAKIS

Department of Mathematics, University of South Carolina
Columbia, SC 29208, USA
giorgis@math.sc.edu

RABINS WOSTI

Department of Computer Science and Engineering, University of South Carolina
Columbia, SC 29208, USA
rwosti@email.sc.edu

Consider a general quantum stochastic source that emits at discrete time steps quantum pure states which are chosen from a finite alphabet according to some probability distribution which may depend on the whole history. Also, fix two positive integers m and l. We encode any tensor product of ml many states emitted by the quantum stochastic source by breaking the tensor product into m many blocks where each block has length l, and considering sequences of m many isometries so that each isometry encodes one of these blocks into the Fock space, followed by the concatenation of their images. We only consider certain sequences of such isometries that we call "special block codes" in order to ensure that the string of encoded states is uniquely decodable. We compute the minimum average codeword length of these encodings which depends on the quantum source and the integers m, l, among all possible special block codes. Our result extends the result of [Bellomo, Bosyk, Holik and Zozor, Scientific Reports 7.1 (2017): 14765] where the minimum was computed for one block, i.e. for $m = 1$. Lastly, we give a simplified non-adaptive compression technique based on constrained special block codes for general quantum stochastic sources. For quantum stationary sources in particular, we show that the minimum average codeword length per symbol computed over all constrained special block codes is equal to the von-Neumann entropy rate of the source for an asymptotically long block size.

Keywords: Indeterminate length, Kraft-McMillan Inequality, uniquely decodable code, Fock space

1   Introduction

Important foundations to the area of quantum encoding were provided by Schumacher who proved the quantum analog of Shannon's noiseless coding theorem for an independent and identically distributed (iid) quantum source, [32]. In this article we study lossless, variable-length block encoding of quantum information which is emitted from a completely general quantum source, and it is encoded into the Fock space. Since the Schumacher's mentioned result, various block encoding schemes for quantum sources have been proposed. We review these schemes in the next few paragraphs of the current section. At the end of the section we describe our contribution in more detail.

Chuang and Modha [12] proposed a source-dependent quantum algorithm to project the block quantum state of an iid quantum source to its "typical subspace". The projected state is compressed using the quantum analog of Shannon-Fano code. According to that code, the number of qubits per signal is only slightly larger than the von Neumann entropy of the source for sufficiently large block size.

A universal fixed-length quantum block compression scheme was introduced by Jozsa et. al. [21] for a family of iid quantum sources such that the von Neumann entropy of each source in the family is known to be upper bounded by a fixed number S. Their encoding is source-independent, in the sense that one does not need to know the average ensemble state which is emitted by the quantum source. In particular, they showed that for a sufficiently large block size, their universal encoding compresses the quantum information from any of these sources to S qubits/signal and is asymptotically optimal. Kalchenko and Yang extended this result in [22, 23]. They proved that for a family of stationary ergodic quantum sources such that the von Neumann entropy rate of each source in the family is upper bounded by a real number $r > 0$, there exists a sequence of universal projectors which compresses the quantum information from any of these sources to r qubits/signal with high fidelity for sufficiently large block size. There have been additional works on ergodic quantum sources [6, 7, 8, 25, 30]. Universal variable-length lossless quantum data compression schemes have been discussed by Hayashi and Matsumoto in [18, 19] and by Hayashi in [17].

Another source-dependent quantum block compression was introduced by Szeto [35] for an iid quantum source. This encoding achieves the same compression performance as Schumacher's noiseless coding theorem [32], but employs a classical block compression alongside, and adopts the assumption that the Hilbert space can be decomposed into two or more mutually orthogonal subspaces and each signal from the source comes from one of these subspaces.

Langford [27] gave a compilation algorithm that takes two inputs. The first input is any classical block compression scheme that uses bounded space and time, and the second input is the density matrix of an iid quantum source. The algorithm outputs a fixed-length quantum block compression algorithm. If the input classical algorithm asymptotically approaches the Shannon entropy of the classical source, then the output quantum block compression algorithm also asymptotically approaches the von Neumann entropy of the input quantum source.

There are several completely lossless compression algorithms in the classical information theory, such as Huffman encoding [20], Shannon-Fano encoding [34, 14]and arithmetic encoding [26]. It is natural to attempt to generalize these classical algorithms to the quantum realm in order to obtain lossless quantum data compression. Classical completely lossless block compression algorithms make use of variable-length codes where sequences of symbols with high probability of occurrence are assigned codewords of shorter length and less probable sequences have longer codewords, thereby reducing the average codeword length. This idea extends to quantum data compression. In the quantum setting, not only the codewords can have different lengths, but also the codewords can be in superposition of states of different length and they are called indeterminate-length codewords. Boström and Felbinger [9] show that it is impossible to achieve a completely lossless quantum block compression scheme by using fixed-length codewords. Therefore, the indeterminate-length codewords form an integral part of lossless quantum data compression, and were investigated first by Schumacher

[31] and later was formalized by Schumacher and Westmoreland [33].

The quantum analogue of classical Huffman code was investigated by Braunstein, Fuchs, Gottesman and Lo [10]. Schumacher and Westmoreland [33] proved that for an iid source, the quantum Huffman code is optimal in the sense that it minimizes the average codeword length which asymptotically approaches the von Neumann entropy of the source as the block size increases to infinity. For their proof they used the formalism of condensable codes. Bellomo, Bosyk, Holik and Zozor [5, Theorem 2] proved later that the quantum Huffman code minimizes the average codeword length of a single block for any given quantum source, (not necessarily iid). There has also been some parallel works [29, 11] towards minimizing the average base length (which is the maximum length of an indeterminate-length codeword used) for quantum sources as opposed to minimizing the average codeword length.

During the recent years, there have been many developments in quantum communication and data compression. We refer the reader to the recent textbooks [24] by Khatri and Wilde and [15] by Hajdušek and Van Meter for a comprehensive review on the subject.

In classical (or quantum) data transmission, the number of bits (or qubits) required to transmit any given classical (or quantum) message is considered as a resource. So, it is desirable to transmit a given classical (or quantum) message using as few bits (or qubits) as possible. Therefore, in this paper, we focus on quantum block data compression in an effort to minimize the average codeword length for a given length of the quantum message emitted by a given quantum source. Precisely, we study *completely lossless* and *indeterminate-length* quantum block data compression of pure states emitted by a *quantum stochastic source*, which is not necessarily iid. We will follow the formalism in [5] as opposed to the zero-extended forms of [33] in order to represent indeterminate-length codewords.

Consider a *quantum stochastic source* (not necessarily iid) and a tensor product of $q \in \mathbb{N}$ many pure states emitted from the source. Fix $m \in \mathbb{N}$ as the number of blocks and l as the block size such that $q = ml$. Consider a sequence of m many isometries (also called quantum codes) $U_1, \ldots, U_m$ such that for $k \in \{1, \ldots, m\}$, each U_k encodes the kth block consisting of the tensor product of l many pure states into the Fock space. The given sequence of m isometries thus encodes the tensor product of ml many states to the state obtained by the concatenation of the images of all U_k's. In order to decode the encoded state in a lossless manner, the sequence of quantum codes must be "uniquely decodable" (also called completely lossless) quantum code. Additionally, this kind of approach of block encoding of the pure states raises some interesting questions: for a given message consisting of the tensor product of $q = ml$ many pure states emitted from a given quantum stochastic source $\mathcal{S}$, what is the minimum average codeword length per symbol computed over all uniquely decodable quantum codes $U_1, \ldots, U_m$ such that each U_k encodes the kth block of l many pure states? How does the minimum average codeword length per symbol computed over all uniquely decodable quantum codes change if we change m and l such that their product ml always remains constant?

To this end, in this paper we extend the notion of uniquely decodable (or completely lossless) quantum codes to be used for quantum block data compression. As the main result (Theorem 2.18), for a fixed ml many pure states emitted by a given quantum stochastic source, we derive the optimal lower bound of the average codeword length over a subset of uniquely decodable quantum codes called "special block codes", which are applied to encode the pure states in m many blocks each of block size l. We hope that our work would be

supportive in the quest towards answering the above questions, which have been left open in the current work.

We explicitly show that in order to achieve the optimal lower bound, for every $k \in \{1, \ldots, m\}$, one needs to dynamically adapt the isometry used for encoding the kth block of l many pure states based on the exact sequence of pure states emitted in the first $(k-1)$ blocks. However, since the pure states emitted by the source are not necessarily orthogonal, the task of state discrimination of non-orthogonal states can be difficult. So, in Section 3, we give a simplified non-adaptive compression technique based on a subset of special block codes called "constrained special block codes" for quantum stochastic sources that emit pure states where the encoder can fix one isometry per block of states ignoring the emissions in the previous blocks, and in Theorem 3.1 we derive the optimal lower bound of the average codeword length over all constrained special block codes which are applied to encode the pure states in m many blocks each of block size l. In the Corollary 3.2, we show that for quantum stationary sources in particular, the optimal lower bound of the average codeword length per symbol computed over all constrained special block codes equals the von-Neumann entropy rate of the source for an asymptotically long block size.

2   Preliminary definitions

A classical binary encoding for a finite alphabet set A is a map from the set A to the set of all finite strings of 0's and 1's. The images of the elements of the alphabet via an encoding are called codewords. Every encoding of an alphabet set A can be extended to the set of finite strings of elements of A by concatenation. An encoding is called uniquely decodable if its extension is one-to-one. An encoding is called instantaneous (or prefix-free) if no codeword is the initial part of another codeword. Instantaneous encodings are special cases of uniquely decodable ones. If the information source is independent and identically distributed (iid), then a probability distribution is assigned to the alphabet, and the average length of the code is defined to be equal to the expectation of the length of its codewords, where the length of each codeword is the number of its bits. The Huffman encoding [20] is an optimal instantaneous classical encoding and its average codeword length l_{avg} satisfies

$$H(X) \leq l_{avg} \leq H(X) + 1$$

where $H(X)$ is Shannon entropy of a classical source.

Now we look at the analogous situation of the quantum data compression. Throughout this article, we fix a Hilbert space $\mathcal{H}$ of dimension d and a set $\{|s_n\rangle\}_{n=1}^N$ of pure states of $\mathcal{H}$. We assume without loss of generality that the vectors $\{|s_n\rangle\}_{n=1}^N$ span the Hilbert space $\mathcal{H}$. In this section we recall the definitions of quantum stochastic sources, quantum stochastic ensembles, quantum codes and average codeword length which have been considered in some of the references mentioned above.

Definition 2.1 *A **quantum stochastic source** $\mathcal{S}$ consists of a set of pure states $\{|s_n\rangle\}_{n=1}^N$ of a Hilbert space $\mathcal{H}$, and a stochastic process $X = (X_n)_{n=1}^\infty$, where each X_n is a random variable which takes values in $\{1, 2, \ldots, N\}$. At every positive integer time n the state $|s_{X_n}\rangle$ is emitted from the quantum source. If p denotes the probability distribution of the stochastic process X, then for every $q \in \mathbb{N}$ and $(n_1, \ldots, n_q) \in \{1, \ldots, N\}^q$, we have that*

$$p(n_1, \ldots, n_q) = \mathbb{P}(X_1 = n_1, \ldots, X_q = n_q). \tag{1}$$

Also, the conditional probability distribution of the stochastic process X is defined for any $q \geq 2$ and $(n_1, \ldots, n_q) \in \{1, \ldots, N\}^q$ by

$$p(n_q | n_{q-1}, \ldots, n_1) = \mathbb{P}(X_q = n_q | X_{q-1} = n_{q-1}, \ldots, X_1 = n_1).$$

Definition 2.2 *A quantum stochastic source $\mathcal{S}$ is **stationary (or translation-invariant)** if the associated stochastic process X is invariant with respect to the translation map, i.e.*

$$\mathbb{P}(X_1 = n_1, \ldots, X_q = n_q) = \mathbb{P}(X_{k+1} = n_1, \ldots, X_{k+q} = n_q) \tag{2}$$

for every $k \in \mathbb{N}$, $q \in \mathbb{N}$ and $(n_1, \ldots, n_q) \in \{1, \ldots, N\}^q$.

Setting $q = 1$ in Equation (2), we get $\mathbb{P}(X_1 = n_1) = \mathbb{P}(X_{k+1} = n_1)$ for every $k \in \mathbb{N}$ and $n_1 \in \{1, \ldots, N\}$, which shows that for every quantum stationary source, the random variables are identically distributed. Notice however that the random variables are not necessarily independent. We will deal with quantum stationary sources in Section 3.

Definition 2.3 *Given a quantum stochastic source $\mathcal{S}$ which consists of a set of pure states $\{|s_n\rangle\}_{n=1}^N$ of a Hilbert space $\mathcal{H}$, and a stochastic process $X = (X_n)_{n=1}^\infty$ with probability distribution p, we define the **ensemble state** ρ_q for $q \in \mathbb{N}$ many emissions from the source as follows:*

$$\rho_q = \sum_{n_1, \ldots, n_q = 1}^N p(n_1, \ldots, n_q) \left| s_{n_1} \cdots s_{n_q} \middle\rangle\!\middle\langle s_{n_1} \cdots s_{n_q} \right|.$$

In this article, we are concerned with quantum block encodings and we assume that each block that we encode is a tensor product of l many pure states which are chosen from the set $\{|s_n\rangle : n = 1, \ldots, N\}$ via the stochastic process $(X_n)_{n\in\mathbb{N}}$. Usually we express that fact by saying that **the block size** is equal to l. Thus, we often look at multiples of l many pure states. This can be observed in the next definition.

Definition 2.4 *Let a quantum stochastic source $\mathcal{S}$ which consists of a set of pure states $\{|s_n\rangle\}_{n=1}^N$ of a Hilbert space $\mathcal{H}$, and a stochastic process $X = (X_n)_{n=1}^\infty$ with probability distribution p. For a fixed block size $l \in \mathbb{N}$, an integer $k \geq 1$, and a fixed sequence $(n_1, \ldots, n_{(k-1)l}) \in \{1, \ldots, N\}^{(k-1)l}$ define the **k^{th} block conditional ensemble state** $\rho^{n_1, \ldots, n_{(k-1)l}}$ to be equal to*

$$\sum_{n_{(k-1)l+1}, \ldots, n_{kl} = 1}^N \frac{p(n_1, \ldots, n_{kl})}{p(n_1, \ldots, n_{(k-1)l})} \left| s_{n_{(k-1)l+1}} \cdots s_{n_{kl}} \middle\rangle\!\middle\langle s_{n_{(k-1)l+1}} \cdots s_{n_{kl}} \right|$$

$$= \sum_{n_{(k-1)l+1}, \ldots, n_{kl} = 1}^N p(n_{(k-1)l+1}, \ldots, n_{kl} | n_1, \ldots, n_{(k-1)l}) \left| s_{n_{(k-1)l+1}} \cdots s_{n_{kl}} \middle\rangle\!\middle\langle s_{n_{(k-1)l+1}} \cdots s_{n_{kl}} \right|, \tag{3}$$

*where each coefficient $p(n_{(k-1)l+1}, \ldots, n_{kl} | n_1, \ldots, n_{(k-1)l})$ above represents the conditional probability that the sequence of states $\left| s_{n_{(k-1)l+1}}, \ldots, s_{n_{kl}} \right\rangle$ gets emitted in the kth block given the emission of the sequence of states $\left| s_{n_1}, \ldots, s_{n_{(k-1)l}} \right\rangle$ in the first $(k-1)$ many blocks. Consequently, the **k^{th} block ensemble state** ρ^k is given by*

$$\rho^k = \sum_{n_1, \ldots, n_{(k-1)l} = 1}^N p(n_1, \ldots, n_{(k-1)l}) \rho^{n_1, \ldots, n_{(k-1)l}} \tag{4}$$

In what follows, for $k = 1$, the state $\rho^{n_1,\ldots,n_{(k-1)l}}$ is interpreted as the ensemble state ρ_ℓ (or equivalently, ρ^1) of the first block, $p(n_1,\ldots,n_{(k-1)l})$ is set to 1, and the string $n_1,\ldots,n_{(k-1)l}$ is the empty string and is dropped from the superscript wherever applicable.

Recall that **Fock space** $(\mathbb{C}^2)^\oplus$ is associated to $\mathbb{C}^2$ and is defined to be the Hilbert space

$$(\mathbb{C}^2)^\oplus = \bigoplus_{k=0}^{\infty}(\mathbb{C}^2)^{\otimes k}, \tag{5}$$

where $(\mathbb{C}^2)^{\otimes 0} = \mathbb{C}$. We simply refer to the Fock space associated to $\mathbb{C}^2$ as the **Fock space**. The **standard basis** of the Fock space $(\mathbb{C}^2)^\oplus$ consists of qubit strings of various lengths, (including the 0-length $|\emptyset\rangle$). If a quantum state $|\psi\rangle$ includes the superposition of only the qubit strings of some length $\ell \in \mathbb{N}$, i.e. $|\psi\rangle \in (\mathbb{C}^2)^{\otimes \ell}$, we refer to such state as **length state**.

Now, we define the notion of quantum codes.

Definition 2.5 *Let $\mathcal{K}$ be a Hilbert space of dimension D. A **quantum code** on $\mathcal{K}$ is a linear isometry $U : \mathcal{K} \to (\mathbb{C}^2)^\oplus$. Thus, for every quantum code U on $\mathcal{K}$, the dimension of its range is equal to D, and if $(|\psi_i\rangle)_{i=1}^{D}$ is any orthonormal sequence in its range, then U has the form*

$$U = \sum_{i=1}^{D} |\psi_i\rangle\langle e_i|, \tag{6}$$

*where $(|e_i\rangle)_{i=1}^{D}$ is an orthonormal basis of $\mathcal{K}$. The quantum state of the form $U|s\rangle$ obtained by applying U to any pure state $|s\rangle \in \mathcal{K}$ is called **codeword**. In particular, if the codeword is a length state, it is called **length codeword**. However, in general, a codeword can be a superposition of lengths states of different lengths, in which case the codeword is called **indeterminate-length codeword**.*

As a result, the range of a quantum code is the finite dimensional subspace of the Fock space that is spanned by the orthonormal vectors $|\psi_i\rangle$'s, possibly of different lengths. The notion of concatenation of codewords is defined as follows:

Definition 2.6 ***Concatenation** of standard basis vectors of the Fock space is defined as the concatenation of classical bit strings. Then, **concatenation** of the linear combinations of basis vectors of the Fock space is defined by making the concatenation distributive over addition and scalar multiplication, as in [28, Definition 2.4] and [29, Section V]. Hence, for $m \in \mathbb{N}$ and $x_1,\ldots,x_m \in (\mathbb{C}^2)^\oplus$, their concatenation $x_1 \circ x_2 \circ \cdots \circ x_m$ is an element of $(\mathbb{C}^2)^\oplus$. Let $\mathcal{K}$ be a finite dimensional Hilbert space, $m \in \mathbb{N}$, and $U_1,\ldots,U_m$ be a sequence of quantum codes with $U_k : \mathcal{K} \to (\mathbb{C}^2)^\oplus$ for all $k = 1,\ldots,m$. Then, the **concatenation** of the sequence $(U_k)_{k=1}^{m}$ is defined to be the linear map $U_1 \circ \cdots \circ U_m : \mathcal{K}^{\otimes m} \to (\mathbb{C}^2)^\oplus$ given by*

$$(U_1 \circ \cdots \circ U_m)|s_1 \otimes \ldots \otimes s_m\rangle = U_1|s_1\rangle \circ \cdots \circ U_m|s_m\rangle \text{ for all pure states } |s_1 s_2 \ldots s_m\rangle \in \mathcal{K}^{\otimes m}, \tag{7}$$

where $\circ$ denotes the concatenation operation.

Remark 2.7 *It is important to note that concatenation of arbitrary pure quantum states does not always preserve the norm, and hence may not be well defined. It's not hard to see that the norm is always preserved if either or both of the pure quantum states being concatenated are length states. However, if both pure quantum states being concatenated are indeterminate-length states, then the norm may not be preserved (for example, concatenation of the pure*

states $\frac{1}{\sqrt{2}}(|0\rangle + |00\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |00\rangle$ is not normalized). This has been well explored in [28, after Definition 2.4]. On the other hand, Remark 4.1 will give that on certain subspaces of the Fock space, concatenations of fixed number of vectors are well defined.

Thus, if $(U_k)_{k=1}^m$ is a finite sequence of quantum codes, then $U_1 \circ \cdots \circ U_m$ in general may not be an isometry. It is important that the concatenation $U_1 \circ \cdots \circ U_m$ remains an isometry, because if U_k is used to encode the kth block for each $k = 1, \ldots, m$, then $U_1 \circ \cdots \circ U_m$ will be used to encode the tensor product of m many blocks of pure states, so its inverse will be used to decode those blocks back in a lossless fashion. This lossless way of decoding is possible if the concatenation $U_1 \circ \cdots \circ U_m$ is uniquely decodable which is explained below.

We now extend the notion, in some sense, of uniquely decodable quantum codes considered in [17, 16, 5, 3] from one isometry to a sequence of isometries.

Definition 2.8 *Let $\mathcal{K}$ be a finite dimensional Hilbert space, and $U_1, \ldots, U_m$ be a finite sequence of quantum codes such that $U_k : \mathcal{K} \to (\mathbb{C}^2)^{\oplus}$ for every $k = 1, \ldots, m$. The sequence of quantum codes $(U_k)_{k=1}^m$ is called **uniquely decodable** if and only if their concatenation $U_1 \circ \cdots \circ U_m : \mathcal{K}^{\otimes m} \to (\mathbb{C}^2)^{\oplus}$ is an isometry.*

We now give a characterization of uniquely decodable quantum codes that follows directly from the Definitions 2.6 and 2.8.

Theorem 2.9 *Let $\mathcal{K}$ be a D-dimensional Hilbert space and $(U_k)_{k=1}^m$ be a sequence of isometries $U_k : \mathcal{K} \to (C^2)^{\oplus}$ with $U_k = \sum_{i=1}^D |\psi_i^k\rangle\langle e_i^k|$ (as in Equation (6)) for $k = 1, \ldots, m$. Then, $(U_k)_{k=1}^m$ is uniquely decodable if and only if the collection*

$$\left\{ |\psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m\rangle : (v_1, \ldots, v_m) \in \{1, \ldots, D\}^m \right\}$$

forms an orthonormal set.

Proof: See Appendix 1.

Motivated from Theorem 2.9 we introduce the following definition.

Definition 2.10 *We say that a sequence of pure states $(|\psi_i\rangle)_{i=1}^D \subseteq (\mathbb{C}^2)^{\oplus}$ is **jointly orthonormal**, if and only if for every $m \in \mathbb{N}, \{|\psi_{v_1} \circ \cdots \circ \psi_{v_m}\rangle : (v_1, \ldots, v_m) \in \{1, \ldots, D\}^m\}$ is an orthonormal set.*

The next remark clarifies Definition 2.10.

Remark 2.11 *Notice that every jointly orthonormal sequence is orthonormal (which is obtained by taking $m = 1$). However, the converse is not always true. Take the states $\frac{1}{\sqrt{2}}(|0\rangle + |00\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |00\rangle)$ as a counter example.*

One can naturally construct a jointly orthonormal sequence by taking each $|\psi_i\rangle$ to be the bit strings produced from a classical uniquely decodable, or a classical prefix-free code. Another possible way to construct these sequences would be by taking the basis of a prefix-free Hilbert space as described in [28].

Consider a quantum stochastic source which contains an alphabet of N many pure states $(|s_i\rangle)_{i=1}^N$ that span a Hilbert space $\mathcal{H}$ of dimension d. Let $l, m \in \mathbb{N}$ where l denotes the block size, and m denotes the number of blocks. Fix a jointly orthonormal sequence $\{|\psi_i\rangle\}_{i=1}^{d^l}$ such that each $|\psi_i\rangle$ is a length state. Consider the tensor product state $|s_{n_1}\rangle |s_{n_2}\rangle \ldots |s_{n_{ml}}\rangle$ emitted from the quantum stochastic source, for some sequence $n_1, n_2, \ldots, n_{ml} \in \{1, \ldots, N\}$. Lastly, consider a sequence of quantum code $(U_k)_{k=1}^m$ where each isometry U_k for $k \in \{1, \ldots, m\}$

is constructed by taking an arbitrary orthonormal basis $\{e_i^k\}_{i=1}^{d^l}$ but the same jointly orthonormal sequence $\{|\psi_i\rangle\}_{i=1}^{d^l}$ such that $U_k = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^k|$, and is used to encode the kth block consisting of a tensor product of l many pure states. Hence, the tensor product state $|s_{n_1}\rangle |s_{n_2}\rangle \ldots |s_{n_{ml}}\rangle$ is encoded as the concatenation

$$U_1(|s_{n_1}\rangle \cdots |s_{n_l}\rangle) \circ U_2(|s_{n_{l+1}}\rangle \cdots |s_{n_{2l}}\rangle) \circ \cdots \circ U_m(|s_{(m-1)l+1}\rangle \cdots |s_{ml}\rangle).$$

By the forward direction of quantum Kraft-McMillan Inequality (Theorem 4.3), the sequence of quantum code $(U_k)_{k=1}^m$ is uniquely decodable, and therefore the concatenation above is well defined.

Since we are interested in minimizing the average codeword length, each isometry will be used to encode the average state of a block as opposed to a specific tensor product of l many pure states. Since the probability distribution for each block emitted by a quantum stochastic source is dependent on the entire sequence of states emitted up to the previous block, any k^{th}-block conditional ensemble state $\rho^{n_1,\ldots,n_{(k-1)l}}$ (as in Definition 2.4) depends on the distinct sequence of states $|s_{n_1} \cdots s_{n_{(k-1)l}}\rangle$ which is emitted up to that particular block, (without including it). Therefore, we collect the isometries of the form $U^{n_1,\ldots,n_{(k-1)l}} = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^{n_1,\ldots,n_{(k-1)l}}|$, one for each distinct k^{th}-block conditional ensemble state to form a special block code, as we specify next.

Definition 2.12 *Consider a quantum stochastic source $\mathcal{S}$ which contains an alphabet of N many pure states $(|s_i\rangle)_{i=1}^N$ that span a Hilbert space $\mathcal{H}$ of dimension d. Let $l, m \in \mathbb{N}$ where l denotes the block size, and m denotes the number of blocks. A **special block code** is a family of isometries*

$$\mathcal{U} = \left\{ U^{n_1,\ldots,n_{(k-1)l}} : 1 \leq k \leq m,\, n_1, \ldots, n_{(k-1)l} \in \{1, \ldots, N\} \right\},$$

such that every isometry used in the family $\mathcal{U}$ has a common jointly orthonormal sequence of length codewords. Thus more explicitly, there exists a jointly orthonormal sequence of length codewords $(|\psi_i\rangle)_{i=1}^{d^l} \subseteq (\mathbb{C}^2)^\oplus$, (see Definitions 2.10 and 2.5), and an orthonormal sequence $(|e_i^{n_1,\ldots,n_{(k-1)l}}\rangle)_{i=1}^{d^l}$ for $1 \leq k \leq m$ and $n_1, \ldots, n_{(k-1)l} \in \{1, \ldots, N\}$ such that

$$U^{n_1,\ldots,n_{(k-1)l}} = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^{n_1,\ldots,n_{(k-1)l}}|.$$

*As a special case, for every $k \in \{1, \ldots, m\}$, if $U^{n_1,\ldots,n_{(k-1)l}} = U_k = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^k|$ for all $n_1, \ldots, n_{(k-1)l} \in \{1, \ldots, N\}$ where $(|e_i^k\rangle)_{i=1}^{d^l}$ is an arbitrary orthonormal sequence fixed for the kth block and independent of the sequence of pure states $|s_{n_1} \cdots s_{n_{(k-1)l}}\rangle$ emitted in the first $k-1$ blocks, then we call such a special block code a **constrained special block code**. This notion will be used in Section 3.*

Remark 2.13 *Notice that in Definition 2.12, the superscript $n_1, \ldots, n_{(k-1)l}$ which appears in the isometry $U^{n_1,\ldots,n_{(k-1)l}}$ and the orthonormal basis $(|e_i^{n_1,\ldots,n_{(k-1)l}}\rangle)_{i=1}^{d^l}$ indicates that these quantities are allowed to depend on the sequence of pure states $|s_{n_1} \cdots s_{n_{(k-1)l}}\rangle$ emitted in the first $k-1$ blocks. In that case, such special block codes are termed as **adaptive** (or **dynamic**) special block codes. Our Theorem 2.18 states that in order to achieve the optimal lower bound*

of the average codeword length for encoding m many blocks of states with l many pure states per block emitted by the given quantum stochastic source, one needs to use such an adaptive special block code. In this optimal special block code, the orthonormal basis $(\left|e_i^{n_1,\ldots,n_{(k-1)l}}\right\rangle)_{i=1}^{d^l}$ for each isometry $U^{n_1,\ldots,n_{(k-1)l}}$ is chosen to coincide with the eigenvectors $(\left|\lambda_i^{n_1,\ldots,n_{(k-1)l}}\right\rangle)_{i=1}^{d^l}$ of the kth-block conditional ensemble state $\rho^{n_1,\ldots,n_{(k-1)l}}$.

Remark 2.14 *The constraint that all the isometries in a given special block code $\mathcal{U}$ have a common sequence of jointly orthonormal length codewords ensures that any sequence $(U_k)_{k=1}^m$ of quantum codes, which are chosen from a special block code, is uniquely decodable (by Theorem 2.9). Nevertheless, one could construct more general uniquely decodable quantum codes which do not necessarily meet this requirement, but we do not consider them in this paper. Such an example is given in the remark below.*

Remark 2.15 *Consider a sequence $(U_k)_{k=1}^m$ of m many isometries. Let O and E denote the set of odd and even numbers respectively from 1 through m. Let $U^x = \sum_{i=1}^D |\psi_i^x\rangle\langle e_i^x|$ and $U^y = \sum_{i=1}^D |\psi_i^y\rangle\langle e_i^y|$ such that $|\psi_i^x\rangle = |0\rangle^{\otimes i}$ and $|\psi_i^y\rangle = |1\rangle^{\otimes i}$ for all $x \in O$, $y \in E$, and $i \in \{1,\ldots,D\}$. It's easy to verify that the collection*

$$\left\{ \left|\psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m\right\rangle : (v_1,\ldots,v_m) \in \{1,\ldots,D\}^m \right\}$$

forms an orthonormal set. Therefore, the sequence $(U_k)_{k=1}^m$ of quantum codes is uniquely decodable by Theorem 2.9. However, the quantum codes U_k's do not have a common sequence of jointly orthonormal length codewords and hence the sequence $(U_k)_{k=1}^m$ is not a special block code.

Now we define the indeterminate length of any state on the Fock space.

Definition 2.16 *Define the **length observable** Λ acting on $(\mathbb{C}^2)^\oplus$ to be the unbounded linear operator given by*

$$\Lambda = \sum_{\ell=0}^\infty \ell \Pi_\ell, \tag{8}$$

where Π_ℓ is the orthogonal projection from $(\mathbb{C}^2)^\oplus$ onto the subspace $(\mathbb{C}^2)^{\otimes\ell}$.

*For any state ρ on the Fock space, define its **indeterminate length** to be the expectation of the length observable with respect to the state ρ, i.e.*

$$\mathrm{Tr}(\rho\Lambda),$$

where Tr denotes the trace on the Fock space.

This notion of indeterminate length of a state can be used to define the average codeword length of the special block code.

Definition 2.17 *Consider a quantum stochastic source $\mathcal{S}$ which contains an alphabet of N many pure states $(|s_i\rangle)_{i=1}^N$ which span a Hilbert space $\mathcal{H}$, and a stochastic process X with probability mass function p. Let $l,m \in \mathbb{N}$ where l denotes the block size, and m denotes the number of blocks. Consider a special block code*

$$\mathcal{U} = \left\{ U^{n_1,\ldots,n_{(k-1)l}} : 1 \le k \le m, n_1,\ldots,n_{(k-1)l} \in \{1,\ldots,N\} \right\},$$

as defined in Definition 2.12. Recall from Definition 2.3 that the ensemble state for the total of ml many emissions by the quantum stochastic source is equal to

$$\rho_{ml} = \sum_{n_1,\dots,n_{ml}=1}^{N} p(n_1,\dots,n_{ml}) \, |s_{n_1}\cdots s_{n_{ml}}\rangle\langle s_{n_1}\cdots s_{n_{ml}}|.$$

*We will denote by $L(\mathcal{U})$ the **average codeword length of the special block code $\mathcal{U}$**, which is defined to be equal to*

$$\sum_{n_1,\dots,n_{ml}=1}^{N} p(n_1,\dots,n_{ml})$$

$$\mathrm{Tr}\left(\left|U(s_{n_1}\cdots s_{n_l})\circ U^{n_1,\dots,n_l}(s_{n_{l+1}}\cdots s_{n_{2l}})\circ\cdots\circ U^{n_1,\dots,n_{(m-1)l}}(s_{n_{(m-1)l+1}}\cdots s_{n_{ml}})\right\rangle\right.$$

$$\left.\left\langle U(s_{n_1}\cdots s_{n_l})\circ U^{n_1,\dots,n_l}(s_{n_{l+1}}\cdots s_{n_{2l}})\circ\cdots\circ U^{n_1,\dots,n_{(m-1)l}}(s_{n_{(m-1)l+1}}\cdots s_{n_{ml}})\right|\Lambda\right).$$

*Let $ILS(\mathcal{S},m,l)$ denote the **infimum** of the set containing $L(\mathcal{U})$ for every **special** block code $\mathcal{U}$ that is used to encode ml many states emitted by $\mathcal{S}$ into m blocks each of size l. Similarly we define $ILC(\mathcal{S},m,l)$ for **constrained** special block codes.*

Finally, we are ready to state the main result of our article.

Theorem 2.18 *Consider a quantum stochastic source $\mathcal{S}$ consisting of an alphabet of N many pure states spanning a d-dimensional Hilbert space $\mathcal{H}$, and a stochastic process X having mass function p, as in Definition 2.1. Fix $m,l\in\mathbb{N}$. Then $ILS(\mathcal{S},m,l)$ can be computed as follows: For each $k=1,\dots,m$, and a sequence $n_1,\dots,n_{(k-1)l}$ of integers chosen from the set $\{1,\dots,N\}$, let $(\lambda_i^{n_1,\dots,n_{(k-1)l}})_{i=1}^{d^l}$ be the eigenvalues of the kth block conditional ensemble state $\rho^{n_1,\dots,n_{(k-1)l}}$, arranged in decreasing order, and $(|\lambda_i^{n_1,\dots,n_{(k-1)l}}\rangle)_{i=1}^{d^l}$ be the corresponding eigenvectors.*

Let

$$\mathfrak{L} = \left\{ (\ell_1,\dots,\ell_{d^l}) : \ell_i\in\mathbb{N}\cup\{0\} \text{ for all } i, \ \ell_1\le\ell_2\le\cdots\le\ell_{d^l}, \text{ and } \sum_{i=1}^{d^l} 2^{-\ell_i}\le 1 \right\}.$$

Define a function $\mathcal{F}_{\mathcal{S}}:\mathfrak{L}\to[0,\infty)$ by

$$\mathcal{F}_{\mathcal{S}}((\ell_i)_{i=1}^{d^l}) := \sum_{j=2}^{m}\left(\sum_{n_1,\dots,n_{(j-1)l}=1}^{N} p(n_1,\dots,n_{(j-1)l})\sum_{i=1}^{d^l}\lambda_i^{n_1,\dots,n_{(j-1)l}}\ell_i\right) + \sum_{i=1}^{d^l}\lambda_i\ell_i. \qquad (9)$$

Then,

$$ILS(\mathcal{S},m,l) = \min\{\mathcal{F}_{\mathcal{S}}((\ell_i)_{i=1}^{d^l}) : (\ell_i)_{i=1}^{d^l}\in\mathfrak{L}\}. \qquad (10)$$

Moreover, the infimum defining $ILS(\mathcal{S},m,l)$ is actually a minimum, i.e., there exists a special block code

$$\mathcal{V} = \left\{ V^{n_1,\dots,n_{(k-1)l}} : k\in\{1,\dots,m\}, \text{ and } n_1,\dots,n_{(k-1)l}\in\{1,\dots,N\} \right\},$$

which can be used to encode ml many states emitted by $\mathcal{S}$ into m blocks each of size l such that

$$\min\{\mathcal{F}_\mathcal{S}((\ell_i)_{i=1}^{d^l}) : (\ell_i)_{i=1}^{d^l} \in \mathfrak{L}\} = L(\mathcal{V}). \tag{11}$$

The minimizer $\mathcal{V}$ is given as follows: Assume that $\mathcal{F}_\mathcal{S}$ achieves its minimum on $\mathfrak{L}$ at the point $(\ell_i)_{i=1}^{d^l} \in \mathfrak{L}$. Since the sequence $(\ell_i)_{i=1}^{d^l}$ satisfies the classical Kraft-McMillan inequality, (which is the last condition in the definition of $\mathfrak{L}$), by the converse of classical Kraft-McMillan inequality there exists a sequence $(\omega_i)_{i=1}^{d^l}$ of classical bit strings with corresponding lengths $(\ell_i)_{i=1}^{d^l}$ such that the sequence $(\omega_i)_{i=1}^{d^l}$ forms the image of a classical uniquely decodable code. Then the corresponding sequence of qubit strings $(|\omega_i\rangle)_{i=1}^{d^l}$ forms an optimal jointly orthonormal sequence of length codewords. For each $k \in \{1, \ldots, m\}$, and string $n_1, \ldots, n_{(k-1)l} \in \{1, \ldots, N\}$, define

$$V^{n_1, \ldots, n_{(k-1)l}} : \mathcal{H}^{\otimes l} \to (\mathbb{C}^2)^{\oplus},$$

by

$$V^{n_1, \ldots, n_{(k-1)l}} = \sum_{i=1}^{d^l} |\omega_i\rangle\langle\lambda_i^{n_1, \ldots, n_{(k-1)l}}|.$$

Sketch of the proof: Here's the outline of the proof of the above theorem. Fix m and l as the number of blocks and the block size respectively. Let

$$\mathfrak{L} = \left\{ (\ell_1, \ldots, \ell_{d^l}) : \ell_i \in \mathbb{N} \cup \{0\} \text{ for all } i, \ \ell_1 \leq \ell_2 \leq \cdots \leq \ell_{d^l}, \text{ and } \sum_{i=1}^{d^l} 2^{-\ell_i} \leq 1 \right\}.$$

Fix a special block code $\mathcal{U}$ (Definition 2.12) such that all isometries in $\mathcal{U}$ have $(|\psi_i\rangle)_{i=1}^{d^l}$ as their common jointly orthonormal sequence of length codewords. Suppose that length$(|\psi_i\rangle) = \ell_i \in \mathbb{N}$ and without loss of generality, assume $\ell_r \leq \ell_s$ for $r \leq s$. By the forward direction of quantum Kraft-McMillan inequality (Theorem 4.3), we have that the set of non-negative integer lengths $\{\ell_i\}_{i=1}^{d^l}$ belongs to $\mathfrak{L}$.

For a specific sequence of pure states $|s_{n_1}, \cdots s_{n_{(k-1)l}}\rangle$ as the first $(k-1)l$ emissions, consider the corresponding k^{th} block conditional ensemble state $\rho^{n_1, \ldots, n_{(k-1)l}}$. Also, using the spectral decomposition, $\rho^{n_1, \ldots, n_{(k-1)l}} = \sum_{i=1}^{d^l} \lambda_i^{n_1, \ldots, n_{(k-1)l}} |\lambda_i^{n_1, \ldots, n_{(k-1)l}}\rangle\langle\lambda_i^{n_1, \ldots, n_{(k-1)l}}|$ and without loss of generality, assume that $\lambda_r^{n_1, \ldots, n_{(k-1)}} \geq \lambda_s^{n_1, \ldots, n_{(k-1)l}}$ for $r \leq s$. The codeword length of encoding the state $\rho^{n_1, \ldots, n_{(k-1)l}}$ by the quantum code $U^{n_1, \ldots, n_{(k-1)l}} = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^{n_1, \ldots, n_{(k-1)l}}| \in \mathcal{U}$, which has the length codewords $(|\psi_i\rangle)_{i=1}^{d^l}$ and an arbitrary orthonormal basis $(|e_i^{n_1, \ldots, n_{(k-1)l}}\rangle)_{i=1}^{d^l}$, is given by

$$\mathrm{Tr}\big(U^{n_1, \ldots, n_{(k-1)l}} \rho^{n_1, \ldots, n_{(k-1)l}} (U^{n_1, \ldots, n_{(k-1)l}})^\dagger \Lambda\big). \tag{12}$$

Then, we show by using the Birkhoff-von Neumann theorem and Lemma 4.2 that the codeword length given by the expression (12) is minimized for the given set of lengths $\{\ell_i\}_{i=1}^{d^l}$ when $U^{n_1, \ldots, n_{(k-1)l}} = \sum_{i=1}^{d^l} |\psi_i\rangle\langle\lambda_i^{n_1, \ldots, n_{(k-1)l}}|$. In this case, expression (12) can be re-written as

$$\mathrm{Tr}\big(U^{n_1, \ldots, n_{(k-1)l}} \rho^{n_1, \ldots, n_{(k-1)l}} (U^{n_1, \ldots, n_{(k-1)l}})^\dagger \Lambda\big) = \sum_{i=1}^{d^l} \lambda_i^{n_1, \ldots, n_{(k-1)l}} \ell_i. \tag{13}$$

Since the state $\rho^{n_1,\ldots,n_{(k-1)l}}$ occurs with the probability $p(n_1,\ldots,n_{(k-1)l})$, the minimum average codeword length for encoding the kth block ensemble state

$$\rho^k = \sum_{n_1,\ldots,n_{(k-1)l}=1}^{N} p(n_1,\ldots,n_{(k-1)l})\rho^{n_1,\ldots,n_{(k-1)l}} \quad \text{(Definition 2.4)}$$

is given by $\sum_{n_1,\ldots,n_{(k-1)l}} p(n_1,\ldots,n_{(k-1)l}) \sum_{i=1}^{d^l} \lambda_i^{n_1,\ldots,n_{(k-1)l}} \ell_i$ for the fixed set of lengths $\{\ell_i\}_{i=1}^{d^l}$. Recall that for $k=1$, we use the following two conventions:

$$\lambda_i^{n_1,\ldots,n_{(k-1)l}} = \lambda_i \quad \text{and} \quad p(n_1,\ldots,n_{(k-1)l}) = 1.$$

Summing the minimum average codeword lengths for each block gives the expression in Equation (9), which is the minimum average codeword length for m blocks. It is worth noticing that the minimum average codeword length for m blocks is additive of the minimum average codeword lengths for each block. Refer to the paragraph following Equation (B.21) in the Appendix B for the details. Finally, minimization of Equation (9) over all sets of lengths $(\ell_i)_{i=1}^{d^l} \in \mathfrak{L}$ gives Equation (10) which is the minimum average codeword length over the set of all special block codes for encoding m blocks. Once the minimizer set of lengths is found, one can then construct an optimal jointly orthonormal sequence of length codewords $(|\omega_i\rangle)_{i=1}^{d^l}$ and define a minimizer special block code $\mathcal{V}$ as stated in the theorem. See Appendix B for the full proof of the theorem. $\square$

Our Theorem 2.18 recovers the result of [5, Theorem 2] which was the case for $m=1$, (i.e. when the whole message is encoded as a single block).

3   A simplified compression for stochastic quantum sources and its application to stationary quantum sources

Notice from the statement of Theorem 2.18 that achieving the above theoretical lower bound requires that for every k^{th} block, Alice (the encoder) smartly chooses the optimal isometry $V^{n_1,\ldots,n_{(k-1)l}} = \sum_{i=1}^{d^l} |\omega_i\rangle\langle\lambda_i^{n_1,\ldots,n_{(k-1)l}}|$ to encode each kth block conditional ensemble state $\rho^{n_1,\ldots,n_{(k-1)l}}$, which requires an accurate estimation of the exact sequence of pure states $|s_{n_1,\ldots,n_{(k-1)l}}\rangle$ emitted in the first $(k-1)$ blocks. Since the pure states emitted by the source are not necessarily orthogonal, this task of state discrimination of non-orthogonal states can be difficult. So, we relax this requirement and envision a simplified compression technique where for a given k^{th} block, Alice uses a fixed isometry U_k to encode every k^{th} block conditional ensemble state $\rho^{n_1,\ldots,n_{(k-1)l}}$ regardless of the emissions in the first $(k-1)$ blocks. In other words, we would like to restrict our attention to the set of constrained special block codes (Definition 2.12) where $U^{n_1,\ldots,n_{(k-1)l}} = U_k = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^k|$ for all $n_1,\ldots,n_{(k-1)l} \in \{1,\ldots,N\}$. We then eventually show that when such a compression is applied to the pure states emitted from a stationary quantum source and one appropriately chooses the isometries U_k's, the optimal lower bound of the average codeword length per symbol equals the von Neumann entropy rate of the source for an asymptotically long block size. From here on, in order to distinguish between the two compressions, we will refer to the compression outlined in Theorem 2.18 as the **adaptive** compression while this new simplified compression that we discuss below as the **non-adaptive** compression. It is worth emphasizing that the only difference between the two

compressions is that the adaptive compression minimizes the average codeword length over the set of all special block codes, while the non-adaptive compression minimizes the same over the set of all constrained special block codes.

Theorem 3.1 *Consider a quantum stochastic source $\mathcal{S}$ consisting of an alphabet of N many pure states spanning a d-dimensional Hilbert space $\mathcal{H}$, and a stochastic process X having mass function p, as in Definition 2.1. Fix $m, l \in \mathbb{N}$. Then $ILC(\mathcal{S}, m, l)$ can be computed as follows: For each $k = 1, \ldots, m$, let $(\lambda_i^k)_{i=1}^{d^l}$ be the eigenvalues of the kth block ensemble state ρ^k, arranged in decreasing order, and $(|\lambda_i^k\rangle)_{i=1}^{d^l}$ be the corresponding eigenvectors.*

Let

$$\mathfrak{L} = \left\{ (\ell_1, \ldots, \ell_{d^l}) : \ell_i \in \mathbb{N} \cup \{0\} \text{ for all } i, \ \ell_1 \leq \ell_2 \leq \cdots \leq \ell_{d^l}, \ \text{and} \ \sum_{i=1}^{d^l} 2^{-\ell_i} \leq 1 \right\}.$$

Define a function $\mathcal{F}_{\mathcal{C}} : \mathfrak{L} \to [0, \infty)$ by

$$\mathcal{F}_{\mathcal{C}}((\ell_i)_{i=1}^{d^l}) := \sum_{k=1}^{m} \sum_{i=1}^{d^l} \lambda_i^k \ell_i. \tag{14}$$

Then,
$$ILC(\mathcal{S}, m, l) = \min\{\mathcal{F}_{\mathcal{C}}((\ell_i)_{i=1}^{d^l}) : (\ell_i)_{i=1}^{d^l} \in \mathfrak{L}\}. \tag{15}$$

Moreover, the infimum defining $ILC(\mathcal{S}, m, l)$ is actually a minimum, i.e., there exists a constrained special block code

$$\mathcal{V} = \left\{ V^k : k \in \{1, \ldots, m\} \right\},$$

which can be used to encode ml many states emitted by $\mathcal{S}$ into m blocks each of size l such that

$$\min\{\mathcal{F}_{\mathcal{C}}((\ell_i)_{i=1}^{d^l}) : (\ell_i)_{i=1}^{d^l} \in \mathfrak{L}\} = L(\mathcal{V}). \tag{16}$$

The minimizer $\mathcal{V}$ is given as follows: Assume that $\mathcal{F}_{\mathcal{C}}$ achieves its minimum on $\mathfrak{L}$ at the point $(\ell_i)_{i=1}^{d^l} \in \mathfrak{L}$. Since the sequence $(\ell_i)_{i=1}^{d^l}$ satisfies the classical Kraft-McMillan inequality, (which is the last condition in the definition of $\mathfrak{L}$), by the converse of classical Kraft-McMillan inequality there exists a sequence $(\omega_i)_{i=1}^{d^l}$ of classical bit strings with corresponding lengths $(\ell_i)_{i=1}^{d^l}$ such that the sequence $(\omega_i)_{i=1}^{d^l}$ forms the image of a classical unqiuely decodable code. Then the corresponding sequence of qubit strings $(|\omega_i\rangle)_{i=1}^{d^l}$ forms an optimal jointly orthonormal sequence of length codewords. For each $k \in \{1, \ldots, m\}$, define

$$V^k : \mathcal{H}^{\otimes l} \to (\mathbb{C}^2)^{\oplus},$$

by

$$V^k = \sum_{i=1}^{d^l} |\omega_i\rangle\langle\lambda_i^k|.$$

The following is a corollary of Theorem 3.1.

Corollary 3.2 *Consider a stationary quantum stochastic source $\mathcal{S}$ (Definition 2.2). We are interested in deriving the minimum average codeword length for m blocks each of block size l using constrained special block codes, i.e. $ILC(\mathcal{S}, m, l)$ as in Theorem 3.1. Let the spectral decomposition of the ensemble state ρ_l (or equivalently, ρ^1) (Definition 2.3) for l many emissions from the source to be:*

$$\rho_l = \sum_{i=1}^{d^l} \lambda_i \left| \lambda_i \right\rangle\!\left\langle \lambda_i \right|.$$

Consider the isometry $V = \sum_{i=1}^{d^l} |\omega_i\rangle\langle\lambda_i|$, where $(\omega_i)_{i=1}^{d^l}$ is the sequence of bit strings obtained from the Huffman code based on the probability distribution $(\lambda_i)_{i=1}^{d^l}$ such that $length(\omega_i) = \ell_i$. Then the minimizer constrained special block code $\mathcal{V}$ for encoding m blocks comprises of the identical quantum code given by $\mathcal{V} = V$ and the minimum average codeword length per symbol equals the von-Neumann entropy rate of the source for an asymptotically long block size, i.e. $\lim_{l\to\infty} \frac{ILC(\mathcal{S},1,l)}{l} = \lim_{l\to\infty} \frac{S(\rho_l)}{l} = \lim_{l\to\infty} \frac{1}{l} \sum_{i=1}^{d^l} \lambda_i \ell_i$.
Refer to Appendix C for the proofs of Theorem 3.1 and Corollary 3.2.

4   Conclusion and open questions

Bellomo, Bosyk, Holik and Zozor [5] defined the optimal quantum code for encoding the entire sequence of pure states emitted by a given quantum stochastic source into a single block and showed that the associated minimum average codeword length could be bounded from above and below in terms of the von-Neumann entropy of the source. The present work focuses on the encoding of the sequence of q many pure states from the quantum stochastic source $\mathcal{S}$ into $m \in \mathbb{N}$ blocks, each of block size $l \in \mathbb{N}$ such that $q = ml$ using special block codes, and we gave an optimal lower bound (denoted by $ILS(\mathcal{S}, m, l)$ in Theorem 2.18) of the average codeword length computed over all special block codes. Lastly, we gave a simplified non-adaptive compression technique based on constrained special block codes for quantum stochastic sources that emit pure states where the encoder can fix one isometry per block of states, and showed that if one uses such a compression technique for a stationary quantum source, then the optimal lower bound of the average codeword length per symbol over all constrained special block codes (denoted by $ILC(\mathcal{S}, m, l)$ in Corollary 3.2) equals the von-Neumann entropy rate of the source for an asymptotically long block size. Since we encode the pure states into indeterminate-length codewords which may not have a concrete physical interpretation, whether the derived optimal lower bounds are practically achievable require further investigation. Nevertheless, these theoretical lower bounds can serve as a benchmark for measuring the efficiency of lossless quantum block compression algorithms. The following interesting questions are still open.

For a fixed length message consisting of $q \in \mathbb{N}$ many pure states emitted by a given quantum stochastic source S, what are the optimal values of m and l such that $ml = q$ and $ILS(\mathcal{S}, m, l)$ over special (or $ILC(\mathcal{S}, m, l)$ over constrained special) block codes is minimized? In other words, what is the optimal number of blocks and the optimal block size for encoding a fixed length message of pure states emitted by a quantum stochastic source using special (or constrained special) block codes?

It would be interesting to relate $ILS(\mathcal{S}, m, l)$ over special block codes with an entropic quantity of a state related to the quantum source. Perhaps one needs an alternate entropic quantity (other than von Neumann entropy) in order to tightly bound $ILS(\mathcal{S}, m, l)$, (see [1, 4]).

Notice that special block codes are not the most general uniquely decodable quantum codes. So, it would be interesting to derive a similar optimal lower bound of the average codeword length over all uniquely decodable quantum codes used for block encoding the sequence of pure states emitted by a quantum stochastic source.

One could also pursue answering these questions for a quantum stochastic source which emits mixed states instead of pure states, which is yet another interesting direction of research.

Acknowledgment

References

1. R. Ahlswede and N. Cai. On lossless quantum data compression with a classical helper. *IEEE Trans. Inf. Theory*, 50(6):1208–1219, June 2004. `doi:10.1109/TIT.2004.828071`.
2. R. Alicki and M. Fannes. *Quantum dynamical systems*. Oxford university press, 2001.
3. G. Androulakis and D. Wright. Optimality in quantum data compression using dynamical entropy. *Phys. Rev. A*, 100(3):032301, Sept. 2019. `doi:10.1103/PhysRevA.100.032301`.
4. A. Anshu, A. Garg, A. W. Harrow, and P. Yao. *Lower Bound on Expected Communication Cost of Quantum Huffman Coding*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Sept. 2016. URL: `https://dspace.mit.edu/handle/1721.1/109878`.
5. G. Bellomo, G. M. Bosyk, F. Holik, and S. Zozor. Lossless quantum data compression with exponential penalization: an operational interpretation of the quantum Rényi entropy. *Sci. Rep.*, 7(14765):1–10, Nov. 2017. `doi:10.1038/s41598-017-13350-y`.
6. I. Bjelakovic, T. Krüger, R. Siegmund-Schultze, and A. Szkola. Chained typical subspaces-a quantum version of breiman's theorem. *arXiv preprint quant-ph/0301177*, 2003.
7. I. Bjelaković, T. Krüger, R. Siegmund-Schultze, and A. Szkoła. The shannon-mcmillan theorem for ergodic quantum lattice systems. *Inventiones mathematicae*, 155:203–222, 2004.
8. I. Bjelaković and A. Szkola. The data compression theorem for ergodic quantum information sources. *Quantum Information Processing*, 4:49–63, 2005.
9. K. Bostroem and T. Felbinger. Lossless quantum data compression and variable-length coding. *Phys. Rev. A*, 65(3):032313, Feb. 2002. `doi:10.1103/PhysRevA.65.032313`.
10. S. L. Braunstein, C. A. Fuchs, D. Gottesman, and H.-K. Lo. A quantum analog of Huffman coding. *IEEE Trans. Inf. Theory*, 46(4):1644–1649, July 2000. `doi:10.1109/18.850709`.
11. H.-H. Chou and J. Cheng. New lower bounds on the average base length of lossless quantum data compression. 2006. URL: `https://api.semanticscholar.org/CorpusID:118374660`.
12. I. L. Chuang and D. S. Modha. Reversible arithmetic coding for quantum data compression. *IEEE Trans. Inf. Theory*, 46(3):1104–1116, May 2000. `doi:10.1109/18.841192`.
13. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Apr. 2005. `doi:10.1002/047174882X`.
14. R. M. Fano. *The transmission of information*, volume 65. Massachusetts Institute of Technology, Research Laboratory of Electronics . . . , 1949.

15. M. Hajdušek and R. Van Meter. Quantum Communications. *arXiv*, Nov. 2023. `arXiv:2311.02367`, `doi:10.48550/arXiv.2311.02367`.

16. M. Hayashi. *A Group Theoretic Approach to Quantum Information.* Springer International Publishing, Cham, Switzerland. URL: `https://link.springer.com/book/10.1007/978-3-319-45241-8`.

17. M. Hayashi. Universal Approximation of Multi-copy States and Universal Quantum Lossless Data Compression. *Commun. Math. Phys.*, 293(1):171–183, Jan. 2010. `doi:10.1007/s00220-009-0909-y`.

18. M. Hayashi and K. Matsumoto. Quantum universal variable-length source coding. *Phys. Rev. A*, 66(2):022311, Aug. 2002. `doi:10.1103/PhysRevA.66.022311`.

19. M. Hayashi and K. Matsumoto. Simple construction of quantum universal variable-length source coding. *Quantum Inf. Comput.*, Dec. 2002. URL: `https://doi.org/10.26421/QIC2.s-2`.

20. D. A. Huffman. A Method for the Construction of Minimum-Redundancy Codes. *Proc. IRE*, 40(9):1098–1101, Sept. 1952. `doi:10.1109/JRPROC.1952.273898`.

21. R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki. Universal Quantum Information Compression. *Phys. Rev. Lett.*, 81(8):1714–1717, Aug. 1998. `doi:10.1103/PhysRevLett.81.1714`.

22. A. Kaltchenko and E.-H. Yang. Universal compression of classically correlated stationary quantum sources. In *Proceedings Volume 5105, Quantum Information and Computation*, volume 5105, pages 104–115. SPIE, Aug. 2003. `doi:10.1117/12.485663`.

23. A. Kaltchenko and E.-H. Yang. Universal compression of ergodic quantum sources. *Quantum Inf. Comput.*, 3(4):359–375, July 2003. `doi:10.5555/2011528.2011533`.

24. S. Khatri and M. M. Wilde. Principles of Quantum Communication Theory: A Modern Approach. *arXiv*, Nov. 2020. `arXiv:2011.04672`, `doi:10.48550/arXiv.2011.04672`.

25. C. King and A. Leśniewski. Quantum sources and a quantum coding theorem. *Journal of Mathematical Physics*, 39(1):88–101, 1998.

26. G. G. Langdon. An introduction to arithmetic coding. *IBM Journal of Research and Development*, 28(2):135–149, 1984. `doi:10.1147/rd.282.0135`.

27. J. Langford. Generic quantum block compression. *Phys. Rev. A*, 65(5):052312, May 2002. `doi:10.1103/PhysRevA.65.052312`.

28. M. Müller and C. Rogers. Quantum bit strings and prefix-free Hilbert spaces. In *Information Theory and Statistical Learning*, 2008. URL: `https://api.semanticscholar.org/CorpusID:17405115`.

29. M. Müller, C. Rogers, and R. Nagarajan. Lossless quantum prefix compression for communication channels that are always open. *Phys. Rev. A*, 79:012302, Jan 2009. URL: `https://link.aps.org/doi/10.1103/PhysRevA.79.012302`, `doi:10.1103/PhysRevA.79.012302`.

30. D. Petz and M. Mosonyi. Stationary quantum source coding. *Journal of Mathematical Physics*, 42(10):4857–4864, 2001.

31. B. Schumacher. Presentation at Sante Fe Institute workshop on Complexity, Entropy and the Physics of Information. 1994.

32. B. Schumacher. Quantum coding. *Phys. Rev. A*, 51(4):2738–2747, Apr. 1995. `doi:10.1103/PhysRevA.51.2738`.

33. B. Schumacher and M. D. Westmoreland. Indeterminate-length quantum coding. *Phys. Rev. A*, 64(4):042304, Sept. 2001. `doi:10.1103/PhysRevA.64.042304`.

34. C. E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

35. K. Y. Szeto. Data Compression for Quantum Code. *arXiv*, July 1996. `arXiv:quant-ph/9607010`, `doi:10.48550/arXiv.quant-ph/9607010`.

Appendix A A characterization of uniquely decodable quantum codes

Proof of Theorem 2.9: The "only if" direction:

Suppose the sequence of quantum codes $(U_k)_{k=1}^m$ is uniquely decodable, and each U_k is given

by

$$U_k = \sum_{i=1}^{D} \left| \psi_i^k \middle\rangle\middle\langle e_i^k \right|.$$

Then by Definition 2.8, the linear map $U_1 \circ \cdots \circ U_m : \mathcal{K}^{\otimes m} \to (\mathbb{C}^2)^{\oplus}$ is an isometry. Therefore, it follows that

$$\begin{aligned}
&\left\langle (U_1 \circ \cdots \circ U_m)(e_{v_1}^1 \otimes \cdots \otimes e_{v_m}^m) \middle| (U_1 \circ \cdots \circ U_m)(e_{v_1'}^1 \otimes \cdots \otimes e_{v_m'}^m) \right\rangle \\
&= \left\langle e_{v_1}^1 \otimes \cdots \otimes e_{v_m}^m \middle| e_{v_1'}^1 \otimes \cdots \otimes e_{v_m'}^m \right\rangle \\
&= \delta_{v_1,v_1'} \cdots \delta_{v_m,v_m'}.
\end{aligned} \tag{A.1}$$

On the other hand, by Definition 2.6, $(U_1 \circ \cdots \circ U_m) \left| e_{v_1}^1 \otimes \cdots \otimes e_{v_m}^m \right\rangle = \psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m$. Therefore, by Equation (A.1) we obtain that

$$\begin{aligned}
&\left\langle \psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m \middle| \psi_{v_1'}^1 \circ \cdots \circ \psi_{v_m'}^m \right\rangle \\
&= \left\langle (U_1 \circ \cdots \circ U_m)(e_{v_1}^1 \otimes \cdots \otimes e_{v_m}^m) \middle| (U_1 \circ \cdots \circ U_m)(e_{v_1'}^1 \otimes \cdots \otimes e_{v_m'}^m) \right\rangle \\
&= \delta_{v_1,v_1'} \cdots \delta_{v_m,v_m'},
\end{aligned}$$

i.e. the collection $\left\{ \left| \psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m \right\rangle : (v_1, \ldots, v_m) \in \{1, \ldots, D\}^m \right\}$ is an orthonormal set.
The "if" direction:
Suppose that the collection $\left\{ \left| \psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m \right\rangle : (v_1, \ldots, v_m) \in \{1, \ldots, D\}^m \right\}$ is an orthonormal set. Consider any two sequences of vectors $(|\phi_k\rangle)_{k=1}^m$, and $(|\Phi_k\rangle)_{k=1}^m$ such that $|\phi_k\rangle$ and $|\Phi_k\rangle$ belong in the range of U_k for $k = 1, \ldots, m$. Hence, both $|\phi_k\rangle$ and $|\Phi_k\rangle$ belong in the linear span of $\{|\psi_v^k\rangle\}_{v=1}^D$ for every $k \in 1, \ldots, m$. Thus, there exist scalars $(a_{v_k}^k)_{v_k=1}^D$ and $(A_{v_k}^k)_{v_k=1}^D$ such that

$$|\phi_k\rangle = \sum_{v_k=1}^{D} a_{v_k}^k \left| \psi_{v_k}^k \right\rangle, \quad \text{and} \quad |\Phi_k\rangle = \sum_{v_k=1}^{D} A_{v_k}^k \left| \psi_{v_k}^k \right\rangle.$$

Therefore,

$$\begin{aligned}
|\phi_1 \circ \cdots \circ \phi_m\rangle &= \left(\sum_{v_1=1}^{D} a_{v_1}^1 \left| \psi_{v_1}^1 \right\rangle \right) \circ \cdots \circ \left(\sum_{v_m=1}^{D} a_{v_m}^m \left| \psi_{v_m}^m \right\rangle \right) \\
&= \sum_{v_1,\ldots,v_m=1}^{D} a_{v_1}^1 \cdots a_{v_m}^m \left| \psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m \right\rangle.
\end{aligned} \tag{A.2}$$

and similarly,

$$|\Phi_1 \circ \cdots \circ \Phi_m\rangle = \sum_{v_1',\ldots,v_m'=1}^{D} A_{v_1'}^1 \cdots A_{v_m'}^m \left| \psi_{v_1'}^1 \circ \cdots \circ \psi_{v_m'}^m \right\rangle.$$

Thus,

$$\langle \phi_1 \circ \cdots \circ \phi_m | \Phi_1 \circ \cdots \circ \Phi_m \rangle$$

$$= \sum_{\substack{v_1,\ldots,v_m=1 \\ v_1',\ldots,v_m'=1}}^{D} \overline{a_{v_1}^1} \cdots \overline{a_{v_m}^m} A_{v_1'}^1 \cdots A_{v_m'}^m \left\langle \psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m \middle| \psi_{v_1'}^1 \circ \cdots \circ \psi_{v_m'}^m \right\rangle. \tag{A.3}$$

Since the states $\left| \psi_{v_1}^1 \circ \cdots \circ \psi_{v_m}^m \right\rangle$ form an orthonormal set, we have

$$\langle \phi_1 \circ \cdots \circ \phi_m | \Phi_1 \circ \cdots \circ \Phi_m \rangle = \sum_{\substack{v_1,\ldots,v_m=1 \\ v_1',\ldots,v_m'=1}}^{D} \overline{a_{v_1}^1} \cdots \overline{a_{v_m}^m} A_{v_1'}^1 \cdots A_{v_m'}^m \, \delta_{(v_1,\ldots,v_m),(v_1',\ldots,v_m')}$$

$$= \sum_{v_1,v_1'=1}^{D} \overline{a_{v_1}^1} A_{v_1'}^1 \delta_{v_1,v_1'} \cdots \sum_{v_m,v_m'=1}^{D} \overline{a_{v_m}^1} A_{v_m'}^1 \delta_{v_m,v_m'}$$

$$= \sum_{v_1,v_1'=1}^{D} \overline{a_{v_1}^1} A_{v_1'}^1 \left\langle \psi_{v_1} \middle| \psi_{v_1'} \right\rangle \cdots \sum_{v_m,v_m'=1}^{D} \overline{a_{v_m}^1} A_{v_m'}^1 \left\langle \psi_{v_m} \middle| \psi_{v_m'} \right\rangle$$

$$= \langle \phi_1 | \Phi_1 \rangle \cdots \langle \phi_m | \Phi_m \rangle. \tag{A.4}$$

Since $|\phi_k\rangle$ and $|\Phi_k\rangle$ belong in the range of U_k for $k = 1, \ldots, m$, there exist vectors $(|\chi_k\rangle)_{k=1}^m$, $(|X_k\rangle)_{k=1}^m$ in $\mathcal{K}$ such that $\phi_k = U_k(\chi_k)$ and $\Phi_k = U_k(X_k)$ for all $k = 1, \ldots, m$. Hence,

$$\langle (U_1 \circ \cdots \circ U_m)(\chi_1 \otimes \cdots \otimes \chi_m) | (U_1 \circ \cdots \circ U_m)(X_1 \otimes \cdots \otimes X_m) \rangle$$

$$= \langle U_1\chi_1 \circ \cdots \circ U_m\chi_m | U_1 X_1 \circ \cdots \circ U_m X_m \rangle$$

$$= \langle U_1\chi_1 | U_1 X_1 \rangle \cdots \langle U_m\chi_m | U_m X_m \rangle$$

$$= \langle \chi_1 | X_1 \rangle \cdots \langle \chi_m | X_m \rangle, \tag{A.5}$$

(where the first equality is valid because of the definition of $U_1 \circ \cdots \circ U_m$, the second equality is valid because of Equation (A.4), and the third equality is valid since $U_1, \ldots, U_m$ are isometries). Thus, $U_1 \circ \cdots \circ U_m$ maps an orthonormal basis of $\mathcal{K}^{\otimes m}$ to an orthonormal set of $(\mathbb{C}^2)^{\oplus}$, and thus it is an isometry. In particular, by taking $\phi_k = \Phi_k$, (and hence $\chi_k = X_k$) for $k = 1, \ldots, m$, Equation (A.5) gives that $U_1 \circ \cdots \circ U_m$ satisfies

$$\| (U_1 \circ \cdots \circ U_m)(\chi_1 \otimes \cdots \otimes \chi_m) \| = \| |\chi_1\rangle \| \cdots \| |\chi_m\rangle \| = \| \chi_1 \otimes \cdots \otimes \chi_m \|.$$

$\square$

Notice that Equation (A.4) gives the following important consequence:

Remark 4.1 *Let $(|\psi_i\rangle)_{i=1}^D$ be a jointly orthonormal sequence in the Fock space, and m be a fixed natural number. Then, on the linear span of $(|\psi_i\rangle)_{i=1}^D$ concatenations of m-many vectors are well defined.*

Appendix B The proof of Theorem 2.17

First, we need to introduce a well known lemma which is required in the proof of our main result.

Lemma 4.2 *For $z \in \mathbb{N}$ consider a non-increasing sequence of positive real numbers $Q_1 \geq Q_2 \geq \cdots \geq Q_z \geq 0$. Further, consider another arbitrary sequence of positive real numbers*

$l_1, l_2, \ldots, l_z$ and its non-decreasing enumeration $l'_1 \leq l'_2 \leq \cdots \leq l'_z$. Then,

$$\sum_{i=1}^{z} Q_i l'_i \leq \sum_{i=1}^{z} Q_i l_i.$$

Proof: Consider the two terms $(Q_1 L + q l'_1)$, (associated to Q_1 and to l'_1), in the sum $\sum_{i=1}^{z} Q_i l_i$. Since Q_1 is the largest term in the sequence $\{Q_i\}_{i=1}^{z}$, we have $Q_1 \geq q$. Also, since l'_1 is the smallest term in the sequence $\{l_i\}_{i=1}^{z}$, we have $l'_1 \leq L$. Thus,

$$Q_1(L - l'_1) \geq q(L - l'_1) \Rightarrow Q_1 L + q l'_1 \geq Q_1 l'_1 + qL.$$

Therefore, by replacing the terms $(Q_1 L + q l'_1)$ in the sum $\sum_{i=1}^{z} Q_i l_i$ with the smaller quantity $(Q_1 l'_1 + qL)$ yields a smaller sum $\sum_{i=1}^{z} Q_i l_i - (Q_1 L + q l'_1) + (Q_1 l'_1 + qL)$. By continuing similarly for the next largest number Q_2 in the newly created sum $\sum_{i=1}^{z} Q_i l_i - (Q_1 L + q l'_1) + (Q_1 l'_1 + qL)$ yields an even smaller sum. By continuing this process for all Q_i's one by one in descending order, we obtain the sum $\sum_{i=1}^{z} Q_i l'_i$ which is smaller than the original sum $\sum_{i=1}^{z} Q_i l_i$. $\square$

Another preliminary material that we need for our main result is the following theorem which is a straightforward generalization of the quantum Kraft-McMillan inequality (and its converse) considered in [3].

Theorem 4.3 *Consider a Hilbert space $\mathcal{H}$ of dimension D and a sequence of jointly orthonormal length codewords $(|\psi_i\rangle)_{i=1}^{D} \subseteq (\mathbb{C}^2)^{\oplus}$ such that each $|\psi_i\rangle \in (\mathbb{C}^2)^{\otimes \ell_i}$ for some $\ell_i \in \mathbb{N}$. For $k \in \{1, \ldots, m\}$, if each $U_k = \sum_{i=1}^{D} |\psi_i\rangle\langle e_i^k|$, then the finite sequence of quantum codes $(U_k)_{k=1}^{m}$ is uniquely decodable and the following inequality is satisfied*

$$\sum_{i=1}^{D} 2^{-\ell_i} \leq 1. \tag{B.1}$$

Conversely, assume that $(U_k)_{k=1}^{m}$ is a sequence of quantum codes and each $U_k = \sum_{i=1}^{D} |\psi_i\rangle\langle e_i^k|$, where $(|\psi_i\rangle)_{i=1}^{D} \subseteq (\mathbb{C}^2)^{\oplus}$ is the common sequence of length codewords such that each $|\psi_i\rangle \in (\mathbb{C}^2)^{\otimes \ell_i}$ for some $\ell_i \in \mathbb{N}$ and the lengths $(\ell_i)_{i=1}^{D}$ satisfy the Inequality (B.1). Then, there exist a jointly orthonormal sequence of length codewords $(|\psi'_i\rangle)_{i=1}^{D} \subseteq (\mathbb{C}^2)^{\oplus}$ such that each $|\psi'_i\rangle \in (\mathbb{C}^2)^{\otimes \ell_i}$, and a uniquely decodable quantum code $(U'_k)_{k=1}^{m}$ such that each $U'_k = \sum_{i=1}^{D} |\psi'_i\rangle\langle e_i^k|$.

Proof: Fix a jointly orthonormal sequence $\{|\psi_i\rangle\}_{i=1}^{D}$ such that each $|\psi_i\rangle$ is a length state i.e. $\text{Tr}(|\psi_i\rangle\langle\psi_i| \Lambda) = \ell_i \in \mathbb{N}$ or equivalently, $|\psi_i\rangle \in (\mathbb{C}^2)^{\otimes \ell_i}$. Consider a sequence of quantum codes $(U_k)_{k=1}^{D}$ where each isometry U_k for $k \in \{1, \ldots, m\}$ is constructed by taking an arbitrary orthonormal basis $\{e_i^k\}_{i=1}^{D}$ but the same jointly orthonormal sequence $\{|\psi_i\rangle\}_{i=1}^{D}$ such that $U_k = \sum_{i=1}^{D} |\psi_i\rangle\langle e_i^k|$. By Definition 2.10, it follows that the collection $\{|\psi_{i_1} \circ \psi_{i_2} \circ \cdots \circ \psi_{i_m}\rangle\}$ for $\{i_1, \ldots, i_m\} \in \{1, \ldots, D\}$ is an orthonormal set. Then, by Theorem 2.9, it follows that the sequence $(U_k)_{k=1}^{D}$ is a uniquely decodable quantum code.

For each $n, N \in \mathbb{N}$, let

$$C_n^N = \{|\Psi^n\rangle \in (\mathbb{C}^2)^{\otimes N} : |\Psi^n\rangle = |\psi_{i_1} \circ \psi_{i_2} \circ \cdots \circ \psi_{i_n}\rangle \text{ for some } i_1, \ldots, i_n \in \{1, \ldots, D\}\}$$

be the collection of states formed from concatenation of n-many codewords and having length N, and let

$$d_\ell = \#\{i \in \{1, \ldots, D\} : \psi_i \in (\mathbb{C}^2)^{\otimes \ell}\}$$

be the number of $|\psi_i\rangle$'s in $(\mathbb{C}^2)^{\otimes \ell}$ for each $\ell \in \mathbb{N}$. Since the sequence $(|\psi_i\rangle)_{i=1}^D$ is jointly orthonormal, the elements of C_n^N are pairwise orthogonal, hence linearly independent, and since they belong in $(\mathbb{C}^2)^{\otimes N}$ whose dimension is equal to 2^N, we have

$$\#C_n^N = \sum_{\ell_{i_1}+\cdots+\ell_{i_n}=N} d_{\ell_{i_1}} d_{\ell_{i_2}} \ldots d_{\ell_{i_n}} \leq 2^N.$$

Thus,

$$2^{-N} \sum_{\ell_{i_1}+\cdots+\ell_{i_n}=N} d_{\ell_{i_1}} d_{\ell_{i_2}} \ldots d_{\ell_{i_n}} = \sum_{\ell_{i_1}+\cdots+\ell_{i_n}=N} (2^{-\ell_{i_1}} d_{\ell_{i_1}})(2^{-\ell_{i_2}} d_{\ell_{i_2}}) \cdots (2^{-\ell_{i_n}} d_{\ell_{i_n}}) \leq 1$$

Set $\ell_{max} = \max\limits_{1 \leq i \leq D}\{\ell_i\}$ so that $N \leq n\ell_{max}$. Summing the above inequality over N we obtain

$$\sum_{\ell_{i_1},\ell_{i_2},\ldots,\ell_{i_n}=1}^{\ell_{max}} = (2^{\ell_{i_1}} d_{\ell_{i_1}})(2^{\ell_{i_2}} d_{\ell_{i_2}}) \cdots (2^{\ell_{i_n}} d_{\ell_{i_n}}) = \left(\sum_{\ell=1}^{\ell_{max}} 2^{-\ell} d_\ell\right)^n \leq n\ell_{max}.$$

Notice that the left hand side of this inequality is exponential while the right hand side is linear. This implies that the left hand side is bounded above by 1. Hence, we must have that

$$\sum_{\ell=1}^{\ell_{max}} 2^{-\ell} d_\ell = \sum_{i=1}^D 2^{-\ell_i} \leq 1,$$

and hence the classical Kraft-McMillan Inequality is also valid.

Conversely, suppose that $(U_k)_{k=1}^m$ is a sequence of quantum codes having a common sequence of length codewords $(|\psi_i\rangle)_{i=1}^D$ and let $\ell_i \in \mathbb{N} \cup \{0\}$ be the length of $|\psi_i\rangle$ for all $i = 1, \ldots, D$. Assume that Inequality (B.1) is satisfied. Hence the classical Kraft-McMillan inequality is valid. So, by the converse of the classical Kraft-McMillan Theorem, there exists a set of bit strings $\{\psi_i'\}_{i=1}^D$ such that there are exactly d_ℓ many bit strings with length ℓ for each $\ell \in \mathbb{N}$, and the sequence $(\psi_i')_{i=1}^D$ forms the image of a uniquely decodable (and in fact, instantaneous) classical code. Then, the sequence $(|\psi_i'\rangle)_{i=1}^D$ of the corresponding qubit strings satisfies $|\psi_i'\rangle \in (\mathbb{C}^2)^{\otimes \ell_i}$ and every sequence of isometries $(U_k')_{k=1}^m$ such that each $U_k' = \sum_{i=1}^D |\psi_i'\rangle\langle e_i^k|$ (also called classical-quantum code as in [3]) for $k = 1, \ldots, m$, forms a uniquely decodable sequence of quantum codes with a common sequence of jointly orthonormal length codewords. $\square$

The last ingredient that we need for the proof of our main result is the definition of concatenations of certain rank-1 operators on the Fock space. Notice that Remark 4.1 ensures that the concatenation of rank-1 operators as introduced in the next definition is well defined.

Definition 4.4 *Let $(|\psi_i\rangle)_{i=1}^D$ be a jointly orthonormal sequence of length codewords in the Fock space. If $m \in \mathbb{N}$ and $I = (i_1, i_2, \ldots, i_m) \in \{1, \ldots, D\}^m$, then let*

$$\left|\underset{i \in I}{\circ} \psi_i\right\rangle := |\psi_{i_1} \circ \psi_{i_2} \circ \cdots \circ \psi_{i_m}\rangle.$$

Let $m \in \mathbb{N}$, and $|x\rangle, |y\rangle$ be linear combinations of concatenations of m-many $|\psi_i\rangle$'s. Similarly, let $n \in \mathbb{N}$ and $|z\rangle, |w\rangle$ be linear combinations of concatenations of n-many $|\psi_i\rangle$'s. Thus,

assume that

$$|x\rangle = \sum_i \mu_i \left|\underset{i' \in I_{x,i}}{\circ} \psi_{i'}\right\rangle, \quad |y\rangle = \sum_j \mu_j \left|\underset{j' \in I_{y,j}}{\circ} \psi_{j'}\right\rangle,$$

$$|z\rangle = \sum_k \mu_k \left|\underset{k' \in I_{z,k}}{\circ} \psi_{k'}\right\rangle \ and \ |w\rangle = \sum_\ell \mu_\ell \left|\underset{\ell' \in I_{w,\ell}}{\circ} \psi_{\ell'}\right\rangle, \tag{B.2}$$

*where $\mu_i, \mu_j, \mu_k, \mu_\ell \in \mathbb{C}$, $I_{x,i}, I_{y,j} \in \{1, \ldots, D\}^m$, and $I_{z,k}, I_{w,\ell} \in \{1, \ldots, D\}^n$ for all i, j, k, ℓ. Then, we define the **concatenation of the rank-1 operators** $|x\rangle\langle y|$ and $|z\rangle\langle w|$ by*

$$|x\rangle\langle y| \circ |z\rangle\langle w| := |x \circ z\rangle\langle y \circ w|.$$

As a further clarification of the above definition, notice that under its setting, we have that if $I = (i_1, \ldots, i_m) \in \{1, \ldots, D\}^m$ and $J = (j_1, \ldots, j_n) \in \{1, \ldots, D\}^n$ for some $m, n \in \mathbb{N}$, then let $I \circ J \in \{1, \ldots, D\}^{m+n}$ be defined by $I \circ J = (i_1, \ldots, i_m, j_1, \ldots, j_n)$. Hence, notice that

$$|x \circ z\rangle = \sum_{i,k} \mu_i \mu_k \left|\underset{m \in I_{x,i} \circ I_{z,k}}{\circ} \psi_m\right\rangle \ and \ |y \circ w\rangle = \sum_{j,\ell} \mu_j \mu_\ell \left|\underset{n \in I_{y,j} \circ I_{w,\ell}}{\circ} \psi_n\right\rangle.$$

Thus,

$$|x \circ z\rangle\langle y \circ w| = \sum_{i,k,j,\ell} \mu_i \mu_k \overline{\mu_j \mu_\ell} \left|\underset{m \in I_{x,i} \circ I_{z,k}}{\circ} \psi_m\right\rangle\left\langle\underset{n \in I_{y,j} \circ I_{w,\ell}}{\circ} \psi_n\right|.$$

Now we are ready to present the proof of our main result.

Proof of Theorem 2.18: Let $\mathcal{U} = \{U^{n_1, \ldots, n_{(k-1)l}}\}$ be any arbitrary special block code, which is used to encode m many blocks of pure states emitted by a given quantum stochastic source $\mathcal{S}$ (Definition 2.1), where each block size is l. Let $(|\psi_i\rangle)_{i=1}^{d^l} \subseteq (\mathbb{C}^2)^\oplus$ be a sequence of jointly orthonormal length codewords which belongs in the range of every member of $\mathcal{U}$ such that each $|\psi_i\rangle$ has length ℓ_i for some $\ell_i \in \mathbb{N}$ i.e. $\text{Tr}(|\psi_i\rangle\langle\psi_i| \Lambda) = \ell_i$, or equivalently $|\psi_i\rangle \in (\mathbb{C}^2)^{\otimes \ell_i}$. Let the orthonormal sequences be $(|e_i\rangle)_{i=1}^{d^l}$ for $k = 1$, and $(|e_i^{n_1, \ldots, n_{(k-1)l}}\rangle)_{i=1}^{d^l}$ for $2 \leq k \leq m$ and $n_1, \ldots, n_{(k-1)l} \in \{1, \ldots, N\}$ such that

$$U = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i| \quad and \quad U^{n_1, \ldots, n_{(k-1)l}} = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^{n_1, \ldots, n_{(k-1)l}}|. \tag{B.3}$$

Since we want to minimize the average codeword length, we encode the ensemble state ρ_{ml} (Definition 2.3) for ml emissions from the source $\mathcal{S}$ using the quantum codes of the special block code $\mathcal{U}$. Recall that the average codeword length of $\mathcal{U}$ is defined in Definition 2.17 to be equal to

$$\sum_{n_1, \ldots, n_{ml}=1}^N p(n_1, \ldots, n_{ml})$$

$$\text{Tr}\left(\left|U(s_{n_1} \cdots s_{n_l}) \circ U^{n_1, \ldots, n_l}(s_{n_{l+1}} \cdots s_{n_{2l}}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-1)l}}(s_{n_{(m-1)l+1}} \cdots s_{n_{ml}})\right\rangle\right.$$

$$\left.\left\langle U(s_{n_1} \cdots s_{n_l}) \circ U^{n_1, \ldots, n_l}(s_{n_{l+1}} \cdots s_{n_{2l}}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-1)l}}(s_{n_{(m-1)l+1}} \cdots s_{n_{ml}})\right| \Lambda\right)$$

We will obtain the minimum average codeword length over all special block codes in two steps: in the first step, we optimize the orthonormal basis $\{|e_i^{n_1,\ldots,n_{(k-1)l}}\rangle\}_{i=1}^{d^l}$ for each isometry $U^{n_1,\ldots,n_{(k-1)l}} \in \mathcal{U}$, and in the second step, we optimize the common jointly orthonormal sequence of length codewords $(|\psi_i\rangle)_{i=1}^{d^l}$.

From now on, in the interest of minimizing the horizontal space used by equations, we will abbreviate certain scalars, pure states, mixed states and sums as in the following table for $r \leq t, i \in \mathbb{N}$.

	The expression	is abbreviated by
Scalars:	$p(n_r, \ldots, n_t)$	$p[r,t]$
	$\lambda_i^{n_r,\ldots,n_t}$	$\lambda_i^{[r,t]}$
Pure states:	$\lvert s_{n_r} \cdots s_{n_t} \rangle$	$\lvert s_{[r,t]} \rangle$
	$\lvert \lambda_i^{n_r,\ldots,n_t} \rangle$	$\lvert \lambda_i^{[r,t]} \rangle$
	$\lvert e_u^{n_1,\ldots,n_t} \rangle$	$\lvert e_u^t \rangle$
	$\lvert e_{u_0} e_{u_1}^{n_1,\ldots,n_l} e_{u_2}^{n_1,\ldots,n_{2l}} \cdots e_{u_t}^{n_1,\ldots,n_{tl}} \rangle$	$\lvert e_{[u_0,u_t]}^{[0,tl]} \rangle$
	$\lvert \psi_{u_r} \cdots \psi_{u_t} \rangle$	$\lvert \psi_{[u_r,u_t]} \rangle$
Mixed states:	$\rho^{n_r,\ldots,n_t}$	$\rho^{[r,t]}$
Sums:	$\sum_{n_r,\ldots,n_t=1}^{N}$	$\sum_{[n_r,n_t]}^{N}$
	$\sum_{u_r,\ldots,u_t=1}^{d^l}$	$\sum_{[u_r,u_t]}^{d^l}$
	$\sum_{u_r,\ldots,u_t,u_r',\ldots,u_t'=1}^{d^l}$	$\sum_{[u_r,u_t,u_r',u_t']}^{d^l}$

Table B.1   The list of all abbreviations that are used in this proof.

Also we accept the following convention.

Convention 4.5 *For $t < 0$, the state $\lvert e_{u_0} e_{u_1}^{n_1,\ldots,n_l} e_{u_2}^{n_1,\ldots,n_{2l}} \cdots e_{u_t}^{n_1,\ldots,n_{tl}} \rangle$ and its abbreviation $\lvert e_{[u_0,u_t]}^{[0,tl]} \rangle$ do not exist and these terms should be ignored where they appear. For $t = 0$, the state $\lvert e_{u_0} e_{u_1}^{n_1,\ldots,n_l} e_{u_2}^{n_1,\ldots,n_{2l}} \cdots e_{u_t}^{n_1,\ldots,n_{tl}} \rangle$ and its abbreviation $\lvert e_{[u_0,u_t]}^{[0,tl]} \rangle$ stand for the state $\lvert e_{u_0} \rangle$.*

Using these abbreviations, the average codeword length of $\mathcal{U}$ can be written as:

$$\sum_{[n_1,n_{ml}]}^{N} p[1,ml] \operatorname{Tr} \Big(\lvert U(s_{[1,l]}) \circ U^{n_1,\ldots,n_l}(s_{[l+1,2l]}) \circ \cdots \circ U^{n_1,\ldots,n_{(m-1)l}}(s_{[(m-1)l+1,ml]}) \rangle$$
$$\langle U(s_{[1,l]}) \circ U^{n_1,\ldots,n_l}(s_{[l+1,2l]}) \circ \cdots \circ U^{n_1,\ldots,n_{(m-1)l}}(s_{[(m-1)l+1,ml]}) \rvert \Lambda \Big). \tag{B.4}$$

The expression inside the trace in the above formula of the average codeword length of $\mathcal{U}$ can be written as

$$\lvert U(s_{[1,l]}) \circ \cdots \circ U^{n_1,\ldots,n_{(m-2)l}} s_{[(m-2)l+1,(m-1)l]} \rangle \circ \lvert U^{n_1,\ldots,n_{(m-1)l}}(s_{[(m-1)l+1,ml]}) \rangle$$
$$\langle U(s_{[1,l]}) \circ \cdots \circ U^{n_1,\ldots,n_{(m-2)l}} s_{[(m-2)l+1,(m-1)l]} \rvert \circ \langle U^{n_1,\ldots,n_{(m-1)l}}(s_{[(m-1)l+1,ml]}) \rvert \Lambda.$$

We rewrite the last expression as

$$\lvert U(s_{[1,l]}) \circ \cdots \circ U^{n_1,\ldots,n_{(m-2)l}}(s_{[(m-2)l+1,(m-1)l]}) \rangle$$
$$\langle U(s_{[1,l]}) \circ \cdots \circ U^{n_1,\ldots,n_{(m-2)l}}(s_{[(m-2)l+1,(m-1)l]}) \rvert \circ$$
$$\lvert U^{n_1,\ldots,n_{(m-1)l}}(s_{[(m-1)l+1,ml]}) \rangle \langle U^{n_1,\ldots,n_{(m-1)l}}(s_{[(m-1)l+1,ml]}) \rvert \Lambda, \tag{B.5}$$

where the concatenation of the two rank-1 operators is defined via Definition 4.4. By (B.5), Expression (B.4) becomes:

$$\sum_{[n_1, n_{ml}]}^{N} p[1, ml]$$

$$\mathrm{Tr}\left(\left| U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right\rangle \right.$$

$$\left\langle U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right|$$

$$\left. \circ \left| U^{n_1, \ldots, n_{(m-1)l}}\left(s_{[(m-1)l+1, ml]}\right) \right\rangle\!\!\left\langle U^{n_1, \ldots, n_{(m-1)l}}\left(s_{[(m-1)l+1, ml]}\right) \right| \Lambda \right)$$

$$= \sum_{[n_1, n_{(m-1)l}]}^{N} p[1, (m-1)l]$$

$$\mathrm{Tr}\left(\left| U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right\rangle \right.$$

$$\left\langle U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right| \circ$$

$$\sum_{[n_{(m-1)l+1}, n_{ml}]}^{N} \frac{p[1, ml]}{p[1, (m-1)l]} \left| U^{n_1, \ldots, n_{(m-1)l}}\left(s_{[(m-1)l+1, ml]}\right) \right\rangle$$

$$\left. \left\langle U^{n_1, \ldots, n_{(m-1)l}}\left(s_{[(m-1)l+1, ml]}\right) \right| \Lambda \right)$$

$$= \sum_{[n_1, n_{ml}]}^{N} p[1, (m-1)l]$$

$$\mathrm{Tr}\left(\left| U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right\rangle \right.$$

$$\left\langle U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right|$$

$$\circ U^{n_1, \ldots, n_{(m-1)l}} \left(\sum_{[n_{(m-1)l+1}, n_{ml}]}^{N} \frac{p[1, ml]}{p[1, (m-1)l]} \left| s_{[(m-1)l+1, ml]} \right\rangle\!\!\left\langle s_{[(m-1)l+1, ml]} \right| \right)$$

$$\left. \left(U^{n_1, \ldots, n_{(m-1)l}} \right)^{\dagger} \Lambda \right)$$

$$= \sum_{[n_1, n_{ml}]}^{N} p[1, (m-1)l]$$

$$\mathrm{Tr}\left(\left| U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right\rangle \right.$$

$$\left\langle U(s_{[1,l]}) \circ \cdots \circ U^{n_1, \ldots, n_{(m-2)l}}\left(s_{[(m-2)l+1, (m-1)l]}\right) \right|$$

$$\left. \circ U^{n_1, \ldots, n_{(m-1)l}} \left(\rho^{[1, (m-1)l]} \right) \left(U^{n_1, \ldots, n_{(m-1)l}} \right)^{\dagger} \Lambda \right), \tag{B.6}$$

where $\rho^{[1,(m-1)l]}$ is the mth block conditional ensemble state as in Definition 2.4. If we use Equation (B.3), the abbreviations listed in Table B.1, and the Convention 4.5, this expression becomes equal to the following:

$$\sum_{[n_1, n_{(m-1)l}]}^{N} p[1,(m-1)l]$$

$$\text{Tr}\left(\sum_{[u_0,u_{m-2},u_0',u_{m-2}']}^{d^l} \left\langle e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \Big| s_{[1,(m-1)l]} \right\rangle \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0',u_{m-2}']}^{[0,(m-2)l]} \right\rangle \right. \tag{B.7}$$

$$\left. \Big| \psi_{[u_0,u_{m-2}]} \Big\rangle\!\Big\langle \psi_{[u_0',u_{m-2}']} \Big| \circ U^{n_1,\dots,n_{(m-1)l}} \rho^{[1,(m-1)l]} (U^{n_1,\dots,n_{(m-1)l}})^\dagger \Lambda \right).$$

Consider the spectral decomposition of the quantum state $\rho^{[1,(m-1)l]}$ given as

$$\rho^{[1,(m-1)l]} = \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \left| \lambda_i^{[1,(m-1)l]} \right\rangle\!\!\left\langle \lambda_i^{[1,(m-1)l]} \right| \tag{B.8}$$

where each $\lambda_i^{[1,(m-1)l]}$ is an eigenvalue corresponding to the eigenvector $\left| \lambda_i^{[1,(m-1)l]} \right\rangle$, (see Table B.1). Without loss of generality, we assume that $\lambda_1^{[1,(m-1)l]} \geq \lambda_2^{[1,(m-1)l]} \cdots \geq \lambda_{d^l}^{[1,(m-1)l]} \geq 0$. Thus, using Equation (B.8) we obtain that the average codeword length of $\mathcal{U}$ (i.e. Expression (B.7))is equal to

$$\sum_{[n_1, n_{(m-1)l}]}^{N} p[1,(m-1)l]$$

$$\text{Tr}\left(\sum_{[u_0,u_{m-2},u_0',u_{m-2}']}^{d^l} \left\langle e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \Big| s_{[1,(m-1)l]} \right\rangle \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0',u_{m-2}']}^{[0,(m-2)l]} \right\rangle \Big| \psi_{[u_0,u_{m-2}]} \Big\rangle\!\Big\langle \psi_{[u_0',u_{m-2}']} \Big| \right.$$

$$\left. \circ \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \sum_{[u_{m-1},u_{m-1}']}^{d^l} \left\langle e_{u_{m-1}}^{m-1} \Big| \lambda_i^{[1,(m-1)l]} \right\rangle \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u_{m-1}'}^{m-1} \right\rangle \Big| \psi_{u_{m-1}} \Big\rangle\!\Big\langle \psi_{u_{m-1}'} \Big| \Lambda \right)$$

$$= \sum_{[n_1, n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-2},u_0',u_{m-2}']}^{d^l} \left\langle e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \Big| s_{[1,(m-1)l]} \right\rangle \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0',u_{m-2}']}^{[0,(m-2)l]} \right\rangle$$

$$\sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \sum_{[u_{m-1},u_{m-1}']}^{d^l} \left\langle e_{u_{m-1}}^{m-1} \Big| \lambda_i^{[1,(m-1)l]} \right\rangle \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u_{m-1}'}^{m-1} \right\rangle$$

$$\text{Tr}\left(\Big| \psi_{[u_0,u_{m-1}]} \Big\rangle\!\Big\langle \psi_{[u_0',u_{m-1}']} \Big| \Lambda \right)$$

$$= \sum_{[n_1, n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-2},u_0',u_{m-2}']}^{d^l} \left\langle e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \Big| s_{[1,(m-1)l]} \right\rangle \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0',u_{m-2}']}^{[0,(m-2)l]} \right\rangle$$

$$\sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \sum_{[u_{m-1},u'_{m-1}]}^{d^l} \left\langle e_{u_{m-1}}^{m-1} \Big| \lambda_i^{[1,(m-1)l]} \right\rangle \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u'_{m-1}}^{m-1} \right\rangle$$

$$\sum_{\ell=0}^{\infty} \ell \left\langle \psi_{[u'_0,u'_{m-1}]} \Big| \Pi_\ell \Big| \psi_{[u_0,u_{m-1}]} \right\rangle, \tag{B.9}$$

where in the last equality we used the cyclic property of trace and Equation (8).

Then, we can evaluate part of the Equation (B.9) as follows:

$$\sum_{\ell=0}^{\infty} \ell \left\langle \psi_{[u'_0,u'_{m-1}]} \Big| \Pi_\ell \Big| \psi_{[u_0,u_{m-1}]} \right\rangle = \left(\sum_{j=0}^{m-1} \ell_{u_j} \right) \left(\prod_{k=0}^{m-1} \delta_{u_k,u'_k} \right)$$

since $(|\psi_i\rangle)_{i=1}^{d^l}$ is a jointly orthonormal sequence of length codewords (Definition 2.10). So, Equation (B.9) simplifies to the following expression:

$$\sum_{[n_1,n_{(m-1)l]}}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-2}]}^{d^l} \left(\left| \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \right\rangle \right|^2 \right.$$

$$\sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \sum_{u_{m-1}=1}^{d^l} \left| \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u_{m-1}}^{m-1} \right\rangle \right|^2 \sum_{j=0}^{m-1} \ell_{u_j}. \tag{B.10}$$

Notice that since $\left\{ \left| \lambda_i^{[1,(m-1)l]} \right\rangle \right\}_{i=1}^{d^l}$ and $\left\{ \left| e_{u_{m-1}}^{m-1} \right\rangle \right\}_{u_{m-1}=1}^{d^l}$ are both orthonormal bases of $\mathcal{H}^{\otimes l}$, we have

$$\sum_{i=1}^{d^l} \left| \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u_{m-1}}^{m-1} \right\rangle \right|^2 = \sum_{u_{m-1}=1}^{d^l} \left| \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u_{m-1}}^{m-1} \right\rangle \right|^2 = 1.$$

Therefore, for every $n_1, \ldots, n_{(m-1)l} \in \{1, \ldots, N\}$, the terms

$$\left\{ \left| \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u_{m-1}}^{m-1} \right\rangle \right|^2 : i, u_{m-1} = 1, \ldots, d^l \right\} \tag{B.11}$$

in Expression (B.10) above, form a $d^l \times d^l$ bistochastic matrix $B^{n_1,\ldots,n_{(m-1)l}}$. By the Birkoff-von Neumann Theorem, every $d^l \times d^l$ bistochastic matrix can be written as a convex combination of at most $d^l!$ many permuation matrices. So,

$$B^{n_1,\ldots,n_{(m-1)l}} = \sum_{h=1}^{d^l!} t_h^{n_1,\ldots,n_{(m-1)l}} P_h^{n_1,\ldots,n_{(m-1)l}}, \tag{B.12}$$

where $\{ t_h^{n_1,\ldots,n_{(m-1)l}} \}_{h=1}^{d^l!}$ are convex coefficients and $\{ P_h^{n_1,\ldots,n_{(m-1)l}} \}_{h=1}^{d^l!}$ are permutation matrices. For every fixed $n_1, \ldots, n_{(m-1)l} \in \{1, \ldots, N\}$, define the function $f^{n_1,\ldots,n_{(m-1)l}}$ on the $d^l \times d^l$ matrices $B = (B(i,u_{m-1}))_{i,u_{m-1}=1}^{d^l}$ and taking scalar values, as follows:

$$f^{n_1,\ldots,n_{(m-1)l}}(B) =$$

$$\sum_{[u_0,u_{m-2}]}^{d^l} \left| \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \right\rangle \right|^2 \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \sum_{u_{m-1}=1}^{d^l} B(i,u_{m-1}) \sum_{j=0}^{m-1} \ell_{u_j}. \tag{B.13}$$

Notice that the function $f^{n_1,\ldots,n_{(m-1)l}}$ is affine. Thus, for fixed $n_1,\ldots,n_{(m-1)l} \in \{1,\ldots,N\}$, there exists $h \in \{1,\ldots,d^l!\}$, (which depends on $n_1,\ldots,n_{(m-1)l}$), such that the minimizer $B^{n_1,\ldots,n_{(m-1)l}}$ of $f^{n_1,\ldots,n_{(m-1)l}}$ is equal to $P_h^{n_1,\ldots,n_{(m-1)l}}$, i.e.

$$f^{n_1,\ldots,n_{(m-1)l}}\left(B^{n_1,\ldots,n_{(m-1)l}}\right)$$
$$\geq f^{n_1,\ldots,n_{(m-1)l}}\left(P_h^{n_1,\ldots,n_{(m-1)l}}\right)$$
$$= \sum_{[u_0,u_{m-2}]}^{d^l} \left| \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \right\rangle \right|^2 \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \sum_{u_{m-1}=1}^{d^l} P_h^{n_1,\ldots,n_{(m-1)l}}(i,u_{m-1}) \sum_{j=0}^{m-1} \ell_{u_j},$$

where $P_h^{n_1,\ldots,n_{(m-1)l}}(i,u_{m-1})$ denotes the (i,u_{m-1}) entry of $P_h^{n_1,\ldots,n_{(m-1)l}}$. Since $P_h^{n_1,\ldots,n_{(m-1)l}}$ is a $d^l \times d^l$ permutation matrix, for every $1 \leq i \leq d^l$ there exists a unique $g_i \in \{1,\ldots,d^l\}$ such that $P_h^{n_1,\ldots,n_{(m-1)l}}(i,g_i) = 1$, and all other entries of the ith row are equal to 0. Hence,

$$P_h^{n_1,\ldots,n_{(m-1)l}}(i,u_{m-1}) = \delta_{u_{m-1},g_i}. \tag{B.14}$$

Therefore,

$$\sum_{[u_0,u_{m-2}]}^{d^l} \left| \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \right\rangle \right|^2 \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \sum_{u_{m-1}=1}^{d^l} P_h^{n_1,\ldots,n_{(m-1)l}}(i,u_{m-1}) \sum_{j=0}^{m-1} \ell_{u_j}$$
$$= \sum_{[u_0,u_{m-2}]}^{d^l} \left| \left\langle s_{[1,(m-1)l]} \Big| e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \right\rangle \right|^2 \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \left(\sum_{j=0}^{m-2} \ell_{u_j} + \ell_{g_i} \right). \tag{B.15}$$

For the length sequence $(\ell_i)_{i=1}^{d^l}$ of the jointly orthonormal sequence of length codewords $(|\psi\rangle)_{i=1}^{d^l}$, we can assume without loss of generality that

$$\ell_1 \leq \ell_2 \leq \cdots \leq \ell_{d^l}. \tag{B.16}$$

Since we have also assumed that $\lambda_1^{[1,(m-1)l]} \geq \lambda_2^{[1,(m-1)l]} \geq \cdots \geq \lambda_{d^l}^{[1,(m-1)l]} \geq 0$, we obtain by Lemma 4.2, that Equation (B.15) is minimized when

$$g_i = i \text{ for } 1 \leq i \leq d^l. \tag{B.17}$$

Equations (B.14) and (B.17) imply that the minimizer permutation matrix $P^{n_1,\ldots,n_{(m-1)l}}$ is equal to the identity $d^l \times d^l$ matrix. Since the entries of the input matrices of the function $f^{n_1,\ldots,n_{(m-1)l}}$ are given by (B.11), we obtain that

$$\left| \left\langle \lambda_i^{[1,(m-1)l]} \Big| e_{u_{m-1}}^{m-1} \right\rangle \right| = \delta_{i,u_{m-1}} \quad \text{for } i,u_{m-1} = 1,\ldots,d^l.$$

Since both vectors $\left| \lambda_i^{[1,(m-1)l]} \right\rangle$ and $\left| e_{u_{m-1}}^{m-1} \right\rangle$ are normalized, one can infer that

$$\left| e_i^{m-1} \right\rangle = \alpha_i \left| \lambda_i^{[1,(m-1)l]} \right\rangle \text{ for some } \alpha_i \in \mathbb{C} \text{ with } |\alpha_i| = 1 \text{ for } 1 \leq i \leq d^l.$$

By ignoring the phase factors α_i's, (since they do not affect the average codeword length of $\mathcal{U}$, as it can be seen by the Expression (B.10)), we conclude that the average codeword length of $\mathcal{U}$ will be minimized if we use the following isometry to encode the last (m^{th}) block:

$$U^{n_1,\ldots,n_{(m-1)l}} = \sum_{i=1}^{d^l} \left|\psi_i\right\rangle\left\langle\lambda_i^{[1,(m-1)l]}\right|. \tag{B.18}$$

Equation (B.17) also gives that the minimum of Expression (B.10) is equal to

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-2}]}^{d^l} \left|\left\langle s_{[1,(m-1)l]}\Big|e_{[u_0,u_{m-2}]}^{[0,(m-2)l]}\right\rangle\right|^2 \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \left(\sum_{j=0}^{m-2} \ell_{u_j} + \ell_i\right). \tag{B.19}$$

One can proceed with finding the optimal orthonormal basis for $(m-1)^{th}$ block in a similar manner, as we explain now. Equation (B.19) can be split into two sums as follows:

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-2}]}^{d^l} \left|\left\langle s_{[1,(m-1)l]}\Big|e_{[u_0,u_{m-2}]}^{[0,(m-2)l]}\right\rangle\right|^2 \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i \;+$$

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-2}]}^{d^l} \left|\left\langle s_{[1,(m-1)l]}\Big|e_{[u_0,u_{m-2}]}^{[0,(m-2)l]}\right\rangle\right|^2 \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \left(\sum_{j=0}^{m-2} \ell_{u_j}\right). \tag{B.20}$$

Note that

$$\sum_{[u_0,u_{m-2}]}^{d^l} \left|\left\langle s_{[1,(m-1)l]}\Big|e_{[u_0,u_{m-2}]}^{[0,(m-2)]}\right\rangle\right|^2 = \left\|\left|s_{[1,(m-1)l]}\right\rangle\right\|^2 = 1 \quad \text{and} \quad \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} = 1,$$

hence Expression (B.20) simplifies as follows:

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i \;+$$

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-2}]}^{d^l} \left|\left\langle s_{[1,(m-1)l]}\Big|e_{[u_0,u_{m-2}]}^{[0,(m-2)l]}\right\rangle\right|^2 \left(\sum_{j=0}^{m-2} \ell_{u_j}\right). \tag{B.21}$$

It is worth mentioning at this point that splitting Equation (B.19) followed by some simplifications gave Equation (B.21) which has two parts: a) the first line i.e. $\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i$ represents the minimum average codeword length for the mth (last) block for the fixed set of lengths $(\ell_i)_{i=1}^{d^l}$ and b) the second line represents the sum of the average codeword lengths for the first $(m-1)$ blocks which still needs to be minimized. As we will see below, this second line can be similarly split further into the minimum average

codeword length for the $(m-1)^{th}$ block for the given set of lengths $(\ell_i)_{i=1}^{d^l}$ and the sum of the average codeword lengths for the first $(m-2)$ blocks subject to further minimization. Therefore, the minimum average codeword length for m blocks is the sum of the minimum average codeword lengths for each block.

Notice that the second line of Expression (B.21) can be simplified as follows:

$$\sum_{[u_0,u_{m-2}]}^{d^l} \left| \left\langle s_{[1,(m-1)l]} \middle| e_{[u_0,u_{m-2}]}^{[0,(m-2)l]} \right\rangle \right|^2 \left(\sum_{j=0}^{m-2} \ell_{u_j} \right)$$

$$= \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \sum_{u_{m-2}=1}^{d^l} \left| \left\langle s_{[(m-2)l+1,(m-1)l]} \middle| e_{u_{m-2}}^{(m-2)l} \right\rangle \right|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +$$

$$\sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{u_{m-2}=1}^{d^l} \left| \left\langle s_{[(m-2)l+1,(m-1)l]} \middle| e_{u_{m-2}}^{(m-2)l} \right\rangle \right|^2 \ell_{u_{m-2}} \right)$$

$$= \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left\| \left| s_{[(m-2)l+1,(m-1)l]} \right\rangle \right\|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +$$

$$\sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{u_{m-2}=1}^{d^l} \ell_{u_{m-2}} \left\langle e_{u_{m-2}}^{(m-2)l} \middle| s_{[(m-2)l+1,(m-1)l]} \right\rangle \right.$$

$$\left. \left\langle s_{[(m-2)l+1,(m-1)l]} \middle| e_{u_{m-2}}^{(m-2)l} \right\rangle \right)$$

$$= \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +$$

$$\sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{u_{m-2}=1}^{d^l} \ell_{u_{m-2}} \left\langle e_{u_{m-2}}^{(m-2)l} \middle| s_{[(m-2)l+1,(m-1)l]} \right\rangle \right.$$

$$\left. \left\langle s_{[(m-2)l+1,(m-1)l]} \middle| e_{u_{m-2}}^{(m-2)l} \right\rangle \right).$$

By substituting the last quantity in Expression (B.21) we obtain

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i +$$

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \Big| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +$$

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \Big| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2$$

$$\left(\sum_{u_{m-2}=1}^{d^l} \ell_{u_{m-2}} \left\langle e_{u_{m-2}}^{(m-2)l} \Big| s_{[(m-2)l+1,(m-1)l]} \right\rangle \left\langle s_{[(m-2)l+1,(m-1)l]} \Big| e_{u_{m-2}}^{(m-2)l} \right\rangle \right).$$

(B.22)

Note that since the bras $\left\langle s_{n_{(m-2)l+1}} \cdots s_{n_{(m-1)l}} \right|$ do not appear any more in the second line of Expression (B.22), the probability distributions $p[1,(m-1)l]$ in the second line of the Expression (B.22) when summed over the indices $n_{(m-2)l+1}, \ldots, n_{(m-1)l}$ simplify to $p[1,(m-2)l]$. Thus, Expression (B.22) simplifies to

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i +$$

$$\sum_{[n_1,n_{(m-2)l}]}^{N} p[1,(m-2)l] \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \Big| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +$$

$$\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \Big| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2$$

$$\left(\sum_{u_{m-2}=1}^{d^l} \ell_{u_{m-2}} \left\langle e_{u_{m-2}}^{(m-2)l} \Big| s_{[(m-2)l+1,(m-1)l]} \right\rangle \left\langle s_{[(m-2)l+1,(m-1)l]} \Big| e_{u_{m-2}}^{(m-2)l} \right\rangle \right)$$

$$= \sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i +$$

$$\sum_{[n_1,n_{(m-2)l}]}^{N} p[1,(m-2)l] \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \Big| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +$$

$$\sum_{[n_1,n_{(m-2)l}]}^{N} p[1,(m-2)l] \sum_{[u_0,u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \Big| e_{[u_0,u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{u_{m-2}=1}^{d^l} \ell_{u_{m-2}} \left\langle e_{u_{m-2}}^{(m-2)l} \right| \right.$$

$$\left. \left(\sum_{[n_{(m-2)l+1},n_{(m-1)l}]}^{N} \frac{p[1,(m-1)l]}{p[1,(m-2)l]} \Big| s_{[(m-2)l+1,(m-1)l]} \Big\rangle\!\Big\langle s_{[(m-2)l+1,(m-1)l]} \Big| \right) \Big| e_{u_{m-2}}^{(m-2)l} \right\rangle \right)$$

$$
= \sum_{[n_1, n_{(m-1)l}]}^{N} p[1, (m-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i +
$$

$$
\sum_{[n_1, n_{(m-2)l}]}^{N} p[1, (m-2)l] \sum_{[u_0, u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0, u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +
$$

$$
\sum_{[n_1, n_{(m-2)l}]}^{N} p[1, (m-2)l] \sum_{[u_0, u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0, u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2
$$

$$
\left(\sum_{u_{m-2}=1}^{d^l} \ell_{u_{m-2}} \left\langle e_{u_{m-2}}^{(m-2)l} \middle| \rho^{[1,(m-2)l]} \middle| e_{u_{m-2}}^{(m-2)l} \right\rangle \right). \tag{B.23}
$$

As in Equation (B.8), consider the spectral decomposition of $\rho^{[1,(m-2)l]}$,

$$
\rho^{[1,(m-2)l]} = \sum_{i=1}^{d^l} \lambda_i^{[1,(m-2)l]} \left| \lambda_i^{[1,(m-2)l]} \right\rangle \left\langle \lambda_i^{[1,(m-2)l]} \right|. \tag{B.24}
$$

By substituting Equation (B.24) in Equation (B.23) we obtain:

$$
\sum_{[n_1, n_{(m-1)l}]}^{N} p[1, (m-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(m-1)l]} \ell_i +
$$

$$
\sum_{[n_1, n_{(m-2)l}]}^{N} p[1, (m-2)l] \sum_{[u_0, u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0, u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2 \left(\sum_{j=0}^{m-3} \ell_{u_j} \right) +
$$

$$
\sum_{[n_1, n_{(m-2)l}]}^{N} p[1, (m-2)l] \sum_{[u_0, u_{m-3}]}^{d^l} \left| \left\langle s_{[1,(m-2)l]} \middle| e_{[u_0, u_{m-3}]}^{[0,(m-3)l]} \right\rangle \right|^2
$$

$$
\left(\sum_{i=1}^{d^l} \lambda_i^{[1,(m-2)l]} \sum_{u_{m-2}=1}^{d^l} \ell_{u_{m-2}} \left| \left\langle \lambda_i^{[1,(m-2)l]} \middle| e_{u_{m-2}}^{(m-2)l} \right\rangle \right|^2 \right). \tag{B.25}
$$

As before (see Expression (B.10)), for every $n_1, \ldots, n_{(m-2)l}$, each $\left| \left\langle \lambda_i^{[1,(m-2)l]} \middle| e_{u_{m-2}}^{(m-2)} \right\rangle \right|^2$ in Equation (B.25) above forms the (i, u_{m-2}) element of a $d^l \times d^l$ bistochastic matrix $B^{n_1, \ldots, n_{(m-2)l}}$ for $1 \le i, u_{m-2} \le d^l$. By the Birkoff-von Neumann Theorem and Lemma 4.2, Equation (B.25) attains its minimum at the $d^l \times d^l$ identity permutation matrix. Thus,

$$
\left| e_i^{m-2} \right\rangle = \alpha_i \left| \lambda_i^{[1,(m-2)l]} \right\rangle \text{ for some } \alpha_i \in \mathbb{C} \text{ with } |\alpha_i| = 1 \text{ for } 1 \le i \le d^l,
$$

and hence, by ignoring the phase factors α_i's as before, we conclude that the average codeword length of $\mathcal{U}$ will be minimized if we use the following isometry to encode the last $(m-1)^{\text{th}}$ block:

$$
U^{n_1, \ldots, n_{(m-2)l}} = \sum_{i=1}^{d^l} \left| \psi_i \right\rangle \left\langle \lambda_i^{[1,(m-2)l]} \right|. \tag{B.26}
$$

In this case, Equation (B.25) simplifies to

$$
\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l]\sum_{i=1}^{d^l}\lambda_i^{[1,(m-1)l]}\ell_i +
$$

$$
\sum_{[n_1,n_{(m-2)l}]}^{N} p[1,(m-2)l]\sum_{[u_0,u_{m-3}]}^{d^l}\left|\left\langle s_{[1,(m-2)l]}\Big|e_{[u_0,u_{m-3}]}^{[0,(m-3)l]}\right\rangle\right|^2\left(\sum_{j=0}^{m-3}\ell_{u_j}\right) +
$$

$$
\sum_{[n_1,n_{(m-2)l}]}^{N} p[1,(m-2)l]\sum_{[u_0,u_{m-3}]}^{d^l}\left|\left\langle s_{[1,(m-2)l]}\Big|e_{[u_0,u_{m-3}]}^{[0,(m-3)l]}\right\rangle\right|^2\left(\sum_{i=1}^{d^l}\lambda_i^{[1,(m-2)l]}\ell_i\right)
$$

$$
= \sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l]\sum_{i=1}^{d^l}\lambda_i^{[1,(m-1)l]}\ell_i +
$$

$$
\sum_{[n_1,n_{(m-2)l}]}^{N} p[1,(m-2)l]\sum_{[u_0,u_{m-3}]}^{d^l}\left|\left\langle s_{[1,(m-2)l]}\Big|e_{[u_0,u_{m-3}]}^{[0,(m-3)l]}\right\rangle\right|^2\left(\sum_{i=1}^{d^l}\lambda_i^{[1,(m-2)l]}\ell_i + \sum_{j=0}^{m-3}\ell_{u_j}\right).
$$

$$(B.27)$$

By comparing the Expressions (B.19) and (B.27), one can easily guess the formula of the minimum average codeword length of $\mathcal{U}$ obtained by optimizing the orthonormal basis for each isometry in the family $\mathcal{U}$. We chose to write the details of the first two steps rather than the formal induction step, since it is easier to be followed by the reader. The minimum average length of $\mathcal{U}$ is equal to

$$
\sum_{[n_1,n_{(m-1)l}]}^{N} p[1,(m-1)l]\sum_{i=1}^{d^l}\lambda_i^{[1,(m-1)l]}\ell_i + \sum_{[n_1,n_{(m-2)l}]}^{N} p[1,(m-2)l]\sum_{i=1}^{d^l}\lambda_i^{[1,(m-2)l]}\ell_i +
$$

$$
\sum_{[n_1,n_{(m-3)l}]}^{N} p[1,(m-3)l]\sum_{i=1}^{d^l}\lambda_i^{[1,(m-3)l]}\ell_i + \cdots + \sum_{n_1=1}^{N} p[1,l]\sum_{i=1}^{d^l}\lambda_i^{[1,l]}\ell_i + \sum_{i=1}^{d^l}\lambda_i\ell_i
$$

$$
= \sum_{j=1}^{m-1}\left(\sum_{[n_1,n_{(m-j)l}]}^{N} p[1,(m-j)l]\sum_{i=1}^{d^l}\lambda_i^{[1,(m-j)l]}\ell_i\right) + \sum_{i=1}^{d^l}\lambda_i\ell_i, \tag{B.28}
$$

where $\{\lambda_i\}_{i=1}^{d^l}$ are the eigenvalues of the ensemble state ρ_l, i.e.

$$
\rho_l = \sum_{[n_1,n_l]}^{d^l} p[1,l]\left|s_{[1,l]}\right\rangle\!\left\langle s_{[1,l]}\right| = \sum_{i=1}^{d^l}\lambda_i\left|\lambda_i\right\rangle\!\left\langle\lambda_i\right|,
$$

where λ_i is the eigenvalue corresponding to the eigenvector $|\lambda_i\rangle$.

Similarly to Equations (B.18) and (B.26), we obtain that the average codeword length of $\mathcal{U}$ will be minimized if we use the following isometries to encode each of the blocks:

$$U^{n_1,\ldots,n_{(m-j)l}} = \sum_{i=1}^{d^l} \left| \psi_i \right\rangle\!\left\langle \lambda_i^{[1,(m-j)l]} \right|, \text{ for } j = 1,\ldots,m-1, \text{ and } U = \sum_{i=1}^{d^l} |\psi_i\rangle\langle\lambda_i|. \quad \text{(B.29)}$$

Equation (B.28) can be re-written as

$$\sum_{j=2}^{m} \left(\sum_{[n_1,n_{(j-1)l}]}^{N} p[1,(j-1)l] \sum_{i=1}^{d^l} \lambda_i^{[1,(j-1)l]} \ell_i \right) + \sum_{i=1}^{d^l} \lambda_i \ell_i, \quad \text{(B.30)}$$

Now we move to the optimization of the sequence of jointly orthonormal length codewords $(|\psi_i\rangle)_{i=1}^{d^l}$. By the forward direction of quantum Kraft-McMillan inequality (Theorem 4.3), any finite sequence of quantum codes from the special block code is uniquely decodable and the length sequence $(\ell_i)_{i=1}^{d^l}$ of the jointly orthonormal sequence of length codewords $(|\psi\rangle)_{i=1}^{d^l}$ must satisfy the classical Kraft-McMillan inequality

$$\sum_{i=1}^{d^l} 2^{-\ell_i} \leq 1. \quad \text{(B.31)}$$

Also recall that without loss of generality, the Inequality (B.16) is satisfied. Thus, in order to minimize the average codeword length of $\mathcal{U}$, we need to choose the optimal length sequence $(\ell_i)_{i=1}^{d^l}$, which in addition to satisfying Equations (B.31) and(B.16), also minimizes Equation (B.30). Notice that a complete binary Kraft tree of height $\lceil \log_2 d^l \rceil$ has at least d^l many leaf nodes which are enough to assign uniquely decodable codewords to any set of d^l many symbols. Therefore, for any optimal length sequence $(\ell_i)_{i=1}^{d^l}$, we have $1 \leq \ell_i \leq \ell_{max} = \lceil \log_2 d^l \rceil$ for every i. So, it is clear that one needs to test finitely many such sequences of lengths $(\ell_i)_{i=1}^{d^l}$ to find the optimal length sequence, and therefore the minimizer that we seek is well defined.

Once the sequence $(\ell_i)_{i=1}^{d^l}$ has been chosen as described in the above paragraph, then choose a classical binary sequence $(\omega_i)_{i=1}^{d^l}$ using the binary Kraft tree as shown in [13, Page 107, Theorem 5.2.1] such that the length of the codeword ω_i (i.e. the number of its binary digits), is equal to ℓ_i. Now, choose the sequence $(|\psi_i\rangle)_{i=1}^{d^l}$ to be equal to the sequence of qubit strings $(|\omega_i\rangle)_{i=1}^{d^l}$. It is worth mentioning that the codewords $(\omega_i)_{i=1}^{d^l}$ obtained from the construction given in [13, Page 107, Theorem 5.2.1] are classical prefix-free codes. However, it is well known by the classical Kraft-McMillan inequality (and its converse) that for every uniquely decodable codeword sequence with the lengths $(\ell_i)_{i=1}^{d^l}$, there also exists a prefix-free codeword sequence with the same lengths. So, it suffices to construct the sequence of jointly orthonormal length codewords $(|\psi_i\rangle)_{i=1}^{d^l}$ from the classical prefix-free codes without affecting the optimality. Additionally, since $(\omega_i)_{i=1}^{d^l}$ is prefix-free (and hence, uniquely decodable), their concatenations denoted by $\omega_{i_1} \circ \cdots \circ \omega_{i_m}$ for $i_1, \ldots, i_m \in \{1, \ldots, d^l\}$ are pairwise distinct for any $m \in \mathbb{N}$. Since each $|\omega_i\rangle \in (\mathbb{C}^2)^{\otimes \ell_i}$, the indeterminate length of $|\omega_i\rangle$ is equal to $\text{Tr}(\Lambda |\omega_i\rangle\langle\omega_i|) = \ell_i$. Hence, each $|\omega_i\rangle$ is a length codeword. Therefore, their concatenations $|\omega_{i_1} \circ \cdots \circ \omega_{i_m}\rangle$ of length m are always well defined and orthonormal for every $m \in \mathbb{N}$. So, the sequence $(|\omega_i\rangle)_{i=1}^{d^l}$ forms a sequence of jointly orthonormal length codewords. Lastly, notice that this sequence $(|\omega_i\rangle)_{i=1}^{d^l}$ is not unique as one can find several such jointly orthonormal

sequences of length codewords each having the same optimal length sequence. Finally, the optimal special block code is then given by

$$\mathcal{V} = \left\{ V^{n_1,\ldots,n_{(k-1)l}} : k \in \{1,\ldots,m\}, \text{ and } n_1,\ldots,n_{(k-1)l} \in \{1,\ldots,N\} \right\},$$

(where for $k = 1$, the string $n_1,\ldots,n_{(k-1)l}$ is the empty string, and $V^{n_1,\ldots,n_{(k-1)l}}$ is simply denoted by V), defined by

$$V^{n_1,\ldots,n_{(k-1)l}} = \sum_{i=1}^{d^l} |\omega_i\rangle\langle\lambda_i^{n_1,\ldots,n_{(k-1)l}}|,$$

(where for $k = 1$, the vectors $(|\lambda_i^{n_1,\ldots,n_{(k-1)l}}\rangle)_{i=1}^{d^l}$ stand for the eigenvectors $(|\lambda_i\rangle)_{i=1}^{d^l}$ of the ensemble state ρ_l).

Since the length sequence $(\ell_i)_{i=1}^{d^l}$ of the obtained sequence of jointly orthonormal length codewords $(|\omega_i\rangle)_{i=1}^{d^l}$ minimizes Equation B.30, it follows that Equation (11) is satisfied and therefore, $L(\mathcal{V})$ is the lower bound on the average codeword length over all special block codes. □

Appendix C Proofs of Theorem 3.1 and Corollary 3.2

The proof of Theorem 3.1 proceeds similarly as that of Theorem 2.18.

Sketch of the proof of Theorem 3.1: As before, fix m, l, and an arbitrary k^{th} block conditional ensemble state $\rho^{n_1,\ldots,n_{(k-1)l}} = \sum_{i=1}^{d^l} \lambda_i^{n_1,\ldots,n_{(k-1)l}} |\lambda_i^{n_1,\ldots,n_{(k-1)l}}\rangle\langle\lambda_i^{n_1,\ldots,n_{(k-1)l}}|$ such that $\lambda_r^{n_1,\ldots,n_{(k-1)}} \geq \lambda_s^{n_1,\ldots,n_{(k-1)l}}$ for $r \leq s$. Define $\mathfrak{L}$ as before. Fix a constrained special block code $\mathcal{U} = \{U_k\}_{k=1}^{m}$ such that each isometry $U_k = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^k| \in \mathcal{U}$ has $(|\psi_i\rangle)_{i=1}^{d^l}$ as common jointly orthonormal sequence of length codewords and an arbitrary orthonormal sequence $(|e_i^k\rangle)_{i=1}^{d^l}$. For every k^{th} block, Alice uses the isometry U_k to encode the symbols. Suppose that $\text{length}(|\psi_i\rangle) = \ell_i \in \mathbb{N}$ and without loss of generality, assume $\ell_r \leq \ell_s$ for $r \leq s$. By the forward direction of quantum Kraft-McMillan inequality (Theorem 4.3), we have that the set of non-negative integer lengths $\{\ell_i\}_{i=1}^{d^l}$ belongs to $\mathfrak{L}$.

In the sketch of the proof of Theorem 2.18, by using the Birkhoff-von Neumann Theorem and Lemma 4.2, we argued that for a given set of lengths $(\ell_i)_{i=1}^{d^l} \in \mathfrak{L}$, in order to achieve the absolute minimization of the average codeword length, every $\rho^{n_1,\ldots,n_{(k-1)l}}$ needs to be encoded using its corresponding optimal isometry $U^{n_1,\ldots,n_{(k-1)l}} = \sum_{i=1}^{d^l} |\psi_i\rangle\langle\lambda_i^{n_1,\ldots,n_{(k-1)l}}|$. However, Alice now favors simplicity at the cost of higher average codeword length, and uses a fixed isometry $U_k = \sum_{i=1}^{d^l} |\psi_i\rangle\langle e_i^k|$ to encode each $\rho^{n_1,\ldots,n_{(k-1)l}}$ for every sequence of states $|s_{n_1},\ldots,s_{n_{(k-1)l}}\rangle$ as the first $(k-1)l$ emissions. So, we can assert that

$$\text{Tr}\left(U_k \rho^{n_1,\ldots,n_{(k-1)l}} U_k^\dagger \Lambda\right) \geq \text{Tr}\left(U^{n_1,\ldots,n_{(k-1)l}} \rho^{n_1,\ldots,n_{(k-1)l}} (U^{n_1,\ldots,n_{(k-1)l}})^\dagger \Lambda\right) \qquad \text{(C.1)}$$

Now the average codeword length for encoding the kth block ensemble state

$$\rho^k = \sum_{n_1,\ldots,n_{(k-1)l}=1}^{N} p(n_1,\ldots,n_{(k-1)l})\rho^{n_1,\ldots,n_{(k-1)l}}$$

is given by

$$\sum_{n_1,\ldots,n_{(k-1)l}} p(n_1,\ldots,n_{(k-1)l}) \operatorname{Tr}\left(U_k \rho^{n_1,\ldots,n_{(k-1)l}} U_k^\dagger \Lambda\right) \tag{C.2}$$

for the fixed set of lengths $\{\ell_i\}_{i=1}^{d^l}$. Since U_k is independent of $\left|s_{n_1},\ldots,s_{n_{(k-1)l}}\right\rangle$, using the linearity of trace and the definition of ρ^k (Definition 2.4), Equation (C.2) can be re-written as

$$\operatorname{Tr}\left(U_k\left(\sum_{n_1,\ldots,n_{(k-1)l}} p(n_1,\ldots,n_{(k-1)l})\rho^{n_1,\ldots,n_{(k-1)l}}\right)U_k^\dagger \Lambda\right) = \operatorname{Tr}\left(U_k\rho^k U_k^\dagger \Lambda\right). \tag{C.3}$$

Consider the spectral decomposition of the kth block ensemble state, $\rho^k = \sum_{i=1}^{d^l} \lambda_i^k \left|\lambda_i^k\right\rangle\!\left\langle\lambda_i^k\right|$ such that $\lambda_r^k \geq \lambda_s^k$ for $r \leq s$. Then, it can be shown again by using the Birkhoff-von Neumann Theorem and Lemma 4.2 that $\operatorname{Tr}\left(U_k\rho^k U_k^\dagger \Lambda\right)$ is minimized for the given set of lengths $\{\ell_i\}_{i=1}^{d^l}$ when $U_k = \sum_{i=1}^{d^l} \left|\psi_i\right\rangle\!\left\langle\lambda_i^k\right|$. If we use $U_k = \sum_{i=1}^{d^l} \left|\psi_i\right\rangle\!\left\langle\lambda_i^k\right|$, then $\operatorname{Tr}\left(U_k\rho^k U_k^\dagger \Lambda\right)$ simplifies to $\sum_{i=1}^{d^l} \lambda_i^k \ell_i$. Summing such length for each ρ^k gives Equation (14), which is the minimum average codeword length of encoding m blocks for a fixed set of lengths $\{\ell_i\}_{i=1}^{d^l}$. Refer to the paragraph following Equation (B.21) in the Appendix B for the details. Minimizing Equation (14) over all sets of lengths in $\mathfrak{L}$ gives Equation (15), which is the minimum average codeword length of encoding m blocks over the set of all constrained special block codes. Once the minimizer set of lengths is found, one can then construct an optimal jointly orthonormal sequence of length codewords $(\left|\omega_i\right\rangle)_{i=1}^{d^l}$ and define a minimizer special block code $\mathcal{V}$ as stated in the theorem. $\square$

From Equation (C.1), it should now be clear that for the same set of lengths $\{\ell_i\}_{i=1}^{d^l}$, the minimum average codeword length produced by this non-adaptive compression for encoding ρ^k is at least as large as that produced by the adaptive compression for every k^{th} block. Precisely, this means that for the same set of lengths $\{\ell_i\}_{i=1}^{d^l}$,

$$\sum_{i=1}^{d^l} \lambda_i^k \ell_i \geq \sum_{n_1,\ldots,n_{(k-1)l}} p(n_1,\ldots,n_{(k-1)l}) \sum_{i=1}^{d^l} \lambda_i^{n_1,\ldots,n_{(k-1)l}} \ell_i$$

for every k^{th} block. Consequently, it's not hard to see that the minimum average codeword length produced by the non-adaptive compression over the set of all constrained special block codes for encoding m blocks of pure states is at least as large as that produced by the adaptive compression over the set of all special block codes.

The proof of Corollary 3.2 proceeds as follows: If the quantum source is stationary, then we can show as below that all the kth block ensemble states ρ^k's are identical. Indeed, from Equations (3) and (4), we have $\rho^k =$

$$\sum_{n_1,\ldots,n_{(k-1)l}=1}^{N} \sum_{n_{(k-1)l+1},\ldots,n_{kl}=1}^{N} p(n_1,\ldots,n_{kl}) \left|s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right\rangle\!\left\langle s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right|$$

$$= \sum_{n_{(k-1)l+1},\ldots,n_{kl}=1}^{N} \left|s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right\rangle\!\!\left\langle s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right| \sum_{n_1,\ldots,n_{(k-1)l}=1}^{N} p(n_1,\ldots,n_{kl})$$

$$= \sum_{n_{(k-1)l+1},\ldots,n_{kl}=1}^{N} \left|s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right\rangle\!\!\left\langle s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right|$$

$$\mathbb{P}(X_{(k-1)l+1} = n_{(k-1)l+1},\ldots,X_{kl} = n_{kl})$$

$$= \sum_{n_{(k-1)l+1},\ldots,n_{kl}=1}^{N} \left|s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right\rangle\!\!\left\langle s_{n_{(k-1)l+1}}\cdots s_{n_{kl}}\right| \mathbb{P}(X_1 = n_{(k-1)l+1},\ldots,X_l = n_{kl})$$

$$\text{(from Equation (2))}$$

$$= \sum_{n_1,\ldots,n_l=1}^{N} \left|s_{n_1}\cdots s_{n_l}\right\rangle\!\!\left\langle s_{n_1}\cdots s_{n_l}\right| \mathbb{P}(X_1 = n_1,\ldots,X_l = n_l) \quad \text{(re-writing above equation)}$$

$$= \sum_{n_1,\ldots,n_l=1}^{N} \left|s_{n_1}\cdots s_{n_l}\right\rangle\!\!\left\langle s_{n_1}\cdots s_{n_l}\right| p(n_1,\ldots,n_l) \quad \text{(from Equation (1))}$$

$$= \rho_l \quad \text{(from Definition 2.3)}.$$

Therefore, we have $\rho^1 = \cdots = \rho^k = \rho_l$. Consider the spectral decomposition of ρ_l and the isometry $V = \sum_{i=1}^{d^l} |\omega_i\rangle\langle\lambda_i|$ as in the statement of the corollary. This isometry V is the optimal isometry [5, Theorem 2] for encoding the state ρ_l. As a result, it turns out that in order to optimally encode every ρ^k, Alice can use the same isometry V. In other words, the minimizer constrained special block code for a quantum stationary source is $\mathcal{V} = V$. Therefore, the minimum average codeword length for encoding m blocks of states will be $ILC(\mathcal{S},m,l) = m\,\mathrm{Tr}\big(V\rho_l V^\dagger\Lambda\big) = m\sum_{i=1}^{d^l}\lambda_i\ell_i$. Then the minimum average codeword length per block is $\frac{ILC(\mathcal{S},m,l)}{m} = \mathrm{Tr}\big(V\rho_l V^\dagger\Lambda\big) = \sum_{i=1}^{d^l}\lambda_i\ell_i = ILC(\mathcal{S},1,l)$. Notice that the minimum average codeword length per block depends only on the block size l and is independent of the number of blocks m. We can now directly apply [5, Theorem 3] to bound this minimum average codeword length per block from above and below in terms of the von-Neumann entropy of the source as follows:

$$S(\rho_l) \le \sum_{i=1}^{d^l} \lambda_i\ell_i \le S(\rho_l) + 1, \tag{C.4}$$

where $S(\rho_l)$ represents the von-Neumann entropy of ρ_l. Dividing the above Inequality (C.4) throughout by the block size l, one obtains

$$\frac{S(\rho_l)}{l} \le \frac{1}{l}\sum_{i=1}^{d^l} \lambda_i\ell_i \le \frac{S(\rho_l)}{l} + \frac{1}{l}, \tag{C.5}$$

where $\frac{1}{l}\sum_{i=1}^{d^l}\lambda_i\ell_i = \frac{ILC(\mathcal{S},1,l)}{l}$ represents the minimum average codeword length per symbol over all constrained special block codes. Finally, taking the limit $l \to \infty$ in the above Inequality C.5, we get

$$\lim_{l\to\infty} \frac{ILC(\mathcal{S},1,l)}{l} = \lim_{l\to\infty} \frac{S(\rho_l)}{l} = \lim_{l\to\infty} \frac{1}{l}\sum_{i=1}^{d^l} \lambda_i\ell_i \tag{C.6}$$

where $\lim_{l \to \infty} \frac{S(\rho_l)}{l}$ represents the von-Neumann entropy rate of the stationary source which is known to always exist [2, Theorem 9.14]. $\square$